FUNCTIONAL CHARACTERISATION of *CONSTITUTIVE EXPRESSER* of *PATHOGENESIS-RELATED GENES 5*

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Molecular Biology

at Massey University, Palmerston North, New Zealand.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Muhammad Faisal

2017

ABSTRACT

As reported previously, CPR5 negatively regulates the onset of leaf death, hypersensitive response, disease resistance and early leaf senescence. *cpr5* plants contain aberrant trichomes and higher levels of ROS, SA and JA. Cell-cycle, JA/ET, ABA and sugar signalling are also affected in *cpr5* plants. These results suggest that CPR5 is a master regulator of multiple processes. However, how CPR5 manages to exert pleiotropic effects is still poorly understood. The first objective of the current study was the purification of the CPR5 protein to solve its crystal structure. Extensive *in silico* analyses were carried out and the results showed that CPR5 is predicted to be a membrane protein with 4 or 5 transmembrane (TM) domains. Additionally, CPR5 contains intrinsically disordered regions (IDRs) at its N-terminus. Proteins containing IDRs and TM domains are often difficult to purify for crystallization studies. Therefore, the undesirable regions of CPR5 such as, IDR and TM domains were deleted and a set of 24 constructs were developed. Despite several efforts, none of the CPR5 recombinant proteins were isolated. In addition to predicting IDR and TM domains, *in silico* results also predicted three NLS-encoding clusters, casein kinase phosphorylation sites, multiple start codons, coiled-coil domains and glycine motifs. To find out the roles of these putative structural elements on CPR5 functions, firstly a *CPR5* cDNA was synthesised and termed as *SynCPR5*. Subsequently, predicted sites or motifs were mutated in *SynCPR5* through site-directed mutagenesis and a set of 25 mutated *CPR5* transgenes (cDNA constructs) were developed. Using a complementation strategy, all the constructs were transformed into *cpr5-2* plants. The results show that the complementation of *cpr5-2* plants with *SynCPR5*, fully restored HR-like lesions, wildtype-like trichomes and leaves on *SynCPR5* plants. Further physiological characterization such as, transcript abundance of *SynCPR5*, *PR1*, *PR5* and *PDF1.2*, leaf area measurements and ploidy levels showed that *CPR5* regulates some of its functions and phenotypes quantitatively as well as qualitatively. When compared with the wildtype, better growth (larger leaves) but enhanced disease susceptibility was found in *metCPR5* transgenic lines (in which putative start codons were mutated), indicating that CPR5 regulates a balance between growth and resistance. Functional characterization of NLS mutants (*nlsCPR5*) showed that NLS-encoding clusters are important for CPR5 proper

functions. However, current evidence is insufficient to relate their role in CPR5 localization. Moreover, *in silico* results show that putative NLS clusters are present in the region of CPR5 which were annotated as intrinsically disordered region (IDR). Similar phenotypes shown by both *nlsCPR5* and *Del63CPR5* (in which the first 63 amino acids of CPR5 including putative NLS were deleted), indicate that the putative NLS clusters could be part of IDR and may have dual functions. Loss-of-function phenotypes shown by coiled-coil domain mutants (*ccdCPR5*) reinforce the role of coiled-coil domains in CPR5 homo-dimerization. Moreover, in contrast to previous reports, the downregulation of *PDF1.2* in the majority of *CPR5* complementation lines proposes CPR5 to be a positive regulator of *PDF1.2*. Based on the results presented in the current study, putative CPR5 IDRs and coiled-coil domains are proposed to facilitate CPR5 dimerization in order to restrict the entry of deregulated cargos into the nucleus. Moreover, these results uncover a novel role of CPR5 in the regulation of balance between plant growth and resistance. Furthermore, this study, for the first time, reports evidence of the requirement of NLS clusters for CPR5 functions.

ACKNOWLEDGEMENTS

*All glory and praises for **ALLAH**, The most Beneficent, The most Merciful, Who is the only source of entire and complete knowledge and wisdom endowed to mankind. It is a matter of great honour and pleasure for me to express my ineffable gratitude and profound indebtedness to the Higher Education Commission (HEC), Pakistan and Massey University, New Zealand for providing me the best educational and research facilities. I offer my humblest gratitude to my revered supervisor Dr **PAUL DIJKWEL** and mentors Drs **GILLIAN NORRIS** and **JASNA RAKONJAC** for their skilful guidance, enlightened views, valuable suggestions, constructive criticism, unfailing patience and inspiring attitude during my studies and research.*

*I feel proud to have nice and helpful colleagues and friends in C5.19 and X-lab and I am grateful for their contributions, moral support and valuable suggestions. I also feel highly indebted to **VICTORIA SIBLEY** for her time and efforts spent in proofreading my dissertation. I would like to extend my gratitude and the deepest appreciation to my mother for her ever encouraging and supporting role in achieving my goals of life. Words are lacking to express my obligations to my affectionate mother, Ms **NASEEM AKHTAR**, for her love, best wishes, inspirations and unceasing prayers, without which the present destination would have been a mere dream. I am highly indebted to my brother, **MUHAMMAD ASLAM MUNNA** who exhibited great patience and good will throughout this study, and looked after the whole family.*

*I deem it my utmost pleasure to avail myself of opportunity to express the heartiest gratitude to my wife, **Dr TINA SEHRISH** for her love, patience, encouragement, good wishes and understanding that inspired me to accomplish this humble effort. I also owe a lot of love, care and time to my children, **AYESHA** and **MUHAMMAD** who could not get the proper time, love and attention during the course of my research and studies. I have no words to express as thanks to my family except to say them aloud, I love you all and I am lucky to have you all. Thank God for blessing me with a great, loving family.*

MUHAMMAD FAISAL

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

aa	Amino acid
ABA	Abscisic acid
ABI5	ABA-INSENSITIVE 5
acd6	*ACCELERATED CELL DEATH 6*
APC/C	anaphase-promoting complex/cyclosome
BRI	BRASSINOSTEROID-INSENSITIVE
BRs	Brassinosteroids
Ca	Calcium
CaM 35S promoter	Cauli Mosaic Virus 35S promoter
CC-domains	Coiled-coil domains
CDC*20*	*CELL DIVISION CYCLE 20*
CDH	*CDH HOMOLOG 1*
CDK	Cyclin-Dependant-Kinase
CK	Casein kinase
CKI	CDK inhibitor
CNGC	*Cyclic Nucleotide Gated Channels*
COI1	*CORONATINE-INSENSITIVE 1*
Col-0	*Arabidopsis* ecotype Columbia
CPR5	*CONSTITUTIVE EXPRESSER of PATHOGENESIS-RELATED GENES 5*
DAS	Days after sowing
EDR1	*ENHANCED DISEASE RESISTANCE 1*
EDS5	*ENHANCED DISEASE SUSCEPTIBILITY 5*
EIN2	*ETHYLENE INSENSITIVE 2*
ET	Ethylene
ETI	Effector-Triggered Immunity
FZR	*FIZZY-RELATED*
GA	Gibberellin
GeBP	*GL1 ENHANCER BINDING PROTEIN*
GFP	Green Fluorescent Protein
GIG1	*GIGAS cell 1*
GPLs	*GeBP*-like proteins
GST	Glutathione-S-transferases
H_2O_2	Hydrogen Peroxide
HR	Hypersensitive Response
HXK	*Hexokinase*
HYS1	*HYPERSENESCENCE1*
IDP	Intrinsically Disordered Protein
IDR	Intrinsically Disordered Region
JA	Jasmonic Acid
JAR1	*JASMONATE RESISTANT 1*
JAZ1	*JASMONATE-ZIM-DOMAIN PROTEIN 1*

K	Potassium
KRPs	*KIP-RELATED PROTEINS*
L*er*-0	*Landsberg erecta*
LOX	Lipoxygenases
LSD	*LESIONS SIMULATING DISEASE*
MeJA	Methyl jasmonate
MoRFs	Molecular Recognition Features
mRNA	Messenger RNA
NDGA	Nordihydroguaiaretic acid
NE	Nuclear Envelope
NLS	Nuclear Localization Signal
NO	Nitric oxide
NPC	Nuclear Pore Complex
NPR1	*NON-EXPRESSER of PATHOGENESIS-RELATED GENES 1*
OE lines	Overexpression lines
OLD1	*ONSET OF LEAF DEATH1*
OSD1	*OMISSION of SECOND DIVISION 1*
PAD4	*PHYTOALEXIN DEFICIENT 4*
PAMPs/ MAPMs	Pathogen- or microbe-associated molecular patterns
PCD	Programmed Cell Death
PCR	Poly Chain Reaction
PDF1.2	*PLANT DEFENSIN 1.2*
PIF	*PHYTOCHROME INTERACTING FACTOR*
PR1	*PATHOGENESIS-RELATED 1*
PRRs	Pathogen Recognition Receptors
PstDC3000	*Pseudomonas syringae* pv *DC3000*
PTI	PAMP-Triggered Immunity
qRT-PCR	Quantitative Reverse-Transcriptase Poly Chain Reaction
R proteins	Resistance proteins
RB	Retinoblastoma
ROS	Reactive Oxygen Species
RPM1	*RESISTANCE TO PSEUDOMONAS SYRINGAE PV MACULICOLA1*
RPS2	*RESISTANCE TO PSEUDOMONAS SYRINGAE*
SA	Salicylic Acid
SAGs	*SENESCENCE-ASSOCIATED GENEs*
SAR	Systemic Acquired Resistance
SIM	*SIAMESE*
SMR	*SIAMESE-RELATED*
SOD	Superoxide Dismutase
TFs	Transcription Factors
TM	Transmembrane
tRNA	Transfer RNA
UVI4	*UV INSENSITIVE 4*

Chapter 1 CPR5 as a master regulator- review of literature

1.1 INTRODUCTION

During the plant-pathogen arms-race, plants evolved a number of defence strategies and mechanisms in order to defend them from an array of pathogens. In contrast to animals, plants have rigid cell walls, a waxy cuticle and spikey trichomes, which help to protect them from pathogen ingress and attack (War et al., 2012). Plants produce various antimicrobial chemicals, such as, phytoalexins. Additionally, resistance is regulated in response to pathogen attack. Resistance can be classified as; 1) general resistance, the activation of which confers resistance to a broad spectrum of pathogens, and 2) specific resistance, which renders resistance to specific pathogens. In 1942, Flor showed that the inability of pathogen to infect host depends on the presence and detection of the pathogenesis gene by the corresponding gene present in the host. If the host has no corresponding gene, the pathogen is able to infect and spread. Whereas a host restricts pathogen ingress, upon the successful detection of pathogen presence through the corresponding gene and then activates defence-related responses. In cases where the presence of the corresponding gene is found in the host and there is unsuccessful infection growth, the pathogenesis gene is known as the *avirulence* gene. Whereas, in the event of the absence of the corresponding gene in the host and successful infection growth, the pathogenicity gene is known as the *virulence* gene (Flor, 1971). Overall, genes from pathogens or microbes are called pathogen- or microbe-associated molecular patterns (PAMPs/ MAPMs). Corresponding genes in the host are referred to as pathogen recognition receptors (PRRs), which are generally localised in the plasma membrane of the host cell. Upon pathogen attack, the PAMPs are recognised by the PRRs and trigger resistance, and hence immunity is conferred (Jones and Dangl, 2006). This type of immunity is known as PAMP-triggered immunity (PTI) and is believed to represent the first line of defence in the host (Jones and Dangl, 2006). In order to avoid PAMP-mediated detection, pathogens have managed to dampen PTI through the delivery of effectors (Jones and Dangl, 2006). However, over time, hosts have evolved to detect effectors through resistance (R) proteins, which have a specific type of receptors known as, TNL-NBS-LRR or CC-NBS-LRR. (McHale et al., 2006) Thus, plants detect the effectors delivered by the pathogens and confer resistance, known as effector-triggered immunity, which then renders the second layer of defence (Jones and Dangl, 2006). Additionally, defence responses are regulated through numerous biotic and abiotic factors, and have separate defence

systems for biotrophic and necrotrophic pathogens in order to reduce fitness cost (Wang et al., 2014b). For example, resistance to biotrophic pathogens is generally mediated through salicylic acid, whereas resistance to necrotrophic pathogens is mediated by jasmonic acid or ethylene (Jing et al., 2007).

1.1.1 CONSTITUTIVE EXPRESSER OF PATHOGENESIS-RELATED GENES (CPR) mutants

To combat pathogens, hosts recognise pathogen attack and respond promptly and efficiently through the production of antimicrobial proteins (phytoalexins) and/or the activation of defence responses and the execution of hypersensitive responses (Dangl et al., 1996). In order to study signal transduction and the activation of resistance pathways, a number of mutants were isolated through mutagenesis. Of these mutants, the first group displayed loss-of-resistance phenotypes e.g. *npr1* (*non-expresser of pathogenesis related genes*; (Cao et al., 1994, Delaney et al., 1995, Shah et al., 1999) and *eds5* (*enhanced disease susceptibility*) mutants (Rogers and Ausubel, 1997, Nawrath and Métraux, 1999). Both *npr1* and *eds5* displayed hyper-susceptibility to pathogen attack. Whereas the second class of mutants showed constitutive activation of defence-related genes and enhanced defence responses, such as, *acd6* (*accelerated cell death6*) and *cpr1* (*constitutive expresser of pathogenesis-related genes 1*; (Bowling et al., 1994). In addition to *cpr1*, there are a number of other loci which are known to show constitutive expression of *PATHOGENESIS-RELATED GENES* as well, such as, *cpr2* and *cpr6* (Clarke et al., 2000). All of the *cpr* mutants were found to be very important candidates for the dissection of resistance signaling and pathways, since *cpr5* mutants display upregulation of *PR* genes and resistance responses without pathogen attack.

1.1.2 CONSTITUTIVE EXPRESSER of PATHOGENESIS-RELATED GENES 5 (CPR5) and other CPR5 mutant alleles

In addition to the isolation of *cpr1*, *cpr2* and *cpr6* (Bowling et al., 1994), Bowling and co-workers found another *cpr* mutant, which also displayed the constitutive expression of *pathogenesis-related genes PR1* and *PR5* (Bowling et al., 1997). In addition to *PR1* and *PR5*, the levels of another plant defence gene, *PDF1.2*, were also upregulated in this mutant. Additionally, mutant plants also showed HR-like spontaneous lesions on their leaves without pathogen attack. Thus, Bowling et al., (1997) named this locus, *CONSTITUTIVE EXPRESSER OF PATHOGENESIS-RELATED GENES 5* (*CPR5*). They isolated two mutant alleles of *cpr5*;

cpr5-1 and *cpr5-2*, which have mutations at different positions. While screening T-DNA mutagenized plants in order to find out genes necessary for trichome development, Bechtold and co-workers (Bechtold and Pelletier, 1998) also found a mutant whose phenotypes were similar to *cpr5* mutants. Therefore, the mutants were referred to as *cpr5-T1* and *cpr5-T2*, both of which have T-DNA insertions in the first intron. Similarly, Jing and co-workers (Jing et al., 2002) found a mutant which displayed hypersensitive (HR)-like lesions and the onset of leaf death. Therefore, they named this locus, *ONSET of LEAF DEATH1* (*OLD1*). Three independent mutant alleles (*old11-1, old11-2 and old11-3*) of *OLD1* are reported by Jing et al., (2002). Despite mutations at different positions, all of *old1* mutants displayed similar phenotypes. In the same year, a mutant was discovered by Yoshida and co-workers (Yoshida et al., 2002) which showed hypersensitivity to sugar and accelerated leaf senescence, which they referred to as *HYPERSENESCENCE1* (*HYS1)*. They reported two mutant alleles, *hys1-1* and *hys1-2*, of *HYS1*. When mapped, *CPR5*, *OLD1* and *HYS1* were all found to be the same locus (gene) but were also found to contain mutations at different positions and share many phenotypes in common (Bowling et al., 1997; Jing *et al*., 2002; Yoshida et al., 2002). In 2011, Borghi and coworkers (Borghi et al., 2011) found another *CPR5* allele, "*cpr5-3"*, which was found to have a T-DNA insertion in the fourth exon. The *cpr5-3* plants showed K^+ reductions and other phenotypes in a similar fashion to *cpr5-1* and *cpr5-2*. *CPR5* alleles (*cpr5-1*, *cpr5-2*, *hys1-1* and *hys1-2*) are in *Arabidopsis* Columbia (Col-0) ecotype background, whereas *OLD1* (*old1-1*, *old1-2*, *old1-3*) and *cpr5-3* are in *Arabidopsis Landsberg erecta* (L*er*-0) ecotype backgrounds (Figure 1.1). *cpr5-2* mutant allele was used in the current study because 1) it is in Col-0 ecotype background, 2) whole genome sequence of Col-0 is available and 3) Col-0 is better ecotype for *Agrobacterium*-mediated transformations compared to L*er*-0 (Desfeux *et al.* 2000).

The CPR5 gene belongs to *Arabidopsis thaliana*, which is a small weed but has served as a model plant for decades (TAIR; http://arabidopsis.org/). According to the information available at TAIR, the genomic sequences of *CPR5*, *OLD1* or *HYS1* genes (*AT5G64930*) are comprised of 2630 nucleotides (Kirik et al., 2001, Yoshida et al., 2002). Compositionally, the *CPR5* gene is comprised of 4 exons and 3 introns, whereas the *CPR5* coding sequence consists of 1692 nucleotides which encodes 564 amino acids (Figure 1.1). The CPR5 protein is predicted to contain a bipartite nuclear localisation signal (NLS) at its amino terminus and 5 transmembrane (TM) domains predicted in the carboxyl terminus (Figure 1.1).

1.1.3 CPR5 Localizes in nuclear membrane

As mentioned above, the CPR5 protein is predicted to contain a bipartite nuclear localisation signal (Yoshida et al., 2002, Kirik et al., 2001). Based on the presence of a bipartite nuclear localisation signal (NLS), CPR5 was proposed to be targeted to the nucleus (Yoshida et al., 2002). However, the majority of the cytoplasm of the cell was found to be fluorescent when GFP was fused to the C-terminus of the CPR5 protein. Thus, Gao and co-workers (Gao et al., 2011) suggested that CPR5 embeds either in the plasma or in the outer-nuclear membrane. In the same year, Perazza and co-workers (Perazza et al., 2011), found CPR5 to be localized in the nucleus when CPR5 was fused to GFP at the N-terminus of CPR5. As shown (Gao et al., 2011), all of the cytoplasm, including the nucleus, appears to be full of CPR5 protein. It could be possible that the overexpression of *CPR5* by the 35S promoter could have resulted in the non-specific (everywhere) localisation of the CPR5 protein in Gao et al., (2011) studies. Whereas, GFP could have escorted CPR5 to the nucleus and resulted in the non-native localisation of CPR5 in the studies conducted by Perazza et al., (2011). Surprisingly, both *CPR5* transgenes used in the studies Gao et al., (2011) and Perazza et al., (2011), managed to rescue the majority of *cpr5* phenotypes. Recently, CPR5 is shown to be located in the nuclear membrane as one of the nucleoporin (Wang et al., 2014b, Gu et al., 2016), which supports the Perazza et al,. (2011) results. CPR5 is predicted to contain a bipartite nuclear localisation signal (Kirik et al., 2001) however the role of putative bipartite NLS in CPR5 localisation has not been characterized in any of the aforementioned studies.

1.1.4 *CPR5* is constitutively expressed in the majority of plant tissues

In order to check the expression levels of *CPR5*, the transcript abundance of *CPR5* was quantified in various plant tissues. *CPR5* is shown to be expressed in the majority of developmental stages, such as, rosette leaves, seedlings, roots, siliques, stems and flowers (Kirik et al., 2001; Gao et al., 2011). Since CPR5 is a lowly expressed gene or protein, the quantification of CPR5 protein, and a comparison between CPR5 and cpr5 recombinant protein, have not been shown by any of the aforementioned studies. However, Jing et al., (2002) for the first time, quantified the transcript abundances of *CPR5* and *cpr5* mutant genes using real-time qPCR. Their results showed that mutation in *CPR5* resulted in a significant reduction of *cpr5* transcript abundance as compared to the wildtype. Based on the loss-of-functions phenotypes shown by the *cpr5* alleles, these mutations were considered as null mutations, and

resultant cpr5 mutant proteins were assumed to be non-functional. In summary, these results show that *CPR5* is constitutively expressed in the majority of tissues.

1.2 Role of CPR5 in programmed cell death, hypersensitive response and pathogen resistance

Whether a cell should stay alive or die, is of central importance for the survival of multicellular organisms and largely depends on the type, function and/or need of cells (Bruggeman et al., 2015). Cells die as a result of overwhelming stress or damage, to which the organisms or cells have no resilience to survive (Okada and Mak, 2004, Degterev and Yuan, 2008). Cells also die in a very sophisticated, efficiently regulated and orchestrated process, called programmed cell death (PCD). PCD is highly conserved and is a completely genetically directed mechanism across many species (Gilchrist, 1998, Collazo Cordero et al., 2006). PCD can be divided into two types, autolytic and non-autolytic PCD. Autolytic PCD involves the formation of a large lytic vacuole and the prompt clearance of cytoplasm when the tonoplast is ruptured and hydrolases are released (van Doorn, 2011). On the other hand, rapid cytoplasm clearance is not accomplished in the non-autolytic type of PCD. Additionally, autolytic PCD commonly occurs during development, whereas non-autolytic PCD is frequently observed in response to stresses (van Doorn, 2011). PCD can be regulated in many ways (Collazo Cordero et al., 2006). Firstly, PCD is inherently initiated by default in order to shape the different organs of an organism during differentiation and organogenesis (Norbury and Hickson, 2001, Elmore, 2007, Christofferson and Yuan, 2010). Secondly, PCD can be triggered in response to stresses (Gilchrist, 1998, Fesik, 2000, Khurana et al., 2005, Elmore, 2007). Thirdly, programmed cell death can also occur in response to a number of external cues, such as, radiation exposure, drugs and/or different hormones (Elmore, 2007).

Programmed cell death in animals is referred to as apoptosis (Kerr et al., 1972). Apoptosis is characterized as the disassembly of various systems e.g. the blebbing of plasma membrane, shrinkage of cytoplasm, loss of cell-cell to contact, and/or the shearing and fragmentation of DNA in unwanted cells (Ellis et al., 1991). Apoptosis includes the elimination of cells, which are: 1) primarily formed for temporary functions e.g. a tadpole's tail; 2) epigenetically produced in abundance, such as, vertebrate neurons; 3) Müllerian duct cells, which are required for sex determination in females but not in males; and/or 4) cells which are destined to die during cell specialization e.g. keratinocytes (Jacobson et al., 1997).

PCD is inevitable for normal development and stresses in plants (Ellis et al., 1991, Fesik, 2000, Elmore, 2007, Williams and Dickman, 2008) and plays an important role during various vegetative and reproductive phases of plant growth and development (Greenberg et al., 2000, Collazo Cordero et al., 2006). For example, PCD is of central importance in: (1) the degeneration of pollen nuclei to prevent self-pollination in out-crossing plant species (Wilkins et al., 2015, Thomas and Franklin-Tong, 2004); (2) the formation of vascular bundles i.e. xylem and phloem (Bonke et al., 2003); (3) the abscission of leaves, petals and fruit from plants (Yamada et al., 2007); (4) the death of petals after fertilization (Pennell and Lamb, 1997); (5) root cap development (Wang et al., 2014a)); (6) post-embryonic decay of the suspensor and aleuronic layers (Yeung and Meinke, 1993, Kuo et al., 1996); (7) somatic and zygotic embryogenesis (McCabe et al., 1997); (8) leaf senescence (Jing et al., 2003), and; (9) sex determination in plants(Dellaporta and Calderon-Urrea, 1994). Programmed cell death in plants can also be activated during hypoxia and in waterlogged conditions (Armstrong, 1980) in order to form aerenchyma cells (Jackson et al., 1985).

The execution of the hypersensitive response is a classic example of programmed cell death in plants (Khurana et al., 2005). Upon pathogen challenge, plants are expected to recognize pathogen(s) through receptors, known as pathogen- or microbe-associated molecular patterns (PAMPs or MAMPS) and restrict pathogen growth through hypersensitive response (HR) and the induction of defence responses (Dangl et al., 1996). Upon successful pathogen recognition, plant defence system immediately induce the production of reactive oxygen species (ROS) and nitric oxide (NO) to restrict the growth of infection by a robust and local collapse of infected cells; the hypersensitive response (HR) (Dangl et al., 1996). HR-induced cell death is a genetically mediated process (Lamb and Dixon, 1997, Moeder and Yoshioka, 2008) and is largely similar to programmed cell death (Khurana et al., 2005, Collazo Cordero et al., 2006). HR is reported to transduce the plant defence system and systemic acquired resistance (SAR) (Kovalchuk et al., 2003). HR-like lesions are also mimicked by another process known as, necrosis, a type of cell death induced by damage or injury (Mizuno et al., 2010, Inaba et al., 2011). The regulation of lesions induced by HR or PCD is highly complex, and it is very difficult to discern one from the other (Mur et al., 2008).

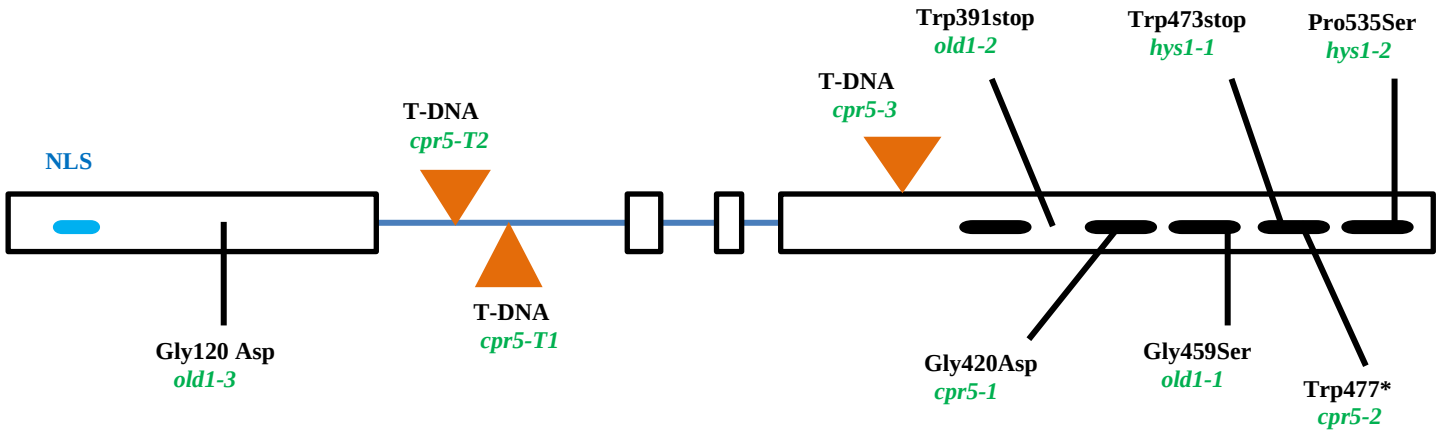


Figure 1.1 Annotated architecture of *CPR5* Gene

This figure represents position of annotated NLS, TM domains, T-DNA insertions, and various mutations reported in *CPR5*. Black rectangular boxes represent exons, whereas the blue lines between the boxes shows introns. Black bold bars in the fourth exon represent the position of annotated 5 TM domains, whereas, the blue bar in the first exon indicates the position of the predicted nuclear localisation signal (NLS). The orange-coloured pointed arrows in the first intron and fourth exon indicate the positions of T-DNA insertion in *CPR5*. The asterisk represents stop codon resulted by a mutation in *cpr5-2* allele of *CPR5*, and the numbers shows the positions of the amino acids changed due to mutations in the said allele.

HR is generally executed in response to pathogen recognition, defence responses (Mur et al., 2008) the generation of reactive oxygen species (ROS) and/or nitric oxide (NO) (Pennell (Pennell and Lamb, 1997), and increased Ca^{+2} and H^{+} influx or K^{+} efflux (Levine et al., 1996). A number of mutants have been isolated, which display spontaneous (HR) lesions without pathogen attack. For example, *Arabidopsis* mutant *acd2* (*accelerated cell death2*), *lsd* (*lesions simulating disease*), and *cpr5* (*constitutive expresser of pathogenesis related genes 5*), all of which display HR-like lesions without pathogen attack (Dietrich et al., 1994, Greenberg, 1996, Bowling et al., 1997, Jing et al., 2002). The exhibition of HR lesions without pathogen attack makes such genes (*cpr5*) unique from other resistance genes and important candidates for the elucidation of regulation of defence-related pathways in the absence of external stimuli.

As reported, *cpr5* is shown to have spontaneous lesions, early leaf senescence, enhanced resistance and increased ROS levels. Thus, *CPR5* could have a direct or indirect suppressive role in the regulation of the aforementioned processes. Since HR-like lesions, senescence and SAR are types of PCD regulation, their activation in *cpr5* plants suggests that CPR5 has a direct or indirect role in the regulation of programmed cell death. Moreover, all of these processes are integrally interconnected with each other. Therefore, it is impossibly difficult to show which process is induced in the first place and which in turn is regulated in second place. For example, none of the studies showed that HR-lesions in *cpr5* are the consequence of SAR or that SAR is the result of HR activation. A number of efforts have been made to understand the nature, signalling and accomplishment of resistance in *cpr5*. For example, *PR1* and *PR5* are known markers for the confirmation of the upregulation of SA-mediated resistance in plants (Clarke et al., 2000). Similarly, *PDF1.2* is a marker gene for the confirmation of induction of resistance against necrotrophic pathogens. Additionally, *cpr5* plants confer resistance to both bacterial (hemi-biotrophic) and fungal (necrotrophic) pathogens. The upregulation of *PR1*, *PR5* and *PDF1.2* genes in *cpr5* (Bowling et al., 1997) suggests that both types of resistance i.e. resistance to biotrophic and necrotrophic pathogens, are upregulated in *cpr5*. These results indicate that *CPR5* may be involved in the suppression of non-specific (general) type resistance instead of resistance against specific pathogens.

To gain deeper insight into how resistance in *cpr5* is regulated, levels of salicylic acid were measured, and results showed that the level of salicylic acid (SA) in *cpr5* is higher than in the wildtype (Bowling et al., 1997). Thus, it was assumed that resistance in *cpr5* could be mediated by SA. The *npr1* (*non-expresser of pathogenesis-related proteins1)*, and *eds5* (*enhanced*

disease susceptibility5) mutants are defective in SA signaling and SA biosynthesis respectively, and *npr1* and *eds5* both display enhanced susceptibility when challenged with pathogens (Bowling et al., 1997; Clarke et al., 2000). The retention of resistance by *cpr5npr1* and loss of resistance in *cpr5eds5* double mutant plants, show that SA is required for *cpr5* resistance, however, *cpr5* resistance is independent of the SA-mediated pathway (Clarke et al., 2000, 2001). This notion was further strengthened when *cpr5*-mediated resistance was lost in *cpr5NahG* plants in which SA was supposed to be converted to catechol by a bacterial *NahG* gene (Bowling et al., 1997; Clarke et al., 2000).

In addition to the SA-mediated resistance pathway, generally, resistance to necrotrophic fungal pathogens is mediated by ethylene and/or jasmonic acid (Clarke et al., 2000). Since *cpr5* plants show elevated levels of the *PDF1.2* gene and JA hormone (Jing et al., 2007), the JA/ET-mediated signalling pathway was also explored. For this purpose, *cpr5jar1* (*JAR1*, JA signalling gene) and *cpr5ein2* (*EIN2*, ET signalling gene) double mutants were hypothesised to show susceptibility if *cpr5* resistance to a fungal pathogen was mediated by JA or ET. However, both of these double mutants exhibited resistance to bacterial and fungal pathogens in a similar fashion to *cpr5* (Clarke et al., 2000). These results suggest that the resistance activated in *cpr5* is independent of ethylene and jasmonic acid signalling pathways as well. Furthermore, the relationship of resistance (R) proteins with CPR5 was also studied. *EDS1* (*ENHANCED DISEASE SUSCEPTIBILITY1*), *RPS2* (*RESISTANCE TO PSEUDOMONAS SYRINGAE*), *PAD4* (*PHYTOALEXIN DEFICIENT 4*) and *RPM1* (*RESISTANCE TO PSEUDOMONAS SYRINGAE PV MACULICOLA1*) are some of the known R-genes whose mutant versions exhibit susceptibility (Clarke et al., 2001, Jirage et al., 2001). As shown by Clarke *et al.*, (2001) and Jirage *et al.*, (2001), *cpr5eds1*, *cpr5rpm1* and *cpr5pad4*, all showed susceptibility, except *cpr5rps2*, which displayed resistance in a *cpr5*-like fashion (Boch et al., 1998). These results show that functional R-genes, except *RPS2*, are required for *cpr5*-mediated resistance.

CPR5 is shown to be indirectly implicated in effector-triggered immunity (ETI) (Wang et al., 2014b). It was further proposed that CPR5 protein guards the cell-cycle inhibitors, *SIAMESE* (SIM) and *SIAMESE-RELATED1* (SMR1) proteins, which dissociate from CPR5 upon recognition of effectors by NB-LLR receptors and subsequently, phosphorylate CDK, which is required for cell cycle progression. Concomitantly, SIM and SMR1 also activate another unknown kinase, which results in retinoblastoma (RB) hyperphosphorylation and E2F

overactivation, leading to effector-triggered PCD and disease resistance (Wang et al., 2014b). Recently, CPR5 has been shown to be embedded in the nuclear membrane and has been proposed to act as one of the nucleoporins (Gu et al., 2016). CPR5 essentially acts as a gate and controls the entry of an array of cargoes e.g. NPR1, JAZ1 and ABI5 (Gu et al., 2016). This gating appears to be disturbed in *cpr5*, which allows the entry of a number of deregulated molecules. These deregulated cargoes, in turn, regulate different pathways and processes. Taken together, these results conspicuously propose a central role of *CPR5* in the execution of PCD, HR and resistance to bacterial and fungal pathogens. Despite higher SA, *PDF1.2* and JA, *cpr5*-mediated resistance is shown to be independent of the SA- or JA-mediated signalling pathways, which then raises the question as to how CPR5 suppresses SA and JA production. Alternatively, resistance induced in *cpr5* could be the consequence of the activation of some other process or pathway, for instance, leaf senescence (Section 1.2.1) or oxidative stress.

1.3 Leaf senescence regulation and CPR5

In addition to HR, leaf senescence is also a type of programmed cell death process, which is highly coordinated and genetically controlled (Buchanan-Wollaston, 1997). Leaf senescence is a terminal developmental stage of plant life, in which older leaves die and their nutrients are translocated to the young actively growing tissues, seeds, and storage organs (Buchanan-Wollaston, 1997). As discussed above, defence-related responses are activated in *cpr5* yet the exact mechanism of resistance regulation is poorly understood. As shown, early leaf senescence is reported in *cpr5* (Jing et al., 2002, Yoshida et al., 2002). The concurrent activation of defence and senescence pathways in *cpr5* indicates that the upregulated defence-related responses could be the result of induced leaf senescence in *cpr5*. Alternatively, leaf senescence could be the consequence of activation of resistance in *cpr5*, since senescence is also shown to be modulated in response to pathogen attack. For instance, biotrophic pathogens are known to promote senescence, whereas necrotrophic pathogens accelerate senescence (Häffner et al., 2015). Plant resistance and senescence are integrally interconnected with each other and share the same signalling cascade and many signalling components (Love et al., 2008, Rivas-San Vicente and Plasencia, 2011), *PHYTALEXINS DEFICIENT4* (*PAD4*) is reported to be involved in leaf senescence regulation, whereas *PAD4*-mediated premature leaf senescence in *Arabidopsis* contributed to basal resistance in response to the green peach aphid "GPA" (Pegadaraju et al., 2005). In addition to *PAD4*, *ENHANCED DISEASE RESISTANCE 1* (*EDR1*) is also shown to negatively regulate disease resistance and ethylene-induced leaf senescence

(Tang et al., 2005). These results suggest the existence of overlapping signalling pathways between resistance and leaf senescence.

Leaf senescence initiation and execution is shown to be dependent on a number of intrinsic and extrinsic stimuli and various age-related signals (Guo and Gan, 2012, Kim et al., 2015). The majority of plant hormones, such as, ethylene, SA, JA and ABA including brassinosteroids and oxidative stress (ROS), have been shown to play important roles in the regulation of growth and resistance, and are also key players for the mediation of leaf senescence regulation (Häffner et al., 2015). For instance, ethylene, which is reported to be implicated in growth and developmental stages, such as, seed germination, fruit ripening and resistance to necrotrophic pathogens (Masood et al., 2012, Nazar et al., 2014, Iqbal et al., 2017), also promotes leaf senescence (Jing *et al.*, 2002, 2005, 2007). Ethylene accelerates leaf senescence by activating the *EIN3* transcription factor, which in turn represses micro-RNA *miRNA164*, allowing the accumulation of *ANAC092* and *ORE1* TFs, the induction of which, positively regulates leaf senescence (Balazadeh et al., 2010a). Consistent with these studies, leaf senescence was found to be accelerated in *cpr5* by ethylene when applied exogenously (Jing et al., 2002) despite of having wildtype-like levels of ethylene (Jing et al., 2007). These results suggest that CPR5 has no direct role or effect on the regulation of ethylene biosynthesis, ethylene-mediated resistance and senescence signalling, since *cpr5ein2* showed no difference in leaf senescence or resistance compared to *cpr5* (Clarke et al., 2001).

ET and JA both act synergistically with regards to the regulation of resistance and leaf senescence (Penninckx et al., 1998). JA promotes senescence and this was first reported by Ueda and Kato (Ueda and Kato, 1980). Several JA-biosynthetic genes are found to be upregulated during senescence (He et al., 2002). Similarly, leaf senescence was not shown to be affected in *coi1* (defective in jasmonate signalling) when *coi1* plants were induced for senescence by JA application (He and Gan, 2002). However, JA appears to have no role in the regulation of said processes in *cpr5*, since *cpr5jar1* displayed senescence and resistance indifferently from *cpr5* (Clarke et al., 2001). Since *cpr5* seedlings show hypersensitivity to ET and JA, Jing et al., (2007) concluded that *cpr5* plants are shown to be defective in the perception, coordination and integration of ethylene signalling. Additionally, they proposed that *CPR5* ensures normal growth and development by repressing cell death during the early life of the plant (Jing et al., 2007).

Similar to ET and JA, SA is also shown to have a dual function in regards to resistance and leaf senescence activation. For example, senescing leaves have higher SA levels, and leaf longevity has been shown to be mediated through SA modulation by SA-3-hydroxylase (Zhang et al., 2013). Additionally, the *SENESCENCE-ASSOCIATED GENE101* (*SAG101*) is reported to interact with *PAD4* and *EDS1* and controls senescence (Zhu et al., 2011). Similarly, senescence was observed to be delayed in plants carrying a bacterial gene, *NahG*, which converts SA into catechol (Zhang et al., 2013). Thus, these results show that SA can regulate senescence. As discussed and established above in Section 1.2, SA is essential for *cpr5*-mediated resistance. Thus, SA could have induced both of the processes simultaneously, or the activation of one pathway, for instance disease resistance, could have transduced the other process i.e. senescence or *vice versa*.

The execution of resistance and senescence is shown to be affected by sugar levels (Yoshida et al., 2002). For example, higher sugars levels resulted in accelerated senescence in tobacco leaf discs (Wingler et al., 1998). Similarly, sugar levels and leaf senescence were found to be enhanced when plants were grown under higher CO_2 conditions (Miller et al., 1997). Enhanced leaf senescence was observed in plants in which *HEXOKINASE* (a sugar sensing receptor) was constitutively expressed. In addition to leaf senescence, higher levels of sugar also correlated with enhanced resistance (Herbers et al., 1996, Salzman et al., 1998). As shown by Yoshida et al., (2002), *hys1* (*HYPERSENESCENCE1*), one of the *cpr5* alleles, shows hypersensitivity to sugars in a HEXOKINASE expression dependant manner and also displays early leaf senescence. Thus, Yoshida et al., (2002) proposed CPR5 to be a repressor of disease resistance and leaf senescence. Moreover, they concluded that the sensitivity of the HEXOKINASE receptor has somehow been enhanced in *cpr5* and therefore, *cpr5* plants are hypersensitive to sugars. Similar results were also found by Anderson (Anderson, 2006) and Jing et al., (2007, 2008). Keeping in view the overlapping in signalling of resistance and senescence, it is very difficult to pinpoint the regulatory nodes required for hexokinase, CPR5 and sugars signalling, and their redundant roles in this said signalling.

As mentioned above, the levels of reactive oxygen species (ROS) are also generally found to be upregulated in senescent tissues. ROS act as signalling molecules for pathogen resistance and leaf senescence, whereas they also act as antimicrobial compounds during PCD and necrosis (Lamb and Dixon, 1997, Van Breusegem and Dat, 2006). ROS are generated during numerous biochemical processes, such as, photosynthesis, respiration, metabolism,

senescence, and hormone signalling (Tripathy and Oelmüller, 2012). The generation of ROS is essential for normal plant growth and development, however their relentless high levels in cells are deleterious to cells and sometimes result in cell death (Pallavi et al., 2012, Tripathy and Oelmüller, 2012). This kind of stress on the cell is known as oxidative stress (Jing and Dijkwel, 2008). Oxidative stress covers the production of superoxide, nitric oxide, triplet and singlet of oxygen, salicylic acid, and hydroxyl radicals (Jing et al., 2007, Hong-Ying et al., 2010). Despite their harmful effects, ROS act as stimulators for the activation of systemic acquired resistance (SAR), programmed cell death, cell cycle and responses to various abiotic and biotic stresses (Vanacker et al., 2006, Zimmermann et al., 2006, Pitzschke and Hirt, 2010, Möller et al., 2001). ROS levels in cells are regulated by antioxidant enzymes, such as, catalases, superoxide dismutases (SOD), peroxidases and glutathione-S-transferases (GST), and antioxidant compounds like ascorbic acid and glutathione (De Gara et al., 2003, Gill and Tuteja, 2010). Antioxidant activity is reduced and ROS levels are increased to execute senescence (Navabpour et al., 2003, Wu et al., 2012). Likewise, the majority of senescence inducing transcription factors are shown to be upregulated when plants are challenged with H_2O_2 (Miao et al., 2004, Balazadeh et al., 2010b, Balazadeh et al., 2011, Garapati et al., 2015).

As shown by the results (Jing et al., 2008), *cpr5* contain elevated levels of ROS. Their results showed that the majority of genes, which were found to be differentially expressed in *cpr5*, belonged to a stress-related class of genes and *SENESCENCE-ASSOCIATED GENES* (*SAGs*). Furthermore, transcriptomic studies confirmed that *cpr5* plants have fivefold higher transcripts of the ROS gene network compared to the wildtype (Jing et al., 2008). Thus, they concluded that early leaf senescence in *cpr5* plants is essentially a consequence of altered redox levels in *cpr5* plants (Jing et al., 2008). Their results were further reinforced when the expression of *SOD* genes was found to be slightly higher in *cpr5*. Based on the upregulation of ROS and *SAGs*, early leaf senescence and HR phenotypes could be the result of altered ROS levels in *cpr5* (Jing et al., 2008). In addition to transcriptomic analyses, the photo-oxidation stability of *cpr5* was also measured in terms of photosynthesis system stability in response to H_2O_2 treatment, i.e. how stable is the photosynthesis system of *cpr5* as compared to wildtype when plants were exposed to H_2O_2. The photosynthesis system II of *cpr5* was found to be resistant to H_2O_2 compared to the wildtype (Hong-Ying et al., 2010). The decrease in Fv/Fm ratios was found to be 25-fold lower in *cpr5* leaves than in the wildtype (Hong-Ying et al., 2010). Membrane leakage was also reported to be five-fold lower in *cpr5* leaves than in the wildtype.

An increase in the expression level and activities of antioxidants, such as, GSH, SOD, and APX, was also observed in *cpr5* compared to the wildtype (Aki et al., 2007, Hong-Ying et al., 2010). In summary, these results indicate an imbalance of ROS levels in *cpr5*, which could be the result of mutation in CPR5 or a consequence of disturbance(s) in other integral processes. Based on these results, it is tempting to propose a role of CPR5 in maintaining ROS balance in the cell, possibly by modulating antioxidant activity in order to supress leaf senescence or disease resistance.

Taken altogether, leaf senescence and defence responses both share many physiological events, such as, increased levels of SA, JA and ethylene (Ryals et al., 1996, Morris et al., 2000), the generation of ROS (Levine et al., 1994, Pastori and Del Rio, 1997) and the accumulation of *SENESCENCE-ASSOCIATED GENES* (Yoshida et al., 2001, Jing et al., 2007). Thus it is difficult to nominate which one of these processes results in the activation of the other processes, and which of the regulators, such as, SA, JA, ET, or ROS is produced first in response to mutation in *CPR5*.

1.4 Cell cycle, cell growth and development, and CPR5

Normal cell cycle progression is crucial for proper plant growth and development (De Veylder et al., 2011). The cell cycle chiefly uses the same cellular machinery for meiosis, mitosis and endoreduplication processes (Breuer et al., 2010, Chevalier et al., 2011, De Veylder et al., 2011). The cell cycle is generally divided into four stages; G1 (gap 1), S (DNA synthesis), G2 (gap 2) and M (mitosis) (Breuer et al., 2010; Chevalier et al., 2011; De Veylder et al., 2011). Cells grow in number by cell division and expand in size through endoreduplication (De Veylder et al., 2011). In endoreduplication, genome (DNA) of the cell replicates, however, the cell does not divide and re-enters into the cell cycle for the next round of replication, resulting in the duplication of the original genome (Chevalier et al., 2011; De Veylder et al., 2011). Cell cycle progression is integrally regulated and progressed through the activities of a number of cyclin-dependent kinases (CDK) and their respective cyclins. CDK inhibitors (CKIs), such as, KIP-RELATED PROTEINS (KRPs), SIAMESE *(SIM)* and SIAMESE-RELATED (SMR), also play a central role in the regulation and progression of cell cycle (Kasili et al., 2010). CDK activities are shown to be higher during G1/S and G2/M transitions, however, mitotic CDK activities are repressed during endoreduplication (Inagaki and Umeda, 2011).

Despite many studies, the switching from the mitotic to the endocycle is poorly understood. Mitotic CDK activity is believed to be reduced or repressed, and this repression of mitotic CDK activity is shown to be achieved through the activation of the anaphase-promoting complex/cyclosome (APC/C). Thus, APC/C is shown to be an important point of convergence for cell division and endoreduplication signalling switching (Peters, 2006, Marrocco et al., 2010, Inagaki and Umeda, 2011). APC, a multi-subunit E3 ubiquitin ligase, essentially controls cell cycle progression through the degradation of cell cycle proteins (Peters, 2006, Marrocco et al., 2010). *FIZZY-RELATED* (*FZR*), *CELL DIVISION CYCLE 20* (*CDC20*) and *CDH HOMOLOG 1* (*CDH1*) are shown to be positive activators of APC (Peters, 2006, Larson-Rabin et al., 2009, Marrocco et al., 2010). Five homologues of *CDC20* (*CDC20.1-CDC20.5*) and three homologues of *CDH1* (*CCS52A1, CCS52A2, and CCS52B*) are reported in *Arabidopsis*. Of three *CDH1* homologues, *CCS52A1* and *CCS52A2* both are implicated in endoreduplication (Lammens et al., 2008, Larson-Rabin et al., 2009, Vanstraelen et al., 2009, Kasili et al., 2010). On the other hand, a number of APC negative regulators such as, *OSD1* (*OMISSION OF SECOND DIVISION 1*), *GIG1* (*GIGAS cell 1*) and *UVI4* (*UV INSENSITIVE 4*) have also been found (Hase et al., 2006, Heyman et al., 2011, Iwata et al., 2011). In conclusion, the proper progression of cell division and endoreplication is of key importance for normal cell growth and development.

As shown, *cpr5* plants show smaller leaves, aberrant trichomes and reduced ploidy levels, all of which are dependent on cell cycle for their cell division and expansion through endoreduplication (Bowling et al., 1997, Kirik et al., 2001). These results further highlight that cell cycle (cell division or endoreduplication) appears to be disturbed in *cpr5*, since epidermal pavement cells of *cpr5* were found to be significantly (70 %) smaller than wildtype. A comparison of ploidy levels (DNA content) of *cpr5-2* and *CPR5* plants revealed that the number of cells having 16C DNA content is greatly (~90 %) reduced in *cpr5* compared to wildtype (Kirik et al., 2001). Similarly, trichomes on *cpr5* leaves were found to have either none, or contain one appendage and showed reduced ploidy levels (Brininstool et al., 2008). Since cell size and ploidy levels are positively correlated with endoreplication (Brininstool et al., 2008), smaller cells and reduced ploidy levels in *cpr5* show that the regulation of cell cycle especially the endocycle, appear to be distorted in *cpr5* plants. In contrast to the wildtype, the majority of *cpr5* leaf cells and trichomes contain 8C DNA content, while few cells contain 16C ploidy levels (Kirik et al., 2001). These results (the presence of 8C and 16C nuclei) indicate

that cell cycle or cells are able to enter into the endocycle, however, their entry is somehow restricted past the second round of replication, allowing only a few *cpr5* cells to enter in the third round of endoreduplication. How CPR5 exerts an effect on the restriction of the endocycle in *cpr5* is not shown yet, and is an interesting question to be answered. A number of studies focus on understanding the role or effect of CPR5 in the regulation of cell cycle, as discussed below.

Endoreduplication is found to be promoted in *osd1* and *uvi4* loss-of-function mutants when these genes are mutated individually. However, the lethality of female gamete is observed when mutated collectively in *osd1uvi4* double mutants (Kasili et al., 2010). Epistatic analyses were carried out between *osd1*, *uvi4* and *cpr5* in order to find out the role of CPR5 in endoreduplication. The results show that *cpr5* managed to supress the majority of *uvi4* ectopic phenotypes (enhanced endocycles) and lethality of *osd1uvi4* double mutants. The severity of aberration of trichome appendages and resistance observed in *cpr5* was found to be increased in *cpr5uvi4* (Bao and Hua, 2014). In addition, cell cycle (cyclin B) genes, *CYCB1;1*, *CYCB1;2* and *CYCB1;4*, were shown to be upregulated in *cpr5*, which indicates that mitotic activity was higher in *cpr5*. Based on their results, Bao and Hua, (2014) concluded that CPR5 interacts with OSD1 and UVI4 in a highly complex (unknown) network where CPR5 appears to affect cell cycle signalling downstream of *OSD1*, *UVI4* and *FZR* (*CCS52A*) genes.

GL1 ENHANCER BINDING PROTEIN (*GeBP*) or *GeBP*-like proteins (*GPLs*) are putative activators of the *GLABRA1* (*GL1*) gene, which is required for normal cuticle and trichome formation as well as plant defence responses (Xia et al., 2010). GeBL/GPLs overlap with CPR5 signalling and are found to activate a set of genes, which represents a subset of *CPR5* signalling pathway genes induced in *cpr5* (Perazza et al., 2011). This *CPR5* subset of genes is shown to be implicated in cell expansion and endoreplication in a *CPR5* dose-dependent manner (Perazza et al., 2011). Despite the aforementioned studies, little is known about the direct role of *CPR5* in the cell cycle, and specifically in endoreduplication regulation.

In 2014, the CPR5 was shown to physically interact with CKIs, such as, KRP2, SIM and SMR (Wang et al., 2014b). Additionally, it was proposed that CKIs are safeguarded by CPR5 in the nuclear membrane, which are released upon pathogen detection by NB-LLR and subsequently phosphorylate RB-E2F signalling in order to initiate cell cycle progression (Wang et al., 2014b). Recently, CPR5 has emerged as one of the nucleoporins of the nuclear pore complex

(NPC) and has been shown to restrict the entry of an array of nuclear cargoes (Gu et al., 2016). Gu and co-workers also proposed that CKIs are dissociated from CPR5, phosphorylate RB-E2F signalling and induce cell cycle progression upon pathogen recognition (Gu et al., 2016). In *cpr5*, the proper shutting and opening of the NPC is compromised, which in turn results in the deregulated entry of a large number of molecules in the nucleus that could lead to intrusions in a number of pathways (Gu et al., 2016). Aberrant phenotypes and disruptions in various pathways could be a consequence of loss of gating ability of CPR5. To summarize, these results show that cell cycle and endoreduplication is affected in *cpr5*, the exact mechanism of which is not understood yet.

1.5 Potassium homeostasis in *cpr5*

Potassium (K^+) is one of the essential macronutrients required for normal plant growth and development (Borghi et al., 2011). Potassium plays an important role in the regulation of enzyme activity, stomata movement, cell expansion and defence in plants (Ashley et al., 2006). For example, upon challenge, plant-pathogen interaction results in the activation of Ca^+, which in turn triggers anion channels, plasma membrane depolarization and the induction of K^+ permeable efflux channels for the export of K^+, and eventually initiates HR execution (Jabs et al., 1997, Jeworutzki et al., 2010). Jasmonic acid (JA) rapidly detects low K^+ levels (shortage) and promptly responds to low K^+ (Armengaud et al., 2004, Armengaud et al., 2009). Furthermore, ABA application results in K^+ reduction in the xylem during drought exposure (Shabala, 2007). As shown (Bowling et al., 1997; Jing et al., 2007), *cpr5* plants show HR-like spontaneous lesions, contain elevated JA levels and display hypersensitivity to exogenous JA and ABA application. It is possible that *CPR5* could have a direct role in K^+ homeostasis. The reduced growth, or cell expansion, and the basal resistance of *cpr5* could be the consequence of low K^+ in *cpr5* cells. Borghi et al., (2011) studied *cpr5* alleles and dissected the role of *CPR5* in potassium homeostasis in the cell. In order to check the role of JA or SA in K^+ misregulation, they carried out epistatic analyses and found that the reduced K^+ levels in *cpr5* leaves are independent of SA, JA and HR signalling. Additionally, they showed that *cpr5* (*cpr5-2* and *cpr5-3*) leaves have higher transcript abundance of *Cyclic Nucleotide Gated Channels* (*CNGC*), such as, *CNGC10*, *CNGC11*, *CNGC12*, *CNGC19* and *CNGC20* compared to the wildtype (Borghi et al., 2011). Since they did not find a direct link between CPR5 and K^+ homeostasis, they proposed that low K^+ in *cpr5* leaves could be a consequence of K^+ efflux, resulted from the constitutive activation of CNGCs. Essentially, the activation of CNGCs may

have enhanced K^+ efflux either indirectly, by increasing cytosolic Ca^{+2} levels, and/or directly, by increasing K^+ transport activity. As mentioned earlier, BA results in the reduction of K+ levels and *cpr5* plants display hypersensitivity to ABA (e.g. *cpr5* shows delayed germination and reduced hypocotyl length, (Jing et al., 2007, Gao et al., 2011), it could be possible that ABA hypersensitivity is the result of low K^+ levels or altered K^+ levels in *cpr5*. Likewise, enhanced defence responses, but reduced growth, could also be the result of altered K^+ levels in *cpr5*. In conclusion, this study was also unable to draw a definite role of CPR5 in K^+ homeostasis.

1.6 Germination and seedling growth response in *cpr5*

Phytohormones are vital for many plant growth and developmental, as well as defence-related stages Dekkers and Bentsink, 2015; (Kucera et al., 2007, McSteen and Zhao, 2008, del Carmen Rodríguez-Gacio et al., 2009, Dekkers and Bentsink, 2015). Any defect in their biosynthesis or signalling leads to detrimental effects (Dekkers and Bentsink, 2015). *cpr5* seedlings were found to be hypersensitive to JA and ET when applied exogenously e.g. *cpr5* seedlings were less green and shorter in length compared to wildtype (Jing et al., 2007). These results were found to be in line with Gao et al., (2011), wherein they studied *cpr5* response to the ABA hormone and found similar effects on seed germination and seedling growth. They further poroposed that *CPR5* controls normal germination and development through the lipoxygenases (LOX) signalling pathway (Gao et al., 2011).

The seedling growth and germination of both wildtype Col-0 and *cpr5* plants was shown to be inhibited at similar levels in response to ABA when applied exogenously. However, *cpr5* exhibited a lower germination rate, reduced cotyledons greening and shorter root lengths at 1% ABA concentrations as compared to the wildtype plants, which displayed the aforementioned phenotypes at 4% of ABA (Gao et al., 2011). In contrast to *cpr5* and the wildtype, *CPR5* overexpression plants were found to be insensitive to higher ABA application (Gao et al., 2011), indicating that *CPR5* may act as a repressor of ABA signalling. Furthermore, the inhibition of germination and seedling growth of *cpr5* in response to NDGA, an ABA biosynthesis inhibitor (Gao et al., 2011), indicates that *CPR5* may be involved in ABA signalling rather than ABA biosynthesis.

Sugar and ABA signalling pathways are reported to crosstalk with each other in regards to plant germination and dormancy (Finkelstein et al., 2008). ABA and sugar are reported to have

antagonistic effects on each other (Dekkers et al., 2008). Physiological studies show that higher sugar concentrations result in the inhibition of seed germination and seedling development (Dekkers et al., 2008). In contrast to ABA, *cpr5* hypersensitivity to sugar was abolished in the *cpr5hxk5* mutant, however, *cpr5hxk5* still displayed hypersensitivity to ABA, JA, ET, and SA (Aki et al., 2007; Yoshida et al., 2002). This observation indicates that the hypersensitivity of *cpr5* plants to sugar is not due to increased sugar levels but rather is due to the deregulation of sugar sensing and signalling (Aki *et al.*, 2007; Yoshida *et al.*, 2002). These results suggest that *CPR5* is involved in the regulation of the sugar signalling response and acts as a negative regulator of sugar-induced inhibition of plant development through the suppression of the glucose dependent HXK signalling pathway. In conclusion, these studies propose CPR5 to be an important component in coordinating signalling pathways required to regulate hormonal homeostasis, yet the exact mechanism of regulation and signalling is poorly understood.

1.7 Plant growth and resistance trade-off and *CPR5*

Unlike animals, plants are sessile and lack a highly specialized memory system, thus they have developed an array of constitutive, as well as induced, defence responses to protect themselves from pathogens (Chisholm et al., 2006). Growth and pathogen resistance are two opposing mechanisms e.g. the activation of defence responses has a significant compromise on growth. Therefore, there is a trade-off between growth and immunity in order to reduce the cost of fitness (Yang et al., 2011). This trade-off is believed to happen largely due to the limited resources that a plant needs to invest either in growth or defence for its survival (Coley et al., 1985, Simms and Rausher, 1987, Herms and Mattson, 1992). The activation of defence-related responses is costly since the expression of resistance that genes exert, negatively effects plant fitness (Simms and Rausher, 1987, Purrington, 2000). The negative correlation between growth rate and defence has been well established and is a well-known phenomenon (Smedegaard-Petersen and Tolstrup, 1985, Fagerstrom and Schneider, 1989, Hoffland et al., 1995). Arabidopsis transgenic lines constitutively overexpressing SAR (systemic acquired resistance) were found stunted in growth and lower in seed yield than wildtype plants (Bowling et al., 1994, Greenberg et al., 2000, Jirage et al., 2001). Similarly, the exogenous application of JA or MeJA on plants in order to increase resistance to herbivores, dramatically reduced seed production (Baldwin, 1998, Redman et al., 2001), leaf growth (Alves et al., 2008) and delayed flowering and fruiting (Redman et al., 2001). Nevertheless, investments in defence regulation always pay off in terms of plant survival, when plants are challenged to pathogens (Baldwin,

1998, Walters and Heil, 2007). Thus, during adaptation and evolution, plants have managed to develop some sophisticated mechanisms to achieve a balance between growth and defence, and reduce fitness costs (Huot et al., 2014). The balance between growth and plant immunity is mainly achieved through the interplay of plant hormones, called phytohormones (Denancé et al., 2013, Yang et al., 2013, Yang et al., 2012).

Phytohormones are intrinsically integrated by complex networks in plants to respond to developmental and environmental cues, hence limiting the defence-related fitness costs. The molecular mechanisms underlying these hormonal networks are largely obscure. During an attack, pathogens target the biosynthesis and/or signalling of plant hormones in order to manipulate and evade the plant defence system (Denancé et al., 2013). Pathogen recognition by plant receptors triggers the biosynthesis of different phytohormones, such as, salicylic acid (SA), ethylene (ET), and jasmonic acid (JA). The roles of salicylic acid (SA), jamonic acid (JA) and ethylene (ET) in the regulation of resistance to biotrophic and necrotrophic pathogens have been extensively studied and are well established (Berrocal-Lobo et al., 2002, Erb et al., 2012, Pieterse et al., 2012, Wasternack and Hause, 2013, Huot et al., 2014). It has been shown that these phytohormones, in addition to the accomplishment of resistance, exert adverse effects on plant growth and development, such as; JA represses growth, flower development and fertility (Wasternack, 2013); SA effects growth and flowering-time regulation (Martínez et al., 2004); and ethylene misregulates fruit ripening, senescence and root hair growth (Abeles et al., 2012). As mentioned above in Section 1.7, SA and JA are upregulated in *cpr5*, and ABA, ET, sugar sensing and their signalling is also distorted in *cpr5*. Therefore, reduced growth phenotypes of *cpr5* could be the consequence of imbalance of phytohormones. In addition, other plant hormones, such as, brassinosteroids, auxins, cytokinins and gibberellins, which were previously known as plant growth promoting hormones, have recently emerged as the plant immunity regulators (Denancé et al., 2013).

Brassinosteriods are of central importance in the regulation of plant growth and development, which are essential throughout plant life, e.g., germination, vegetative growth, etiolation, stomata development, flowering, and fertility (Clouse, 2011, Zhou et al., 2013). A number of studies have proposed their role in modulating the trade-off between plant growth and immunity (Navarro et al., 2008, Albrecht et al., 2012, Belkhadir et al., 2012, Yang et al., 2012, Lozano-Durán et al., 2013, Chandran et al., 2014, Fan et al., 2014, Malinovsky et al., 2014). These studies suggest a unidirectional negative correlation between brassinosteriods (BR) and

plant immunity, for example, plant immunity was compromised when brassinosteriods signalling was activated, either by the exogenous BR application or by the overexpressing BR receptor *BRI1* (*BRASSINOSTEROID INSENSITIVE1*) and/or the BR biosynthetic enzyme *DWF4* (Belkhadir et al., 2012, Sun et al., 2013a, Chandran et al., 2014). Contrastingly, immunity is not always compromised in response to the application of BRs. For instance, rice and tobacco (Nicotiana tabacum) plants conferred resistance when BRs were applied exogenously (Nakashita et al., 2003).

Brassinosteriods (BRs) and PAMP-triggered immunity (PTI), e.g. the activation of the defence response upon *flg22* recognition by the *FLS2* receptor, act through the same signalling cascade, in which many of the signalling components or receptors, such as, BAK1, BSK1, and BIK1 are shared between BRs and PTI (Sun et al., 2013a, Sun et al., 2013b, Lozano-Durán and Zipfel, 2015). Among these co-receptors, BAK1 is the most important receptor, since BR-(BRI1 receptor) and PAMP-triggered signalling (FLS2 receptor) receptors, both compete for the BAK1 co-receptor for downward signalling and transduction (Chinchilla et al., 2007, Belkhadir et al., 2012, Wang, 2012). However, Albrecht *et al.*, (2012) show that BRI1 and FLS2 do not compete for BAK1, and that BAK1 is not a rate-limiting factor for either of these pathways. Similar results were obtained by several studies (Shan et al., 2008, Gimenez-Ibanez et al., 2009, Ranf et al., 2011, Lozano-Durán et al., 2013). The idea of BAK1 competition appears to be weak since the receptors and co-receptors are assembled in separate membrane compartments before binding with cognitive ligands (Bücherl et al., 2013). Thus, the crosstalk between BRs and PTI could be downstream of BAK1 or BIN2 signalling components (Bucherl et al., 2013; Lozano-Duran et al., 2013).

The crosstalk between BRs and PTI could be happening at a transcriptional level, since PTI is compromised when BZR1 becomes activated (Lozano-Duran et al., 2013). Additionally, the activation of WRKY transcription factors (*WRKY 11, 15, 18*) could be part of BRs and PTI crosstalk (Lozano-Duran et al., 2013). *WRKY 11, 15, 18* are a subset of genes that are also regulated through BZR1 and supress defence signalling (Lozano-Duran et al., 2013). BZR1 and BES1 are the two critical transcription factors of the BRI1-mediated BR signalling pathway. In the absence of BRs, BZR1 and BES1 are repressed by BIN2 through phosphorylation and remain in the cytoplasm (Wang et al., 2002, Vert and Chory, 2006, Gampala et al., 2007, Ryu et al., 2007). Upon BR treatment, BZR1 and BES1 both get dephosphorylated, translocate to the nucleus and extensively orchestrate BR-induced

transcriptional reprogramming (Sun et al., 2010b, Tang et al., 2011, Yu et al., 2011). Consequently, this transcriptional reprogramming activates *WRKY* TFs, such as, *WRKY11*, *15* and *18*, which promote growth by repressing defence responsive genes. In addition to BZR1-mediated BR signalling, BR-responsive genes can also be activated through another transcription factor, HBI1, which also represses defence responsive genes. On the other hand, pathogen (*flg22*) recognition by the *FLS2* receptor leads to the activation of defence responsive genes and inhibits growth by repressing BR-responsive growth promoting genes as well as *HBI1* TF. Concomitantly, the activation of *WRKY11*, *15*, *18* transcription factors supress or control defence responses through a negative feedback loop in order to limit fitness cost (Lozano-Durán and Zipfel, 2015). Based on these results, the activation and regulation of *BZR1*, *HBI1* and *WRKY* TFs (*WRKY11*, *15*, *18*) appear to be of central importance for growth and immunity trade-off. Having established that resistance and BR-mediated growth share the same signalling cascade and components, it is plausible that compromised growth phenotypes of *cpr5* could be a consequence of suppression of growth by the activation of resistance. However, how CPR5 coordinates or participate in BRs and PTI crosstalk is totally obscure. Assuming that CPR5 inhibits resistance via SA suppression or modulation, it appears that SA-mediated resistance in *cpr5* could repress BR-mediated growth and consequently result in enhanced resistance and reduced growth. However, there are no reports on the role of BRs in *cpr5*-mediated resistance and growth, and this needs further investigation and evidence to support my hypothesis.

Gibberellins and their cognitive DELLAs, both play a central role in the regulation of plant growth prompting genes (*PREs*) via BZR1-mediated BR-signalling. For example, DELLAs repress BZR1, which in turn, supress PREs (Claeys et al., 2014, Lozano-Durán and Zipfel, 2015). Additionally, DELLAs also interact with *PHYTOCHROME INTERACTING FACTORs* (*PIF3* and *PIF4*) and inactivate them by interfering with their DNA-binding domain (Achard et al., 2007, de Lucas et al., 2008, Feng et al., 2008). Active or stabilized *PIFs* activate *PRE* genes, which promote growth and repress immune responses. Contrarily, inactive PIFs lead to repression of growth (proliferation) and the activation of immune response genes. In a light-grown condition, PREs are regulated by BRs through *BZR1*, whereas in the dark, *BZR1* interact with *PIFs*, and the *BZR1-PIFs* complex then activates downstream *PREs* genes. *PIFs* transcription factors (*PIF3* and *PF4*) are centrally important in the regulation of balance between growth and immunity (Achard et al., 2007; de Lucas et al., 2008; Feng et al., 2008).

Alternatively, in the absence of GAs, DELLAs regulate plant defence responses through the modulation of salicylic acid (SA) and jasmonic acid (JA) levels in the cell (Navarro et al., 2008). In GA deficient mutants, *ga1-3*, one of the DELLAs mutants displayed an overlapping of 30% transcripts with plants treated with *flg22*. The majority of these (30%) upregulated genes represent defence-related WRKY transcription factors, including WRKY11 (Eulgem et al., 2000, Navarro et al., 2008, Journot-Catalino et al., 2006). WRKY11 is shown to be upregulated in senescent leaves in *Arabidopsis* (Eulgem et al., 2000). These results suggest that GAs and DELLAs are key players in the management of trade-off between plant growth and immunity.

Similarly, several lines of evidence show that auxin and BRs relationships are also very important with respect to their role in the trade-off between growth and immunity (Cohen and Meudt, 1983, Tanaka et al., 2005, Bajguz and Hayat, 2009). For example, studies (Vert and Chory, 2006, Zhao et al., 2002) proposed that dephosphorylated BES1 and BZR1 (two important transcription factors of BR signalling) show a higher binding affinity to the promoters of *CDP* and *SMALL AUXIN UP RNA 1* (*SAUR-AC1*). BRs regulate the hypocotyl length of *Arabidopsis* through auxin signalling components, such as, IAA19 and ARF7 (Zhou et al., 2013). BRs and auxin act synergistically and promote growth via the degradation of ARF2 (auxin) by BIN2 in BR signalling pathways (Cohen and Meudt, 1983, Tanaka et al., 2005, Fridman and Savaldi-Goldstein, 2013). These results indicate that plant growth can be restricted by modulating auxin regulation through BR signalling.

Alternatively, many plant pathogens have the ability to alter auxin levels by themselves or regulate auxin production within the plant by manipulating host growth and defence pathways (Manulis et al., 1998, Glickmann et al., 1998, Vandeputte et al., 2005, Robert-Seilaniantz et al., 2007, Spaepen and Vanderleyden, 2011). For instance, auxin responsive genes were found to be repressed when plants were exposed to SA analogue benzothiodiazole (BTH) or bacterial elicitor *flg22* (Navarro et al., 2006, Wang et al., 2007). SA regulates auxin signalling repression through the stabilization of AUX/IAA repressor proteins and the inhibition of the auxin response (Wang et al., 2007, Fu and Wang, 2011). Similarly, plants were shown to be susceptible to *PstDC3000* when the auxin response was enhanced through the overexpression of the TIR1 auxin receptor in contrast to plants, which displayed increased resistance to the same bacteria when auxin signalling was attenuated through the overexpression of *miR393*

(microRNA393; Navarro et al., 2006). Thus, the misregulation of auxin signalling appears to be one of the bacterial pathogen strategies to evade host defence.

Similar to BRs, GAs and IAAs, cytokinins are also important plant growth hormones that promote cell division, leaf longevity, nutrient mobilization and grain yield (Ashikari et al., 2005, Kim et al., 2006a, Sakakibara, 2006, Choi and Hwang, 2007, Matsumoto-Kitano et al., 2008, Hwang and Sakakibara, 2006, Zhao et al., 2010b, Kurakawa et al., 2007). Cytokinins confer drought resistance in tobacco (Rivero et al., 2007). Cytokinins and auxin both play crucial roles in gall formation by biotrophic bacterial (*Rhodococus facians* and *Agrobacterium tumefaciens*) as well as fungal (*Puccinia striiformis*) pathogens (Robert-Seilaniantz et al., 2007, Walters and Heil, 2007, Pertry et al., 2010, Lee et al., 2009). Similarly, a number of plant pathogens are known to produce cytokinins or cytokinin analogues, which enable them to manipulate the plant host system to activate cytokinins biosynthesis. For example, the formation of green islands and gall formation by biotrophic bacterial as well as fungal pathogens (Robert-Seilaniantz et al., 2007, Walters and Heil, 2007, Pertry et al., 2010, Lee et al., 2009).

A large number of studies propose that cytokinin-mediated growth activation results in the suppression of plant basal defence responses (Robert-Seilaniantz et al., 2007). However, Choi and co-workers (Choi et al., 2010) demonstrated that basal defence suppression, actually depends on the source of origin of cytokinin production. Host-derived cytokinins, in fact, enhance the plant defence system via the activation of SA signalling (Choi et al., 2011). The degradation of RIN4 by bacterial effectors (*AvrRpm1* and *AvrRpt2*) and subsequent recognition of RIN4 degradation by resistance proteins (RPM1 and RPS2) activate the expression of cytokinin-responsive primary genes in addition to the induction of the defence system (Igari et al., 2008). RPM1 and RPS2 proteins are shown to activate cytokinin, as well as SA-mediated responsive genes - *ARABIDOPSIS RESPONSE REGULATOR 5* (*ARR5*) and *PR1* respectively, possibly through the induction of cytokinin biosynthesis. In cytokinin signalling, cytokinin receptors, AHK2 and AHK3 phosphorylate type-B ARR transcription factor (ARR2), result in the induction of type-A ARR transcription factors. When challenged by pathogens, the TGA3 transcription factor that is activated by SA, recruits ARR2 that binds in the *PR1* promoter and upregulates *PR1* levels (Choi et al., 2011). Additionally, cytokinins are also known to regulate nitric oxide (NO), which leads to the execution of the hypersensitive response (HR) when *flg22* is recognised by FLS receptor (Choi et al., 2011). Thus, pathogen-derived cytokinins confer

susceptibility to plants and help gall formation, whereas, cytokinins derived from within plants boost plant immunity.

Based on the above-discussed results, the majority of hormones use the same signalling cascade for their actions in order to promote growth. In the signalling cascade, the activation and regulation of BZR1, HBI1 and WRKY TFs (WRKY11, 15, 18) appear to be of central importance for the growth and immunity trade-off. Having established that resistance and BR-mediated growth share the same signalling cascade and components, it is plausible that compromised growth phenotypes of *cpr5* could be a consequence of suppression of growth by the activation of resistance through the same pathway. However, how CPR5 coordinates or participates in BRs and PTI crosstalk is obscure. Assuming that CPR5 inhibits resistance via SA suppression or modulation, it appears that SA-mediated resistance in *cpr5* could repress BR-mediated growth and consequently result in enhanced resistance and reduced growth. However, there are no reports on the direct implication or role of *cpr5* in BRs -mediated resistance and growth, and further investigation and evidence is needed to support my hypothesis.

Taken altogether, the abovementioned review of growth and immunity trade-off provides an overview of the different possible strategies, which plants have evolved to keep an optimal balance between growth and development. As shown, this trade-off is mainly controlled and regulated through the modulation of biosynthesis and/or signalling of plant hormones(Lozano-Durán and Zipfel, 2015). Moreover, only a few of the hormones have been quantified and studied in relation to CPR5. Despite detailed analyses of CPR5 with SA, JA, ABA and ET, the relationship between these hormones and CPR5 is yet poorly understood. As evident from many studies, CPR5 is shown to take part in a number of physiological processes and signalling pathways (Jing and Dijkwel, 2008). Most importantly, growth and development is compromised, whereas, resistance is gained as a consequence of mutation in *cpr5*. This observation conspicuously indicates an example of growth and immunity trade-off happening in *cpr5*, which needs further detailed studies.

1.8 Outlook and Aims of the Research

CPR5 appears to be a master regulator as proposed by Jing and Paul (Jing and Dijkwel, 2008), since a mutation in *CPR5* results in defects in multiple biological processes and pathways. Despite evidence of *CPR5* implications in numerous processes and pathways, the basic

mechanism of how *CPR5* coordinates or participates in the proposed pathways is yet poorly understood. The function of a protein can be determined from the structure of a protein. It could be possible that CPR5 has a unique structure that helps CPR5 to participate in several pathways or interact with multiple partners. CPR5 displays no homology to any of the proteins with a known function. Thus, the first aim of the current study was to solve the atomic structure of CPR5. In addition to structure, CPR5 protein could have a number of important structural motifs, sites or regions which could facilitate CPR5 to perform multiple functions. Therefore, extensive *in silico* analyses of CPR5 followed by the mutagenic analyses of the identified structural motifs, sites or regions will help to understand pleiotropic effects exerted by CPR5.

CPR5 is shown to be a nuclear membrane protein by different research groups. However, the role of the nuclear localisation signal (NLS) in the transportation of CPR5 in the nucleus has never been characterised and is a long sought-after question. The mutagenic and functional characterisation of putative NLS clusters will show the implication of NLS clusters in CPR5 localization. Thus, the functional characterisation of the 25 transgenic lines of the *CPR5* gene in the current study (Chapter 6) would prove a great resource of information and would highlight different novel roles of *CPR5*.

CPR5 is important for cell growth, cell cycle and trichome development (Kirik et al., 2001); on the other hand, it is shown to negatively regulate resistance, leaf senescence and HR-like lesions, which are all types of programmed cell death. These results suggest that CPR5 is an important candidate implicated in the regulation of growth and resistance trade-off, which needs further investigations. Thus, mutagenic and functional characterizations of developed transgenic lines are expected to uncouple the role of CPR5 in the regulation resistance and growth and understand growth and resistance antagonism.

Chapter 2 Materials and Methods

2.1 *In silico* computer tools used for CPR5 analyses

The nucleotide and deduced amino acid sequence of CPR5 (AT5G64930.1) were downloaded from "The Arabidopsis Information Resource, TAIR" (http://www.arabidopsis.org/) which were used in the current thesis and to design all the primers and the computer-based analyses. CPR5-like proteins (homologues) were identified from the NCBI database using an online tool, Basic Local Alignment Search Tool (BLAST). Using default settings, CPR5 deduced amino acid sequence was inputted into BLAST query search interface, which provided a list of proteins which were found to have amino acid identity with Arabidopsis CPR5. The amino acid sequences of all of these CPR5-like proteins (Appendix 11.2) were downloaded from the NCBI database and subsequently were again aligned and compared using CLUSTALW2 (http://www.ebi .ac.uk/Tools/msa/clustalw2/) in order to find out the level of identity between CPR5 and CPR5-like proteins.

The hydrophobic regions (transmembrane domains) were analysed and selected by TOPCONS(http://topcons.net/) (Bernsel et al., 2009), TMHMM (http:// www.cbs.dtu.dk/ services/TMHMM/) (Möller et al., 2001), SMART (http://smart. embl-heidelberg.de/) (Letunic et al., 2009) and PSIPRED (http://bioinf.cs.ucl.ac.uk/ psipred/) (Nugent and Jones, 2009). The CPR5 nuclear localisation signal (NLS) was predicted using SeqNLS (http://mleg. cse.sc.edu/seqNLS) (Lin and Hu, 2013) and PSORTII (http://psort.hgc.jp/ form.html).

The cytoplasmic domain of the CPR5 was examined using "The Eukaryotic Linear Motif Resource" (ELM; http://elm.eu.org/) and PHYRE2 (Kelley and Sternberg, 2009) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) for the annotation of casein kinase I (CKI) and casein kinase II (CKII) plus other structural motifs. Additionally, the COILS web interface (http://www.ch.embnet. org/software /COILS_form. html) was used to find out the presence of potential coiled-coil motifs in the CPR5 protein. CPR5 subcellular localisation was annotated using SoftBerry (http://linux1.softberry.com/berry.phtml?topic= protcomp&group=help&subgroup=proloc); PredSL (http://aias.biol.uoa.gr/PredSL/); Cello (http://cello.life.nctu.edu.tw/); Protein Prowler (http://bioinf.scmb.uq.edu.au:8080/pprowler_ webapp_1-2/).

In a similar fashion, the presence of intrinsically disordered regions (IDRs) in the CPR5 protein was investigated using the DisProt (http://www.disprot.org/) (Sickmeier et al., 2007), IUPred (http://iupred.enzim.hu/) (Dosztányi et al., 2005), MetPrDOS (http://prdos.hgc. jp/cgi-bin/meta/top.cgi) (Ishida and Kinoshita, 2008) and DisMeta (Huang et al., 2014) web interfaces, whilst the occurrence of alpha-helices and a beta-strand was annotated using PHYRE2, PSIPRED and I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy et al., 2010). The proper position and orientation of reading frames of all the designed constructs were checked and confirmed using the *in silico* cloning option of Geneious R6 (Biomatters Limited) software, prior to the start of the *in vitro* experiments. Lastly, the CPR5 gene or protein sequence was analysed according to the default settings of each *in silico* program used in the current thesis. A number of *in silico* tools were used in order to measure or assess the same parameter of the CPR5 gene or protein, only to gain a better idea or higher probability of the said parameter.

2.2 Primer designing, RNA extractions and cDNA synthesis

2.2.1 Primer designing

All the primer sets which were used for the amplification of the 8 constructs to be used to solve the CPR5 protein structure were designed through PerlPrimer, (http://perlprimer. sourceforge.net/). Restriction sites and a stop codon were added in the primers where needed and mentioned (Appendix 11.1.1, 11.1.2, 11.1.3). However, all the primers used for real-time (qRT-PCR) quantifications were designed using the online tool, QuantPrime (http://www. webcitation.org/query.php?url=http://www.quantprime.de/&refdoi=10.1186/1471-2105-9-465) (Arvidsson et al., 2008). For transcript abundance quantification of *CPR5*, *cpr5-2* and transgenic versions (*SynCPR5*), a separate set of primers were designed for each plant line (Appendix 11.1.4). As shown by Drenkard et al., (2000), a single base pair mismatch within the last three nucleotides (3` end) of a primer make the identification of two alleles possible and easier. Since the *cpr5-2* allele of the *CPR5* gene has a G-to-A mutation, different sets of primers were designed using SNAPER (developed by Ausbel Lab; http://ausubellab.mgh. harvard.edu/) in order to select sets of primers highly specific for *cpr5-2* and *CPR5*. All the sets of primers (designed by SNAPER) were applied and one of the best pairs of primers (which displayed no amplification from the non-specific allele) were selected (Appendix 11.1.5). These pairs of primers were found to be highly specific, such as, the *CPR5* (wildtype) based

primer set, which was unable to amplify a product from *cpr5-2* (mutant) or *SynCPR5* and vice versa. In order to quantify the levels of transcripts, a set of primers for each plant line (*CPR5*, *cpr5-2* and *SynCPR5*) was designed in the same region of At*CPR5* mRNA (Figure 2.1). The nucleotide changes in the *SynCPR5* gene introduced during codon optimization, provided a highly specific set of primers for *SynCPR5* transcript abundance quantification.

2.2.1.1 Primer sets for genotyping

Homozygous T5 transgenic lines of each construct were genotyped for the presence of respective transgene cassettes in their genomes. For constructs in which a long stretch of nucleotides (*Del37CPR5* and *Del63CPR5*) or a few nucleotides (*DelckICPR5*) were deleted, intergenic types of primers were designed (Drenkard et al., 2000). However, nucleotide mismatches were selected and included in the 3` end of the primers used for the genotyping of constructs in which amino acids of CPR5 protein were changed by substitution of nucleotides. It is important to note that only the forward (sense) primer contained mismatches and was unique for each construct whereas the same reverse (antisense) primer was used for all of the genotyping studies (Appendix 11.1.5).

2.2.2 RNA Extractions and cDNA synthesis

Total RNA was extracted from 4-5 week old leaf tissue or from the whole plants (for real-time qRT-PCR analyses) using ZR plant RNA miniprep® (Zymo Research) according to the manufacturer's instructions. The total RNA was treated with DNase I (Roche Applied Sciences) prior to being used as a template for complementary DNA strand (cDNA) synthesis. The first-strand cDNA pool was made from 2 µg of total RNA through reverse transcription using One-Step Blueprint reverse transcriptase (Takara), *CPR5* specific primers and dNTPs (Section 2.2.4.1). This One-Step BluePrint PCR reaction also converted cDNA into *CPR5* double-stranded DNA using single-stranded DNA as a template (Section 2.2.4.2). Thus *CPR5* DNA obtained from this reaction was used for cloning in the pGEM-T easy vector. For real-time quantifications, cDNA was synthesized using Superscript III Reverse transcriptase (Life Technologies-Invitrogen). The cDNA synthesized from the 2 µg of total RNA was further 20-fold diluted and used as a template for quantitative real-time PCR (qRT-PCR) using a LightCycler® 480 SYBR Green 1 Master reaction mix (Roche Life Science) and LightCycler® 480 instrument II (Roche Life Science, USA).

2.2.3 qRT-PCR data analyses

Transcript abundance of each selected gene was quantified using the LightCycler 480 Real-Time PCR (Roche) system and LinReg PCR analysis software. For every biological replicate, three technical replicates were included and analysed, using SYBR green I (Roche) as a florescent dye. A volume of 2.5 µL of 20-times diluted cDNA was used as a template (Section 2.2.4.3). The raw data files resulted by LightCycler 480 qPCR were extracted using LC480 conversion-software, freely available online at http://www.hartfaalcentrum.nl/. Similarly, the primer efficiency of each primer set was calculated using another computer tool, LinReg PCR software (Ruijter et al., 2009). It is important to mention that none of the amplicons were sequenced. Transcript abundance and statistical analyses were performed using Microsoft Office Excel 2010 using the Pfaffl equation as described in their studies (Pfaffl et al., 2004). The expression value (gene of interest e.g. *CPR5*) from each transgenic line was normalized with the average value (geometric means of their C_T values) of the two housekeeping genes, *Ubiquitin Conjugating Enzyme 9* (*AtUBC9; AT4G27960*) and *Actin2* (*AtACT2; AT3G18780*). For housekeeping gene selection, 5 genes were studied and the aforementioned set of genes were selected based on their stable expression patterns in all of the plant lines studied in the current study.

2.2.4 PCR reactions and profiles

2.2.4.1 Amplification of single and double-stranded full length *CPR5*

i.	OneStep BluePrint RT enzyme	2.0 µL
ii.	2x OneStep BluePrint buffer	25.0 µL
iii.	Forward primer (10 µM)	1.0 µL
iv.	Reverse primer (10 µM)	1.0 µL
v.	RNase free ddH_2O	20.0 µL
vi.	RNA	1.0
vii.	Total volume made	**50.0 µL**

PCR profile

a.	cDNA:	50.0˚C	30.0 min	
		94.0˚C	2.0 min	
b.	DNA:	98.0˚C	10.0 sec	
		54.5˚C	30.0 sec	cycles = 30
		72.0˚C	2.15 min	

2.2.4.2 Amplification of *CPR5* constructs for structural studies

a)	DNA template	1.0 µL
b)	Forward primer (10 µM)	1.0 µL
c)	Reverse primer (10 µM)	1.0 µL
d)	Master Mix (Promega)	12.5µL
e)	ddH_2O	9.5 µL
f)	Total volume made	**25.0 µL**

A construct-specific set of primers were used for the amplification of each construct.

PCR profile

i.	94.0˚C	2.0 min	
ii.	94.0˚C	10.0 sec	cycles = 30
iii.	54.0˚C	30.0 sec	
iv.	72.0˚C	2.15 min	
v.	72.0˚C	10.0 min	

2.2.4.3 cDNA synthesis for real-time PCR quantifications

a)	cDNA template	2.5 µL
b)	Forward primer (10 µM)	1.0 µL
c)	Reverse primer (10 µM)	1.0 µL
d)	SYBER green I (Roche)	5.0 µL
e)	ddH_2O	0.5 µL
f)	Total volume made	**10.0 µL**

Real-time PCR profile

i.	96.0˚ C	2.0 min	
ii.	96.0˚ C	10.0 sec	cycles = 40
iii.	60.0˚ C	10.0 sec	
iv.	72.0˚ C	10.0 sec	
v.	96.0˚ C	5.0 min (Melting curve)	
vi.	10.0˚ C	Cooling	

2.3 Cloning strategy and development of *CPR5* constructs for structural studies

In order to make the purification of the *CPR5* truncated versions (Figure 2.2) easier, each of the *CPR5* designed constructs (Figure 5.1) was fused with 6 x His-tag either at the N-terminus or C-terminus of the construct (Figure 2.2C and 2.2D). For this purpose, a modified version of

the pET32 vector (Figure 2.3) was kindly provided by Dr Gillian Norris (Associate Professor, IFS, Massey University, PN, New Zealand). This modified version of pET32 contains two copies of 6 x His-tag as shown in Figures 2.2 and 2.3 and allows the cloning and expression (fusion) of the gene of interest with the N-terminal, C-terminal His-tag or without any tag (Figures 2.2 and 2.3) depending on the cloning strategy. For example, pET32 vector was digested with NdeI and BamHI restriction enzyme and ligated with *CPR5* fragment (construct), which was already digested with the same set of enzymes in order create CPR5 recombinant protein without His-tag (Figure 2.2B). Similarly, pET32 vector was digested with BamHI to create protein with an N-terminal His-tag and with NdeI and XhoI in order to develop protein with C-terminal His-tag (Figure 2.2). A stop codon (TAA) along with the selected restriction sites was incorporated through primers in some of the constructs in order to avoid the fusion of the protein of interest with His-tag (Figures 2.2B and 2.2C).

2.4 Amplification, transformation and confirmation of 8 *CPR5* transgenes

All of the 8 *CPR5* constructs (Figure 5.1) were amplified using construct specific primer sets (Appendix 11.1). The amplified *CPR5* truncated fragments were ligated (Section 2.5.3) into the pGEM-T-easy vector system (Promega). After 2 hours incubation at room temperature, the ligation mix was transformed (Section 2.6) into *E. coli* (DH5α) competent cells (Section 2.8). Using blue-white screening under ampicillin selection pressure, positive transformants (white colonies) were selected (Section 2.6). Plasmids were isolated (Section 2.6) and the nucleotide sequences of *CPR5* fragments (B-H; Figure 5.1) amplified from the isolated pGEM-T easy vector were confirmed by the Sanger sequencing method in order to make sure that no nucleotide addition, deletion or replacement occurred during amplification and cloning. Following sequencing, each of the constructs was then excised from the pGEM-T-easy vector and ligated into the pET32 vector as mentioned in Section 2.3 and Figures 2.2 and 2.3.

Following successful transformation, the transformants were grown under ampicillin selection, in order to choose the right transfromants. Subsequent to selection, plasmids from positive transformants were isolated (Section 2.7) in order to verify their integration in the desired place. The proper integration and orientation of transgenes was confirmed via restriction analysis using the EcoRI restriction enzyme. The presence of the EcoRI site in the transgenes as well as in the pET32 vector allowed the restriction analysis and subsequently confirmed the proper integration and orientation of the transgene cassette of each individual version in the pET32

vector in the right order. Additionally, the proper placement of each transgene was also validated through a colony PCR using a T7 promoter-specific (sense) and construct-specific (antisense) set of primers.

After the successful transformation of all the selected constructs in the final expression vector (pET32a), the bacterial plasmids carrying the transgene cassettes were individually transformed (Section 2.6) into the bacterial strain (*Rosetta gami*) in order to express CPR5 protein variants. Following transformation, the bacteria carrying transgene plasmids were grown under four types of selection pressures i.e. ampicillin, tetracycline, kanamycin and chloramphenicol (Section 2.17). These cells were then induced with IPTG (1.0 mM) when the optical density (measured by a spectrophotometer) of the cells reached between 0.5-0.6. After induction, the cells showed a maximum protein yield after 3-4 hours of incubation at 37˚ C. A decline in protein yield was observed when samples were incubated for a longer period of time e.g. 5 or 6 hours, or overnight. Recombinant proteins were extracted, analysed for solubility and processed for purification and crystals.

2.5 Cloning techniques and protocols

2.5.1 DNA restriction/digestions

Restriction enzymes used for DNA digestions carried out in the study were purchased either from Roche (Roche Diagnostics GmbH, Manheim, Germany) or New England Biolabs (MA, USA). Concentration of DNA to be digested was estimated using Nanodrop 1000 spectrometer (Thermo Fisher Scientific, Wilmington DE. USA) and subsequently appropriate amount of buffer(s), DNA, and appropriate restriction enzyme(s) were added in the reaction mix. Sterile water (ddH_2O) was added into reaction mixture in order to make the required final volume. Digestion reactions were incubated at 37 C^0 for 2 hours or overnight. Recommended or compatible buffers were used for double or sequential digestions.

2.5.2 DNA elution/extractions from agarose gels

2.5.2.1 Preparation of DNA extraction/elution tubes

DNA elution/extraction columns or tubes were manually prepared in the lab as explained here. A hole was made in the bottom of a 0.5 mL micro-centrifuge tube (sterilized) by pricking with a needle, pointing the back of the tube towards myself. 10-15 threads of glass wool were placed

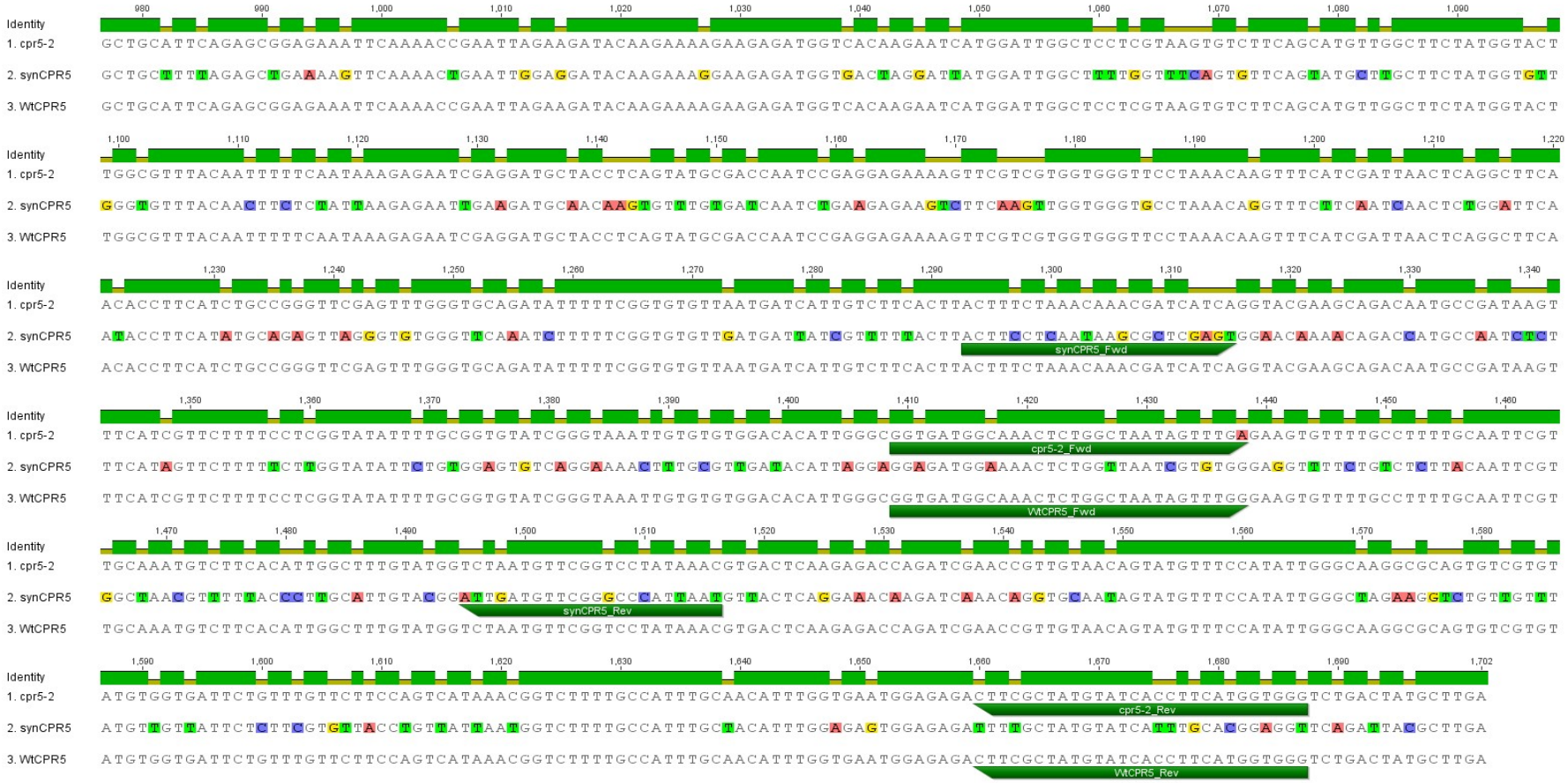


Figure 2.1 Positions of primers used for real-time qRT-PCR quantifications

This Figure represents the positions of the primer sets (sense and antisense) used for the quantification of transcript abundance of *CPR5* (wildtype), *cpr5-2* (mutant) and *SynCPR5* (synthetic) genes, using Geneious. Figure was constructed by aligning nucleotide sequences of *CPR5* (wildtype), *cpr5-2* (mutant) and *SynCPR5* (synthetic) genes; only 3` ends of the genes are depicted in this figure.

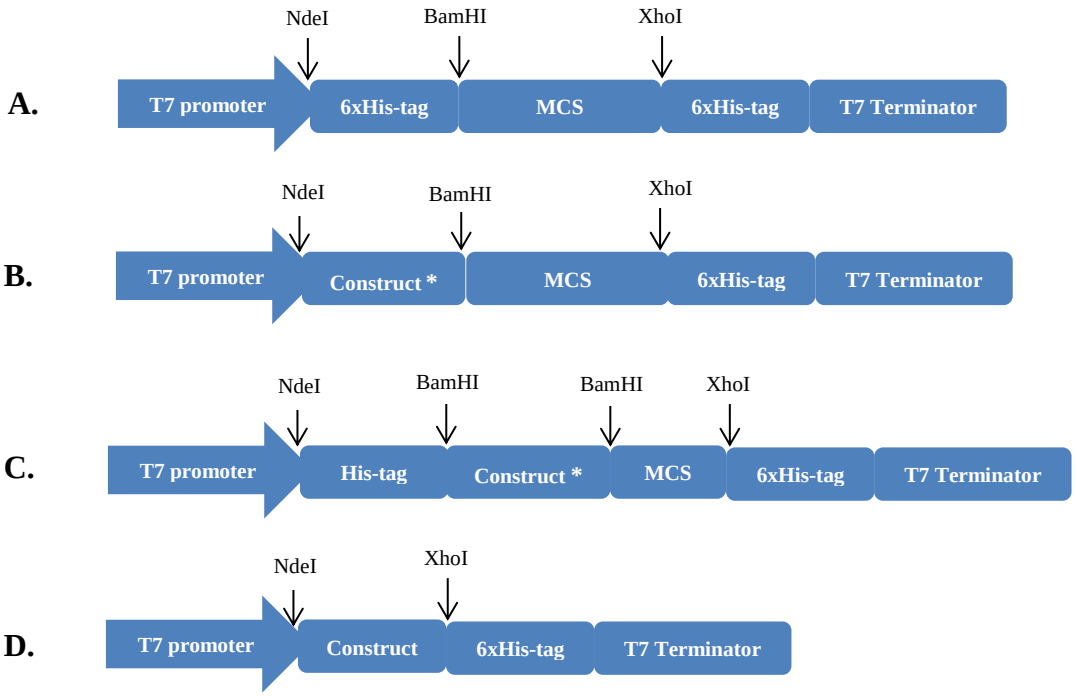


Figure 2.2 Schematic representation of transgene cassette of pET32

This figure represents the positions of the T7 promotor, His-tag(s), multiple cloning sites (MCS), and T7 terminator. Figure **2.1A** displays the original version of pET32. However, **2.1B**, **2.1C** and **2.1D** show the cloning strategy of integration of the construct without His-tag, with N- and C-terminal His-tag, respectively. The asterisk (*) indicates that the said construct contains a stop codon at the end of this construct.

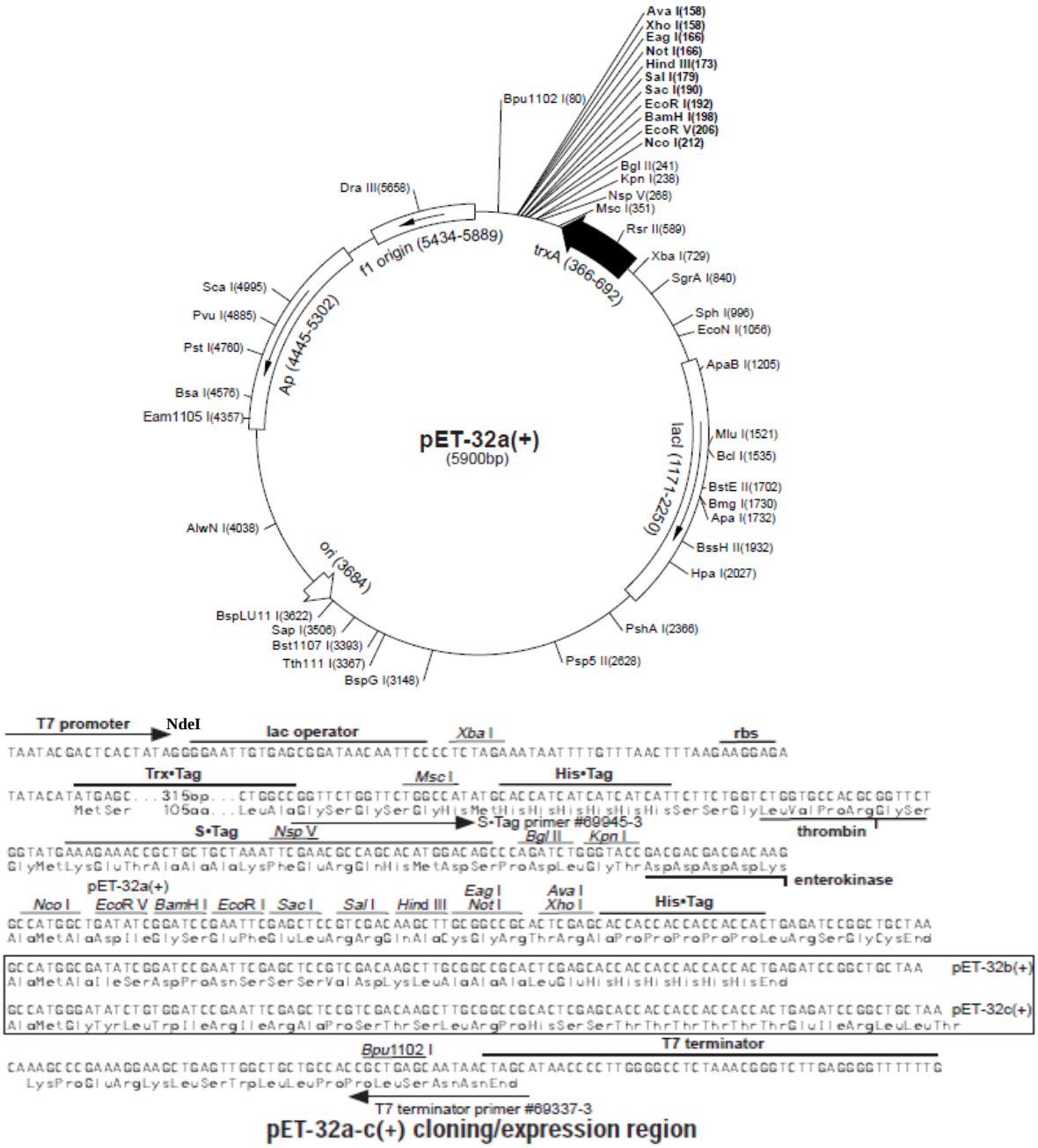


Figure 2.3 pET32 sequence composition and landmark information

This figure shows the pET32 vector map, positions of His-tags and restriction sites which allow the cloning of the *CPR5* constructs to be expressed with His-tag at N- and C-terminal of CPR5 recombinant proteins and CPR5 recombinant CPR5 protein without His-tag. A copy of pET32 vector was kindly given by A/Prof Gillian Norris, IFS, Massey University, New Zealand.

at the mouth of hole with the help of sterilized forceps. This 0.5 mL tube containing wool was then placed inside the 1.7 mL Eppendorf tube. Extraction/elution tubes were prepared under aseptic conditions.

2.5.2.2 DNA elution procedure

The selected DNA fragment was marked and subsequently excised from the 1 % agarose gel using a razor blade or scalpel. It was then transferred to a 0.5 mL tube and placed in a 1.7mL micro-centrifuge tube as prepared above in Section 2.5.2.1. The extraction/elution containing slices of the DNA fragment were spun at 6000 rpm for 15 minutes in a bench top centrifuge tube and DNA was collected in the 1.7 mL tube. Liquid collected in the tube (pass-through supposed to be containing DNA) was measured and transferred into a new Eppendorf tube. The isolated DNA was further purified using ethanol precipitation as described below;

100 % Ethanol	3x of collected liquid
3 M NaOAc	$1/10^{th}$ of collected liquid

and this mixture was incubated at -20°C for at least one hour or overnight. The mixture was spun at full speed for 32 minutes and the pellet was washed with 70 % ethanol two times. Finally, the pellet was air-dried and re-suspended in 10-20 µL of sterile distilled water. When needed, DNA fragments were also eluted from agarose gels using commercially available elution kits (Zymoclean Gel DNA Recovery kit, Zymo Research Corporation, Irvine CA. USA) according to the manufacturer's instructions.

2.5.3 DNA ligation

In order to ligate the DNA fragment of interest (the insert), firstly the concentration of the insert was measured using the Nanodrop 1000 spectrometer and subsequently the following formula;

ng of insert = ng of vector x Kb size of insert x 3:1 ratio/ Kb size of vector

To set up a ligation reaction, ligation regents were thawed, briefly centrifuged and added in the following order.

i.	Insert DNA	3 times of vector molar concentration
ii.	Vector (pGEM)	1.0 µL (50 ng/µL)
iii.	2x buffer	5.0 µL
iv.	T4 DNA Ligase	1.0 µL
v.	ddH_2O	to make up to a total volume of 10 µL

Appropriate amounts of the insert and vector were added and incubated for either 1 hour at room temperature or overnight at 4° C, if a maximum number of transformants was required.

2.6 Bacterial transformation protocol

100 µL of ligation reaction (as prepared in Section 2.5.3) was added in to competent cells (as prepared in Section 2.8), mixed the cells by gently flicking the tube, and incubated on ice for 20-30 minutes. Subsequently, cells were given heat-shock for 30–60 seconds in a water bath set at exactly 42°C and the tubes were then immediately returned to ice for 2 minutes. 900 µL LB medium was added to the tube(s) containing the cells transformed with ligation reactions and these were then incubated for 1.5 hours at 37°C with shaking (~220 rpm). In the meantime, Agar plates (Section 2.6.1) were equilibrated to room temperature. After 1.5 hours incubation, 100 µL of transformation mixture were spread onto LB/ampicillin/IPTG/X-Gal plates (Section 2.6.2). These plates were further incubated overnight (16–24 hours) at 37°C. Bacteria carrying the transgene were able to grow on the agar plates in the form of colonies. For *Agrobacterium*-mediated transformations, the plasmids (pGreen0229) containing the transgenes were transformed into *Agrobacterium* electro-competent strain (*GV3101*) by electric-shock-mediated transformation, and were then subsequently spread on the plate containing the selective media. The plates were incubated for about 40 hours for *Agrobacterium* in order to develop colonies.

2.6.1 Preparation of LB liquid and agar media

	LB Liquid	LB Agar
Peptone	10 g/L	10 g/L
Yeast-extract	5 g/L	5 g/L
Sodium chloride	10 g/L	10 g/L
Agar	---	15 g/L

The said amounts were dissolved in 1 litre of RO water and pH was adjusted to 6.5 and media were autoclaved. Appropriate antibiotic(s) were added before use (Section 2.22). LB agar media were cooled down to 50^0 C and antibiotic(s) were added prior to the preparation of LB agar plates.

2.6.2 LB agar plates for IPTG/X-Gal for white/blue Selection

Similarly, IPTG/X-Gal stocks were prepared and mixed as follows;

- X-gal (0.12M) 40 µL
- IPTG (0.1M) 4 µL
- LB 40 µL

The above-made mix was then spread onto the LB agar plates prepared in Section 2.6.1, and subsequently the plates were left to dry in the laminar flow hood.

2.7 Bacterial plasmid (DNA) isolation protocol (Alkaline lysis)

Single bacterial colonies grown on agar plates (Section 2.6) were picked and added into 3 mL of Luria Broth (LB; Section 2.6.1) along with appropriate antibiotic(s). After 14-16 hours of incubation at 37°C, cells were harvested by spinning at 13,000 x rpm at room temperature using the method developed by Sambrook et al., (2001). The cell pellet was resuspended in 200 µL ice-cold Solution I (50 mM glucose, 25 mM Tris pH 8.0, 10 mM EDTA, 4 mg/ mL lysozyme). To lyse the cells, 300 µL of Solution II (200 mM NaOH, 1% SDS) was added and mixed gently by inversion ~20 times and was incubated for 5 minutes at room temperature. 300 µL of Solution III (3.0 M Sodium Acetate, pH 4.8) was added to neutralize the solution. Samples were centrifuged at 13,000 x rpm for 15 minutes in order to remove cell debris, and supernatants were transferred to fresh tube(s). Optionally, proteins were removed by adding 400 µL of chloroform:isoamyl alcohol (24:1). The top aqueous layer was transferred to the fresh tube(s) after spinning the tubes at 13,000 x rpm. Isopropanol (700 µL; 0.7 of the volume of the transferred liquid) was added to precipitate plasmid DNA, and the precipitated DNA was pelleted by centrifugation at 13,000 x rpm for 15 minutes. The pellet was washed twice with 70 % ethanol and centrifuged at 13,000 x rpm for 2 minutes in order to remove residual isopropanol. Finally, the pellet was air-dried and dissolved in sterile distilled water. Plasmid isolations from *Agrobacterium GV3101* cells were performed by following the method described by Wise and co-workers (Wise et al., 2006). The purity and quality (260/280 and 230/260 nm ratios) of the plasmid DNA was checked by the Nanodrop ND-1000 spectrophotometer.

2.8 Competent cells preparations for *E. coli* and *Agrobacterium*-mediated transformations

E. coli (DH5α strain) cells were streaked from the glycerol stock, which was stored at -80°C, and the ultra-chemical competent cells were prepared following the protocol developed by Inoue Lab (Inoue et al., 1990). These ultra-competent cells were transformed with ligations (Section 2.5.3) by heating at 37°C for 15 minutes. After heat shock, the transformed cells were allowed to grow in 1 mL of Luria Broth (LB) (Section 2.6.1) for 1.5 hours. Then 100 µL of the 1.0 mL transformation volume was uniformly spread on the plate containing the selective media and appropriate antibiotics (Ampicillin; 100 mg/ mL). For *Agrobacterium*-mediated transformations, *Agrobacterium* strain (*GV3101*) electro-competent cells were prepared as described in (Weigel and Glazebrook, 2006) studies.

2.9 Purification of CPR5 protein variants

2.9.1 Sample preparation

Bacteria containing transgenic plasmids were grown in 5 mL LB (Section 2.6.1) containing appropriate antibiotic(s) and incubated overnight at 37°C at 200 x rpm. Having achieved an optical density of 0.5 – 0.6, 2.5 mL, this culture was added into two flasks, each containing 100 mL of LB and appropriate antibiotic(s), and then incubated on 37°C at 200 x rpm. After saving 1.0 mL from each flask as the un-induced control, the rest of the culture was induced by IPTG (1 mM) and incubated at 37°C and 200 x rpm. After 4 hrs, 30 mL from each flask (the induced culture) was taken out and the rest was kept shaking overnight (O/N) under the same conditions as mentioned above, in order to compare protein expression and stability. Cells from each sample (un-induced, induced for 4 hours and overnight) were collected in the form of pellets by spinning at 4000 x rpm at 4°C for 10 minutes and then frozen at -80^{0}C. Pellets of cells from each sample were thawed and re-suspended in 3.0 mL and 1.5 mL of lysis buffer (Section 2.9.5.1). 250 uL from both of these samples (3.0 mL and 1.5 mL tubes) were divided into 1.7 mL tubes and then the cells were lysed by sonication (20%, 100J). The lysed cells were spun at 13,200 rpm for 1 minute, and a supernatant containing the soluble protein was transferred into new tubes, whereas the left-over cell pellet was dissolved in 8.0 M Urea (Section 2.9.5.5). The expression level of each construct in the soluble fraction of protein was checked and

compared to the insoluble fraction of the protein on 12% SDS-PAGE gel (Section 2.9.7) before the purification of the protein through His-column or ion-exchange chromatography.

2.9.2 Resin (sepharose beads) preparation

150 µL of chelating Sepharose (GE healthcare, Life Sciences) was added into a 1.7 mL tube and spun for 7-10 sec. The resin was washed with 70% ethanol (three times) as well as with double distilled water (ddH2O) five times, and the supernatant was discarded each time in order to remove the residual ethanol from the resin. 75 µL of the washed resin was added into the tube containing Ni^+ metal ions (or Co^+, Fe^+, Zn^+, Mg^+; Section 2.9.5.6) and was subsequently washed with double distilled water (ddH2O) five times. Resin was equilibrated with lysis buffer (the same buffer with which the cells were lysed) five times.

2.9.3 Purification using His-tag

Cell lysates (supernatants and pellets prepared in Section 2.9.1) were mixed into the tubes containing sepharose beads (resin) pre-charged with selective metal ions (Section 2.9.5.6) and the mixture was incubated for 1.5 hrs at 4°C and continuously swirled. After 1.5 hours of shaking, the tubes were spun for 30 seconds at full speed and the supernatants were transferred to fresh tubes. Resin containing proteins was washed with 1 mL of elution buffer containing low levels of imidazole (Section 2.9.5.3) and the supernatants were saved. To avoid any protein degradation, protein inhibitors were added to all the supernatants obtained as a result of each washing and kept on ice during the purification process. The resin was washed with elution buffer containing 100 mM of imidazole in order to remove non-specific (bacterial) proteins weakly bounded onto the resins, and the supernatants were transferred to new tubes. In order to elute CPR5 His-tagged proteins, the resin was washed with elution buffers (Section 2.9.5.6) containing 0.5 M, 1.0 M and 1.5 M imidazole in a gradient manner, and all the supernatants were transferred to fresh tubes. All the supernatants isolated from un-induced and empty vector samples were migrated onto SD-PAGE gels at 200 V for 55 minutes. The gels were subsequently stained in Coomassie blue staining for 45 minutes. Gels were de-stained in de-staining buffer and imaged by scanning.

2.9.4 Purification using ion-exchange chromatography

In order to purify the protein using ion-exchange chromatography (without His-tag), the isoelectric point (pI) of each construct (Untagged version) was calculated using ProtParam

(http://expasy.org/). Since most of the constructs displayed an isoelectric value of 6.92-7.58, both anion and cation exchange chromatography were used in order to purify different untagged versions. Based on the pI value of each construct, two types of lysis, loading and elution buffers (Section 2.9.6), such as, sodium acetate (pH 5.0) for anion chromatography and Tris-HCl (pH 8.0) for cation chromatography, were prepared. Samples were prepared and processed for purification in a similar fashion as prepared for the His-tagged version in Section 2.9.3.

2.9.5 List and recipes of buffers used for Ni^{+}-column based purifications

2.9.5.1 Lysis buffer

Chemical	250mL	500mL
50mM K_2HPO_4 (pH=7.5)	2.1773g	4.3545 g
0.25M NaCl	6.13g	12.26 g
10% (v/v) glycerol	25mL	50 mL
0.5 mM TCEP	35.827mg	71.655 mg
10mM Imidazole	170.2mg	340.4 mg

The final pH was adjusted between 7.5 8.0 using KCl. The buffer was filtered and stored without autoclaving.

2.9.5.2 Loading buffer

50 mM K_2HPO_4 (pH=7.5)	2.1773 g	4.3545 g
0.5 M NaCl	12.26 g	24.52 g
0.5 mM TCEP	35.827 mg	71.655 mg
20 mM Imidazole	340.4 mg	680.8 mg

The final pH was adjusted between 7.5 8.0 using KCl. The buffer was filtered and stored without autoclaving.

2.9.5.3 Elution buffer

Chemical	250mL	500mL
50 mM K_2HPO_4 (pH=7.5)	2.1773 g	4.3545 g
0.5 M NaCl	12.26 g	24.52 g
0.5 mM TCEP	35.827 mg	71.655 mg
500 mM Imidazole	8.51 g	17.02 g

The final pH should be 7.5 or 8.0. The buffer was filtered and stored without autoclaving.

2.9.5.4 10x PBS buffer

KH_2PO_4	0.6 g	1.2 g
NaCl	20.0 g	40.0g
KCl	0.5 g	1.0 g
$Na_2PO_4.7H_2O$ / Na_2PO_4	6.7 g/3.6 g	13.4 g/7.2 g

All of the above mentioned chemicals were dissolved in ddH_2O and stored after sterilizing.

Notes

1. K_2HPO_4 is a strong base so adjust pH first with HCl by diluting it to 100 folds
2. Imidazole is light sensitive, so keep out of light by wrapping in aluminium foil
3. Don't forget to wash the filter with ddH_2O after filtering each buffer
4. The pH of all buffers should be the same i.e. 7.5 or 8.0

2.9.5.5 Preparation of 8 M urea

8 M Urea	14.41 g	15 mL of distilled water

Distilled water was added to make up the final volume to 30 mL. It was not filtered or sterilized and was stored at room temperature.

2.9.5.6 Preparation of metal ions used in chromatography

Ion		**Amount (mg)**	**ddH_2O (mL)**
Ni^+	**($NiCl_2$)**	200	1.0
Co^+	**($CoCl_2$)**	200	1.0
Fe^+	**($FeCl_3$)**	200	1.0
Zn^+	**($ZnCl_2$)**	200	1.0
Mg^+	**($MgCl_2$)**	200	1.0

Selected ions were completely dissolved and filtered.

2.9.6 List and Recipes of buffers used for ion-exchange purifications

A. 10 mM Tris-HCl (pH= 8.0 with HCl); dissolved 20 mL of 50 mM Tris-HCl buffer (which also contained 0.05 mM TCEP) in 80 mL ddH_2O
B. 10 mM Tris-HCl + 1M NaCl (pH= 8.0); dissolved 20 mL of 50 mM Tris-HCl buffer in 80 mL ddH_2O in addition to 5.844 g of NaCl
C. 10 mM NaOAc (pH= 5.0 with acetic acid)
D. 10 mM NaOAc + 1 M NaCl (pH= 5.0 with acetic acid)

2.9.7 Preparation of 12% acrylamide gels

2.9.7.1 Preparation of resolving or separating gel

Monomers	5mL	10mL	15ml	20ml
ddH_2O	2.18 mL	4.345 mL	6.53 mL	8.69 mL
1.5 M Tris-HCl	1.25 mL	2.5 mL	3.75 mL	5.0 mL
20 % SDS	25 µL	50.0 µL	75.0 µL	100.0 µL
40 % Acrylamide	1.5 mL	3.0 mL	4.5 mL	6.0 mL
10 % APS	50.0 µL	100.0 µL	150.0 µL	200.0 µL
TEMED	2.5 µL	5.0 µL	7.5 µL	10.0 µL

2.9.7.2 Preparation of stacking gel

Monomers	5mL	10mL	15ml	20ml
ddH_2O	3.12 mL	6.24 mL	9.36 mL	12.48 mL
1.5 M Tris-HCl	1.25 mL	2.5 mL	3.75 mL	5.0 mL
20 % SDS	25.0 µL	50.0 µL	75.0 µL	100.0 µL
40 % Acrylamide	0.55 mL	1.1 mL	1.65 mL	2.2 mL
10 % APS	50.0 µL	100.0 µL	150.0 µL	200.0 µL
TEMED	5.0 µL	10.0 µL	15.0 µL	20.0 µL

WARNING:

- De-gas the mixture after adding 10% APS
- Immediately cast gel after adding TEMED
- Un-polymerized acrylamide is extremely toxic
- Wear gloves and never pipette by mouth

2.10 In-gel protein digestion for mass-spectrometry

2.10.1 Excision of the protein bands of interest from Gel

In order to identify CPR5, CPR5 putative protein bands were excised from the gel with a sterile scalpel blade inside the laminar flow hood under aseptic conditions. These slices were chopped into fine pieces with the scalpel blade, transferred to new LoBind Eppendorf tubes and then processed for Mass-spec identification as described below.

2.10.1.1 De-staining

a. 200 µL of de-staining solution (Section 2.10.2.1) was added into a tube containing sliced gel pieces and was incubated at 37° C for 1 hr.
b. If the gel pieces were found to be not properly de-stained, the de-staining step was repeated for another hour by adding fresh de-staining solution (200 µL).

2.10.1.2 Reduction

200 µL of reduction solution (Section 2.10.2.2) was added and incubated at 56° C for 45 minutes after discarding the de-staining buffer completely.

2.10.1.3 Alkylation

a) 200 µL of alkylating solution (Section 2.10.2.4) was added and mixed well by vortex for 30 min.
b) Mixture was washed with 50 mM NH_4HCO_3 and the supernatant was discarded
c) 150 µL of ACN was added and incubated for 10 minutes
d) ACN was removed and the gel pieces were dried under a vacuum (SpeedVac) for 5 minutes.

2.10.1.4 Tryptic digestion

a) 78 µL of digestion buffer (10 % ACN in 50mM NH_4HCO_3) was added in tubes containing dry gels.
b) Samples were vortexed for 5 mins and then incubated for 16-18 hrs at 37°C

2.10.1.5 Extraction of protein from gel pieces

a) After 16 hrs of incubation, the gel slurry of each sample was centrifuged and the supernatant was transferred to fresh clean tube(s) (Eppendorf).
b) 100 µL of 10 % ACN dissolved in 50 mM $NH4HCO_3$ was added to the gel slurry, vortexed for 10 minutes and then the supernatant was pooled after each wash.
c) Pooled supernatant was centrifuged for 10 minutes by adding 50 µL of ACN/1 % formic acid (ESI-MS) solution to the gel slurry.
d) Finally, the gel slurry was vortexed vigorously after adding 80% ACN and all the gel extracts were pooled again.
e) Samples were concentrated in the speedVac until nearly dry, taking care to ensure that samples were not completely dried out.
f) Samples were submitted for mass-spec (MS/MS) analyses and protein identification.

2.10.2 Solutions and suffers used for In-gel protein digestion

2.10.2.1 De-staining solution

a. 200 mM NH_4HCO_3 0.158 g/10 mL
b. 50 % Acetonitrile (ACN) 5 mL/10 mL

0.158 g of NH_4HCO_3 was dissolved in 5 mL of ACN, mixed well and then double distilled water (ddH_2O) was added to make up the final volume (10 mL).

2.10.2.2 Reduction buffer

i. 50 mM TCEP 0.1433 g/10 mL
ii. 100 mM NH_4HCO_3 0.079 g/10 mL

2.10.2.3 Washing buffer

c. 100 mM NH_4HCO_3 0.079 g/10 mL
d. 50 mM NH_4HCO_3 0.0395 g/10 mL

2.10.2.4 Alkylating solution

e. 360 mM Acrylamide 0.256 g/10 mL
f. 100 mM NH_4HCO_3 0.079 g/10 mL

0.256 g of acrylamide was added in 100 mM NH_4HCO_3 solution.

2.11 General sowing and plant growth conditions

Seeds were washed and disinfected according to the method described by Weigel and Glazebrook (2002), followed by vernalisation in the dark for 3-4 days at 4˚C. The vernalized seeds of the wildtype, mutants and transgenic lines were grown on autoclaved and pre-cooled Daltons Premium Seed Raising Mix®. Unless mentioned, the plants were grown at 22˚C and 65% Relative Humidity (R.H) under long day conditions, such as, 16 hours of day light (280 µE) and 8 hour dark periods. For *Pseudomonas syringae* bacterial infiltrations, plants were grown under short periods of light (10 hours) and longer periods of dark (14 hours) at 21˚C and 65% relative humidity (R.H.).

2.12 Trichome counting and leaf area measurements

Leaves were removed from the plants grown in soil. The whole leaf or circular discs of 0.68 cm diameter were cut from the same place of every selected leaf and then were mounted on the slide. The typical pattern of trichomes and their appendages was investigated under dissecting microscope, and trichome numbers with each type of appendage were recorded. Three to five

leaves were investigated for each transgenic line and significance of the differences was confirmed by performing a t-test.

To measure the leaf size or area, all the leaves including the cotyledons were pulled out, and two leaves were fed to a leaf area machine in the form of pairs, such as, first and second rosette leaves. The leaf area machine, scans, measures and displays leaf size on a separate displaying unit. The areas of leaves from about 10 plants of the same transgenic line were measured. Average leaf areas were calculated by taking the mean values of the same leaves from 10 different plants of the same transgenic line and applying a student t-test in order to show significance of differences between two or more transgenic lines.

2.13 *CPR5* gene synthesis, site-directed mutagenesis and transformations

2.13.1 Gene Synthesis

Based on the *in silico* results, the sequence of *CPR5* CDS and each construct, along with detailed information of positions for selected truncations or nucleotide substitutions, was provided to GenScript, (GenScript®, USA). GeneScript synthesised the *CPR5* gene (termed as *SynCPR5*) and introduced mutations or deletions as per provided instructions. The *CPR5* gene along with *CPR5* native promoter (1.5 Kb upstream of the *CPR5* start codon position) was synthesized and cloned into pGreen0229 by GenScript. All of the constructs were prepared and transformed in the pGreen0229 plasmid. All of the prepared *CPR5* transgenic constructs were sequenced by GeneScript in order to verify the introduction and integration of introduced mutation(s), insertion(s) and deletions(s) in the desired place, and this sequencing information was also sent along with samples. Following the confirmation of integration of each transgene in the proper position and orientation in the pGreeb0229 plasmid, GenScript provided 5 µg of pGreen0229 DNA containing the transgene. Upon reception, all the transgene cassettes were transformed into *E. coli*, and pGreen0229 plasmids were isolated. The presence of transgenes was verified by digestion analyses based on the introduced restriction sites in the *SynCPR5* gene. Additionally, a few of these constructs (transgene cassettes) were re-sequenced and verified in the lab prior to their transformation into the *Agrobacterium GV3101* strain as well. Thirdly, genotyping the set of primers (Appendix 11.1.5) was also used for the verification of introduction of changes in each construct by PCR amplifications.

2.13.2 *SynCPR5* transgenes transformation into *Agrobacterium GV3101*

In order to functionally characterise and confirm how many of the designed *CPR5* transgenes could complement *cpr5-2* compromised phenotypes, all the transgenes including the *CPR5* synthetic gene, were transformed into *cpr5-2* mutant plants using the floral dip method (Zhang et al., 2006). For *Agrobacterium* transformations, all the *CPR5* transgenes in pGreen0229 provided by GenScript, were transformed into the *Agrobacterium* strain *GV3101*. Following successful pGreen0229 transformations in *GV3101*, *Agrobacterium* (*GV3101*) containing transgenes was grown and multiplied in LB (500 mL; Section 2.6.1) supplemented with kanamycin (50 µg/mL), and was incubated for ~40 hours at 28°C with vigorous shaking (200 rpm). After about 40 hours, the cells were harvested and pelleted by spinning at 5000 rpm in a pre-chilled centrifuge (4°C). Cell pellets were resuspended in 200 mL of 10 mM $MgCl_2$, and 70 µL of 100% Silwett L77 (Sigma-Aldrich, USA) was also added as detergent.

2.13.3 *Agrobacterium*-mediated transformations of *cpr5-2* plants

For floral dips, the plants were grown in soil for 4 weeks in long day conditions (light: dark:16:8 hrs) and were transformed when plants had 4-5 inflorescences. The inflorescences were dipped into the bacterial suspension for 30 seconds. After dipping, the pots containing plants were covered with bags, tilted on their sides and kept in the dark for 24 hours at room temperature in order to maintain higher humidity and a favourable environment for bacterial infiltration. In the following week, the same plants were transformed again in order to increase transformation efficiency. Transformed plants were let to senesce naturally and seeds were recovered from transgenic plants when siliques were completely dried out. These seeds were named as T0 seeds, which were surface sterilized prior to sowing. About 3 or 4 day old transgenic seedlings were sprayed with BASTA (125 mg/L) in order to select plants carrying *CPR5* transgenes. BASTA spraying was repeated 3-4 times in order to exclude false positives. Successfully transformed plants were further studied and followed in the second generation for their segregations, and only plants which showed 1:3 (dead: survived) ratios were termed as homozygous transgenic lines and were selected for third generation. These homozygous transgenic lines of each construct were grown for two more generations and subsequently were analysed for physiological analyses.

2.13.4 Confirmation of positive transformants (genotyping)

Homozygous lines (T5 plants) of each construct were investigated for the confirmation of presence of transgenes in these plants. For this purpose, construct specific primers were used for the amplification of each transgene. *CPR5*, *cpr5-2* and transgene (*SynCPR5*) specific sets of primers made the isolation of the wildtype, *cpr5-2* and *SynCPR5* plants from each other, easier.

2.14 *Pseudomonas syringae* infiltrations

Pathogen infection assays were conducted with *Pseudomonas syringae* pv. tomato *DC3000* strain. LB medium, supplemented with rifampicin (Section 2.17), was used to grow bacteria (*PstDC3000*), for 24 hours at 30°C. After incubation, the cells were harvested and diluted in 20 mL of sterile 10 mM $MgCl_2$ in order to initially obtain an optical density (OD) of 0.2. The concentration of bacteria (5×10^6 bacteria per mL; OD = 0.001) was achieved by a serial dilution in 10 mM $MgCl_2$ solution as shown below.

$$Vol\ (to\ be\ required) = \frac{OD\ (desired) - Vol(desired)}{OD\ (current)}$$

e. g. a small scoop of cells was harvested from the plate and dissolved in 20 mL of $MgCl_2$ so the current OD was calculated as below;

OD (current) = 0.407 dissolved in 20 mL of 10 mM $MgCl_2$ = 0.407 x 20 = 8.14

Dilution 1

Desired OD = 0.2 Desired volume = 1 mL

$$Vol = \frac{0.2\ (desired) - 1\ mL\ (desired)}{8.14\ (current)} = 0.0245\ mL$$

Thus, 0.245 mL (24.57 µL) was taken from the initial cell solution (20 mL $MgCl_2$ solution containing cells with an OD of 8.14) and mixed into 1 mL of 10 mM $MgCl_2$.

Dilution 2

$$Vol = \frac{0.02\ (desired) - 3\ mL\ (desired)}{0.2\ (current)} = 0.3\ mL$$

Dilution 3

$$Vol = \frac{0.001\ (desired) - 50\ mL\ (desired)}{0.02\ (current)} = \ 2.5\ mL$$

A final 50 mL of 10 mM $MgCl_2$ was used to infiltrate the plant leaves. Plants were grown for 4 weeks in conditions of 12 hours light and 12 hours dark. The abaxial sides of the leaves of the wildtype as well as of the CPR5 transgenic plants were infiltrated with *PstDC3000* bacteria using a needleless syringe (1 mL). On average, a set of 8-10 plants (4-5 leaves per plant) were infiltrated with bacterial culture (*PstDC3000*) for each line.

The bacterial infections (*PstDC3000* growth) were quantified at 0, 3 and/or 4 days post infiltrations. Leaf discs (0.68 cm in diameter) were excised from three plants of the same line and pooled. Three biological replicates of each sample were taken, and the leaf discs pooled from 3 leaves of the same line were ground and homogenised in 500 µL of sterile 10 mM $MgCl_2$. The homogenised solution was then diluted 10 times [20 µL of solution was diluted in 180 µL of 10 mM $MgCl_2$] and serial dilutions were prepared as shown in Figure 2.4. In order to quantify bacterial growth from each transgenic line, 10 µL from each dilution was spread on LB (supplemented with rifampicin; 50 mg/ mL) agar plates and incubated at 30°C for 48 hours. Bacterial growth/counts were calculated as Colony Forming Units (CFU).

2.15 Electron Scanning Microscopy (SEM) and cell size measurements

Third, fourth, fifth and sixth rosette leaves were excised from 5 week old plants (32 DAS) and immediately fixed in the primary fixative containing 2% formaldehyde, 3% glutaraldehyde, 0.1 M phosphate buffer ($NaPO_4$; pH 7.2) at 4°C and were vacuum infiltrated until the leaves were wet. The fixative was replaced and the samples were fixed for at least 8 hours at room temperature. Samples were washed thrice (10-15 minutes each) in a phosphate buffer (0.1 M $NaPO_4$, pH 7.2) followed by dehydration in gradient ethanol series (25%, 50%, 75%, 95%) for 10-15 minutes each. Subsequently, samples were incubated in 100% ethanol for 1 hour.

Samples were dried to their critical point (CP) using liquid CO_2 as the CP fluid and 100% ethanol as the intermediary (Polaron E3000 series II critical point drying apparatus). Samples were mounted on to aluminium stubs using double sided tape and were sputter coated with approximately 100 nm of gold (Baltec SCD 050 sputter coater), and viewed in the FEI Quanta 200 Environmental Scanning Electron Microscope at an accelerating voltage of 20 kV. The

scanning electron microscopy imaging was carried out at the Manawatu Microscopy Imaging Centre, Massey University, Palmerston North, New Zealand. The images of the leaves were taken at different magnifications, such as, 200, 300, 400, and 600, of the same area of the same leaf in order to cover at least 30 cells per image. Three biological replicates (3 leaves of the same transgenic line) of each sample were fixed and photographed in order to reduce standard errors. Cell areas and perimeters were measured using ImageJ.

2.16 Ploidy level measurement using Flow Cytometer

Kinematic analyses (DNA content measurement) of the *CPR5* transgenic lines, including the CPR5 wildtype and *cpr5-2*, were carried out using flow cytometry as described in Jing et al. (2009). Leaf tissues (~50 – 100 mg) were collected from 3 week old plants sown in soil under long day conditions. Nuclei were isolated by chopping leaves with a razor blade in a glass petri dish containing 500 µl buffer (45 mM magnesium chloride, 30 mM sodium citrate, 20 mM 4-morpholinepropane sulfonate, and 0.1 % Triton X-100, pH 7.0). This finely grounded leaf mince was passed through a filter with a 30 µm pore size (Partec, CyFlow®). Flowthrough containing nuclei was stained with Propidium iodide by adding 1 mL of Propidium iodide solution (20 µg/ mL propidium iodide) and incubating at 4°C for 2 hours or overnight. Ploidy levels (2C, 4C, 8C, 16C, 32C DNA content) of each sample were quantified using Partec cytometer (Partec, CyFlow®) according to manufacturer's instructions. The number of nuclei present in each peak corresponding to respective ploidy levels was calculated using FlowMax® Peak analysis. The percentage of each nuclei population was calculated based on the total number of events (15000) and was displayed in the form of graphs.

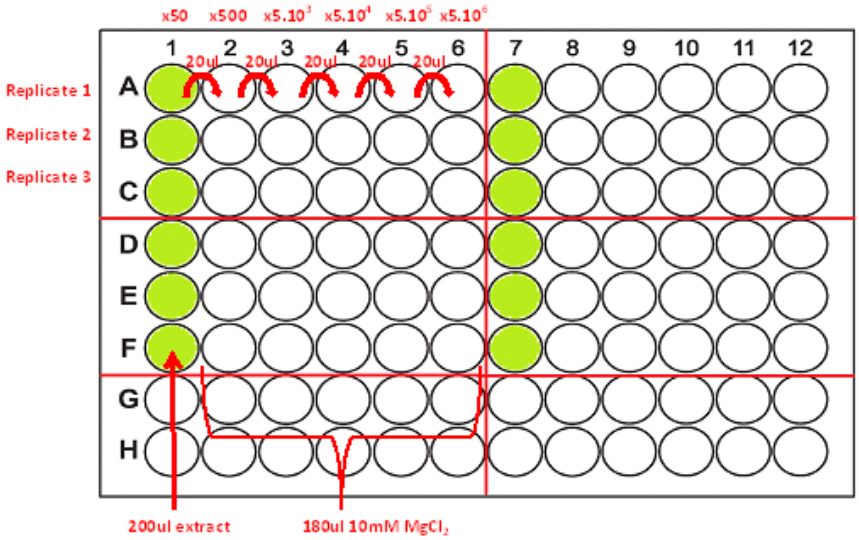


Figure 2.4 Schematic diagram for serial dilutions

This figure shows the serial dilution of leaf extract obtained from infiltrated plants samples, which were grounded in 500 µL of 10 mM MgCl2. The green wells represent 200 µL volume of original 500 µL from where 20 µL was added in the subsequent wells already containing 180 µL of 10 mM MgCl2.

2.17 Preparation of antibiotics

2.17.1 Ampicillin sodium salt (Sigma)

Stock solution: 100 mg/1 mL

Prepared 500 mg/5 mL by dissolving 0.5 g of Ampicillin in 5 mL of double distilled water and filtered through 0.2 μm. Stored at -20°C.

2.17.2 Kanamycin sulphate (Sigma)

Stock solution: 15 mg/1mL

Prepared 75 mg/5 mL by dissolving 0.075 g of kanamycin sulphate in 5 mL of double distilled water and filtered through 0.2 μm. Stored at -20°C.

2.17.3 Tetracycline hydrochloride (Sigma)

Stock solution: 12.5 mg/1mL

Prepared 62.5 mg/5 mL by dissolving 0.0625 g of tetracycline hydrochloride (95%) in 5 mL of 70% ethanol and filtered through 0.2 μm. Stored at -20°C.

2.17.4 Chloramphenicol (Duchefa)

Stock solution: 34.0 mg/1 mL

Prepared 170.0 mg/5mL by dissolving 0.170 g of chloramphenicol in 5 mL of 70% or 100% ethanol and filtered through 0.2 μm. Stored at -20°C.

2.17.5 Rifampicin (Duchefa)

Stock solution: 50.0 mg/1 mL

Working: 50 μL/1 mL

Prepared 500.0 mg/10 mL by dissolving 500 mg of rifampicin in 10 mL of 70% or 100% ethanol and filtered through 0.2 μm. Stored at -20.0°C.

2.17.6 Gentamycin (Duchefa)

Stock solution: 25 mg/mL

Prepared 250 mg/10 mL by dissolving 250 mg of gentamycin in 10 mL of water and filtered through 0.2 μm. Stored at -20.0°C.

Chapter 3 *In silico* Characterization of CPR5

3.1 INTRODUCTION

As mentioned in chapter 1, *cpr5-2* plants are shown to be defective in many developmental and physiological pathways, which indicate that CPR5 is a master regulator (Jing and Dijkwel, 2008). CPR5 is a plant specific protein, which displays no homology to any of the proteins with a known function. In addition to genetic homology, the 3-D structure of the CPR5 protein has also not been resolved yet. The pleiotropic roles of CPR5 suggest that the CPR5 protein may have a unique structure or multiple structural motifs, sites and/or regions that enable CPR5 to be involved in various processes or interactions. In order to test how CPR5 functions, coordinates with its interacting partners and participates in multiple pathways, extensive *in silico* analyses of the CPR5 gene or protein were performed through a number of computer-based bioinformatics tools. The CPR5 protein sequence was also analysed through a number of structure predicting computer programs in order to get an idea of the structural composition of CPR5. In the current study, the *in silico* analyses of CPR5 shows that CPR5 contains important structural motifs, phosphorylation sites, a nuclear localization signal, transmembrane domains, disordered regions, coiled-coil domains and glycine zippers.

3.2 RESULTS

3.2.1 *CPR5* is widely distributed throughout the plant kingdom

AtCPR5 is a master regulator which exert pleiotropic effects (Jing and Dijkwel, 2008). Therefore, it was hypothesised that *CPR5* is an essential gene and is part of many plant genomes. In order to find the presence of *CPR5*-like genes or proteins in different genomes, a deduced amino acid (At5G64930) sequence of AtCPR5 was obtained from TAIR (https://www.arabidopsis.org/). The obtained amino acid sequence of the CPR5 protein was BLASTed (using default settings) in the NCBI database (Section 2.1). The BLAST search displayed a list of about 50 CPR5-like proteins as shown in Appendix 11.2. These CPR5-like proteins are from different species are called CPR5 homologues (putative). As shown in Appendix 11.2, AtCPR5-like proteins are present in a variety of plant species, ranging from weeds, such as, *Eutrema salsugineum*, *Capsella rubella*,*Amborella trichopoda*; crops e.g. rice, maize, barely, cotton, tobacco, turnip, rapeseed, sorghum, clover, soybean; and many vegetable plants, for instance, tomato, potato, cucumber etc. and fruit plants, such as, grapes, muskmelon and date-palm. Additionally, *CPR5*-like genes are also present in some tree plants, such as, poplar and eucalyptus. However, At*CPR5* did not show homology with any genes from animals or human genomes. In conclusion, these results establish the fact that many of the plant species available in the database contain *CPR5*-like genes. Thus, AtCPR5 appears to be widely distributed among plant species.

3.2.2 CPR5 *in plant* homology amongst its homologues

As shown in Section 3.2.1, At*CPR5* is present in the majority of plants whose genomes were studied, which suggests that the AtCPR5 protein could have a high level of homology at the amino acid level. In order to find out the level of homology between AtPR5 and CPR5-like proteins (putative homologues) from different species, protein sequences were aligned using CLUSTAL omega. When compared to Moss, the AtCPR5 protein sequence shows 22% homology with the Moss protein sequence (Figure 3.1A). The CPR5-like protein (1008 aa) from Moss is shown to be 2-times longer than AtCPR5 (564 aa). Moss CPR5 was found to have an additional 200 amino acids at its N-terminus compared to AtCPR5 (Figure 3.1A). Similarly, protein sequences of closely related homologues, such as, *Brassica rapa*, *Camelina sativa*, *Eutrema salsuginum* and *Capsella rubella* show 63.3% homology with AtCPR5 (Figure 3.1B). The level of homology reduces to 42-45% when CPR5-like proteins from *Glycine max*,

Medicago trunculata, and *Ricinus communis* homologues were included in the aforementioned list of homologues and compared (Figure 3.1B and 3.1C). In comparison to CPR5-like proteins from *Glycine max*, *Eutrema salsuginum*, and *Ricinus communis*, AtCPR5 was found to contain an additional region of 37-150 amino acids at its N-terminus, that is highly polymorphic in composition (Figure 3.1B and 3.1C). Compared to AtCPR5, CPR5-like proteins from the monocot plant group, show homology at similar levels (49.7 %) as shown by eudicot homologues (45%, Figure 3.1D). Additionally, the majority of the AtCPR5 homologues belonging to the monocot species, such as, *Sorghum bicolar*, *Oryza sativa*, *Zea mays* etc. lack an N-terminal part of their proteins when compared to the Arabidopsis CPR5 protein sequence. Overall, CPR5-like proteins have relatively higher homology in the middle and C-terminal parts as compared to the N-terminal part. To conclude, Figure 3.2 shows that despite disparities i.e. some gaps or additions, there are many regions, sites and motifs, which are highly conserved among the CPR5 homologues from monocots.

3.2.3 AtCPR5 is prediticted to have 4 or 5 transmembrane regions

AtCPR5 is proposed to contain five transmembrane (TM) domains (Kirik et al., 2001; Yoshida et al., 2002), thus the AtCPR5 protein sequence was investigated in order to validate the proposed number of TM domains. In contrast to previous reports, the i*n silico* analyses of AtCPR5 carried out in the current study showed that AtCPR5 is predicted to contain 4 TM domains; with a failure to annotate the first TM domain from the N-terminus of protein (Figure 3.2B and 3.2C). Concomitantly, some TM predicting algorithms, such as, RHYTHM and TOPCON predicted 5 transmembrane regions (Figure 3.2A & 3.2D). The closer manual scan of the vague TM domain showed that this TM region is actually a mixture of both hydrophobic and hydrophilic amino acids, and hence was not annotated as a transmembrane region by SMART and TMHMM. Compared to AtCPR5, the position of putative TM domains also appears to be conserved (Figure 3.2G; only C-terminal half of the sequences spanning TM domains is shown). Figure 3.2G shows that the positions of the annotated second, third and fifth TM domains are highly conserved amongst selected AtCPR5 homologues in contrast to the first and fourth predicted TM domains. Furthermore, Figures 3.2E and 3.2F show that the N-terminus of the AtCPR5 protein is predicted to reside in the cytoplasm of the cell while C-terminal is buried inside a membrane. To conclude, this study suggests that AtCPR5 may have 4 or 5 putative TM domains with its N-terminus in the cytoplasm of the cell.

A.

```
Moss            MREDYFAWLFPGGRCRVGVGESMPASVPTPDICRRSMIQFRLVVSLPVVVLLLFAQVGPL 60
Arabidopsis     ------------------------------------------------------------

Moss            LAFPHHANSDAELEFGLHQRSHLDAKAGPEGESLAGVPNWNNEAPEWDELTRCIRSSESG 120
Arabidopsis     ------------------------------------------------------------

Moss            DTIARWLNCLESLEKFTAPDSAALSSTGLSSKIASSYADVNDSLQLAEDRMEPFDPLAGG 180
Arabidopsis     ------------------------------------------------------------

Moss            SSKLRKNLNSSDGNSTGGKGLVSARVQDSAPGLADENRTLADADVSREQGSGTSSGVDID 240
Arabidopsis     ----------------------MEALLLPPSPEPQNQITNPANSKPNHQSGD-VHKDETMM 38
                                     :.* :  .:*   ::  . *::. .::.*. . ..  :

Moss            MSENSGVSPSRISSKESSLLGRDNALGDADTARSSASNSGDENVGCLEVLTEASNSTSST 300
Arabidopsis     MKKKKDTNPSNLEKRK--LKGKKKEIMDNDEASSSYCS------------TSSTSNSNST 84
                *.::....**.:..::  * *:.: : * * * ** ..            *.::..:.**

Moss            TRLQNLLDALKLAACLGVVTKLLWGKRAQLLASGQKAQSGLTGQAVECLAPPTKQYSKAG 360
Arabidopsis     KRVTRVVHRLRNPMRLGMARRSVGERQAEKLAK--------------------------- 117
                .*: .::. *: .  **:. : :  ::*: **.

Moss            FSSLNNDVGAPSVKVGGLSTPLGMAIATVMAEALHQEDGRRPPRTADELSKLLEAALIET 420
Arabidopsis     --------------------PLGFSLAAFANMVIARKNAAGQNVYVDDLVEIFATLVEES 157
                                   ***:::*:.   .: :::.      .*:* ::: : : *:

Moss            LSPVMGSECPGFVEQFKQSFSSTYRLLGSVRSASSRNFASSFANHSRDVSQRRPEAPSRE 480
Arabidopsis     LANVYGNKLGSFATNFEQTFSSTLKILKLTN----------------------------E 189
                *: * *.:  .*. :*:*:**** ::*  ..                            *

Moss            CHRQTDAEEDDSSGLTKKANLDLAFNSRFRLPNRKHPDPQADLADQFSKLTLPVHGLEAV 540
Arabidopsis     CANPHQSNNNDGG------------------------------------------------ 202
                * .  :::::*..

Moss            AKSVDVKGASSDGENSSCVDQEVFGEPDDISSSVGLDDVSEGLSEKGTGHRVPLAVHPIG 600
Arabidopsis     --SCNLDRSTIDG----CSDTELFER-ETSSATSAYEVMQGSATATSLMNELALFEETLQ 255
                  * ::. :: **    * * *:* . :  *:: . : :. . : ..  :.:.*  ..:

Moss            IRNRLERRGLEKQTANQQIYRQERGSSLEPGDADLAMQPFETTSNDQLVAVFESSLAHYD 660
Arabidopsis     LSCVPPRSSAMALTTDERFLKEQT----------------------------------- 279
                :     * .    *::::: :::

Moss            SGTCYQQQKILLSQQEMELRKMELNLKFRELHLKEANLHLASENNTLMREHVDLSRNKVE 720
Arabidopsis     --------------RANDLKTVEIGLQIRELRCKETALGLKFESNNLGKAALELDVSKAA 325
                             :  :*:.:*:.*::***: **: * *  *.*.* :  ::*. .*.

Moss            FKEAEYQDRKLTAAHAIFAKRCADEMVAGLCVMLCALFFGAWKYSHERLVDIVSTCQPTL 780
Arabidopsis     FRAEKFKTELEDTRKEEMVTRIMDWLLVSVFSMLASMVLGVYNFSIKRIEDATSVCDQSE 385
                *:  ::: .   : :  :..*  * ::..:  **.::.:*.:::* :*: * .*.*: :

Moss            YVSNPLVPLLDSVFIVLPHVTLWMEPSNSYVPGFNTVVNSLNHMSSRLHMFVCEVTVASR 840
Arabidopsis     EKS----------------------SSWWVP------KQVSSINSGFNTFICRVRVWVQ 416
                  *                     *. :**      :.:. :.* :: *:*.* *  :

Moss            MLLGLGIISFVAAALLRKSVASNSQAMPATILIVVLGGICGFAGKLAVDSLGGSGYHWLL 900
Arabidopsis     IFFGV-LMIIVFTYFLNKRSSGTKQTMPISFIVLFLGIFCGVSGKLCVDTLGGDGKLWLI 475
                :::*: :: :* : :*.*  :...*:** :::::.** :**.:***.**:***.*  **:

Moss            MWEAFCLVHTLGTCFTSTLYRILNGTPESVTGKWQWQRQTYFMKPWLRRFWFHVTMVLVL 960
Arabidopsis     VWEVFCLLQFVANVFTLALYGLMFG-PINVTQETRSNRCNSMFPYWARRSVVYVVILFVL 534
                :**.***:: :.. ** :** :: * * .** : : :* . ::  * **  .:*.:::**

Moss            PILTGLIPFAHLREVVQLVAFNFWHLAYSPIESYLRKNHSQLTSVKEI 1008
Arabidopsis     PVINGLLPFATFGEWRDFAMYHLHGGSDYA------------------ 564
                *::.**:*** : *  ::. :::   :  .
```

B.

```
Capsella        MEALLLLSPSPEPQNPITDPANSKSHHQ---SDEEVNYETMIK---KKTKPSNSEKRKLK 54
Camelina        MEALLLLSSSPEPQKSITDPANSKSDHQ---SDEEVKDETMMKKKKKKTKQSDSGKRKHK 57
```

```
Arabidopsis      MEALLLP-PSPEPQNQITNPANSKPNHQ---SGDVHKDETMMMKKKKDTNPSNLEKRKLK 56
Brassica         MDALLLP-PSPDPEESITDPANQRANHRPENSSDELKDETMRK----KMKPSPLTKRKLK 55
Eutrema          ---------------------------------------MMT---KKIKPSPLEKRNLK 17
                                                         *      . : *   **: *
Capsella         GKKKEINMNDDEASSSSC---STSNSNSTRRVSRVAHRLRNPTVRLGMSRRSVGERQAEA 111
Camelina         GKKK--EMNNDEASSSSCSISSTSNSNSTRRVSRVAHRLRNPTVRLGMSRRSVAERQAEA 115
Arabidopsis      GKKKE-IMDNDEASSSYCSTSSTSNSNSTKRVTRVVHRLRNP-MRLGMARRSVGERQAEK 114
Brassica         AKKKE----SDEASSSSCSSSSASNLNSTRRVSRVAHRLRSPPVRLGFTRRSVGERQAEA 111
Eutrema          PKKT---MNNDEASSSCS---SSSNPNYTRRVSRVAYRLRNPTVRLGMARRSVGERQAEA 71
                  **.     .****** .   *:** * *:**:**.:***.* :***::****.*****

Capsella         LALPLG----------MSFAAFAILVLQRKNAAGQNVNIDDLAGISATAVTESFANVFGD 161
Camelina         LAFPLG----------MSFAALANLVLQRKNAAGQNVFVDDLAMISARAATETVASVYGY 165
Arabidopsis      LAKPLG----------FSLAAFANMVIARKNAAGQNVYVDDLVEIFATLVEESLANVYGN 164
Brassica         LALPLG----------MSFAAFANLVLQRKSADGQNVYVDDLAGISASAVKESLANVYGD 161
Eutrema          LARPLVYIICSVYILFMRYLTFVHMVLQR-NAAGQNLHVDDLAVISTSAVKESLANVYGD 130
                 ** **           :   ::. :*: * .* ***: :***. * :  . *:.*.*:*

Capsella         KLGSFSTNFEKSFNSTLKILNLINESANPPQLKNNDVGSCNLDRSTIDGYSDTELYAKET 221
Camelina         KLGSFATTFEKSFNHTLRILKVTNESANPRQLNNDDVGRCNLDCSIIDEYSDTELSAKET 225
Arabidopsis      KLGSFATNFEQTFSSTLKILKLTNECANPHQSNNNDGGSCNLDRSTIDGCSDTELFERET 224
Brassica         KLGSFATNFEKSFNSTLKILKLINESTSPHRLDNTNVASCDLDRSTLDGCSDSELSAKET 221
Eutrema          KLGSFATNFERSFKSTLSILKLTNEYTFPQHSNNN-----NVATSTIDECSDTELSAMET 185
                 *****:*.**::*. ** **:: ** : * : .*      ::  * :*  **:**   **

Capsella         SSAMSAYEEIQG--SATATSLMNELVLHEETRQLSCAPRRSSAMSLTTAERFFGEQVQEA 279
Camelina         SSSTSVYEVIQG--SAIATTSMNELVLHEPSRQLSCIPPLSSEMSSTAVVRCL----TEA 279
Arabidopsis      SSATSAYEVMQG--SATATSLMNELALFEETLQLSCVPPRSSAMALTTDERFLKEQ-TRA 281
Brassica         SSTTSAYEAIQGGTRTTTTSSMNEIVLHEETRQLSTVSRTSSAMPLTTFEKALEEQ-ARA 280
Eutrema          SSATSAYEVIQG----------------------------ERSEMSLTTFERRRLEEKARA 218
                 **: *.** :**                              * *. *:  :       .*

Capsella         NDLKKVELGLKKIELSLKESELGLRYDSNNLGKRKLEIGVEQAAFQADKFKTKLEDTRKA 339
Camelina         NDLQREANSLKKIGLTYKKMELGINYESNSLQKSQLEVDVSKTAFKKEKFKTELEDTRKA 339
Arabidopsis      NDLKTVEIGLQIRELRCKETALGLKFESNNLGKAALELDVSKAAFRAEKFKTELEDTRKE 341
Brassica         NDLKSMEIGLTMRKLRLKETELALSYESNNLGKSKLEMNVSKAAFRAEKFKTEQEDSKKA 340
Eutrema          NDLKTMEIGLKMTKIAISS-------ESNNLEKAKLEMAASKAAFKDEKFKTELEDSRRD 271
                 ***:    .*    :  ..       :**.* *  **: ..::**: :****: **:::

Capsella         EMVTKIMDCLVVSVFSMLASMLIGAYNFSQKRIADATSICEKPEEKSSSWWVPKQVSNLN 399
Camelina         EMVTKIMDWLGLSVFSMSVSMFIGAYNFSLKRIEDATAACEPSEEPSSTWWVPKHVSTLN 399
Arabidopsis      EMVTRIMDWLLVSVFSMLASMVLGVYNFSIKRIEDATSVCDQSEEKSSSWWVPKQVSSIN 401
Brassica         DMVHKIMDWLVLSVFTMLASMLYGAYVFSHQRITEAASICEPSEEKTSSWWVPKQVSSIN 400
Eutrema          DFVVKTMDWLGVSVFIMVACMIYGASVFSQEKIKGATSICEPSKEKSYLWWG---FSWLN 328
                 ::* : ** * :*** * ..*. *.  ** ::*  *:: *: .:* :  **    .* :*

Capsella         SD-FSAMICRIRVWVQMCFGFVMICVFIYFMMQRSSGRTQTMPISFLLIFLGIVCGLPVK 458
Camelina         SE-FHTMICRIRVWVQFVFGLIMLLVLAYFIMQRSSGRTQAMPISFIVIVLGIFCGLPVK 458
Arabidopsis      SG-FNTFICRVRVWVQIFFGVLMIIVFTYFLNKRSSGTKQTMPISFIVLFLGIFCGVSGK 460
Brassica         AE-FNILLCRLRVWVQSFFGVLMILVITYFIIQSGT-AKQTMPVTFIVIFLGVVCGLPGK 458
Eutrema          TESINILICRLGVWLK-YFGVLMILVFTYLIMSRSS---QTRPVTFIVVFLSIVCGLAGE 384
                 :  :  ::**: **::  **.:*: *: *:: . .:   *: *::*:::.*.:.**:. :

Capsella         VCVDTLGGNGKLWLILWQVFCVFQFVANVCTLALYRLMYGPISVTEGTKRSRWS--TIFP 516
Camelina         VCVDTLGGDGKLWLMFWEVFCLFHLFANVFTLALYGLMHGPINVTQVTKSSRCK--TLFP 516
Arabidopsis      LCVDTLGGDGKLWLIVWEVFCLLQFVANVFTLALYGLMFGPINVTQETRSNRCN--SMFP 518
Brassica         FCVDTLGGDGKLWLLLWETFCVLQFVANVFTLAFYRLMYGPV----TTQGGRCD--AMVP 512
Eutrema          FCVETLGGNSDLWLYLWQVFCVLLFLANVFTDVVYGLIYGPVTVTQGTRLSRRDKNTWVP 444
                 .**:****:..*** .*:.**:: :.*** * ..* *:.**:     *: .* .  : .*

Capsella         YWARRTVLYVVFMLALPAITGLLPFATVGEWRDLAMYNFLGESASIPEV 565
Camelina         YWARRSVFYVVFIFVLPAITGLLPFATFGEWRDLAMDNFFGGSASDSKV 565
Arabidopsis      YWARRSVVYVVILFVLPVINGLLPFATFGEWRDFAMYHLHGGSDYA--- 564
Brassica         YWARRSVFFALVVLALPVMNGLLPFATYGEWREHWATWWAGPGSDD--- 558
Eutrema          YWARRCFFFGLIRFVLPAITGLLAFATFGEW----IDHFFGGRGSKD-- 487
                 ***** ..: :. :.**.:.***.*** ***         *
```

C.

```
Eutrema          ---------------------------------------------MMTKKIKPSPLEKRNLK        17
Brassica         -MDALLLPPSPDPEESITDPANQRANHRPENSSDELKDETM----RKKMKPSPLTKRKLK        55
Arabidopsis      -MEALLLPPSPEPQNQITNPANSKPNHQSGD---VHKDETMMMKKKKDTNPSNLEKRKLK        56
Capsella         MEALLLLSPSPEPQNPITDPANSKSHHQSDE---EVNYETMIK---KKTKPSNSEKRKLK        54
Medicago         -MLLLFIETPMA-------------------E-IETCSSQ-----------SA------S        22
Gylcine          ------------------------------------------------------------        0
Ricinus          ------MEPPPS-----VTPSNSTEHSNPTAD-NPSCLPTMIKN-------KN------R        35
```

```
Eutrema          PKKTM---NNDEA-SSSCS--SS---SNPNYTRRVSRVAYRLRNPTVRLGMAR---RSVG         65
Brassica         AKKKE----SDEASSSSCSSSSA---SNLNSTRRVSRVAHRLRSPPVRLGFTR---RSVG        105
Arabidopsis      GKKKEIMDND-EASSSYCSTSST---SNSNSTKRVTRVVHRLRNP-MRLGMAR---RSVG        108
Capsella         GKKKEINMNDDEASSSSCS---T---SNSNSTRRVSRVAHRLRNPTVRLGMSR---RSVG        105
Medicago         KNRAG---------LEICETNSYNSSQLQRTKVKGKGIACKRRNPRVSV-------RRTR         66
Gylcine          ------------------------------------------------------------         0
Ricinus          KKKKV---------FKTCAQTSSSSSA-------ATRFPYKRRNSKVVLAPVRRGGRSFG         79

Eutrema          ERQAEALARPLVYIICSVYILFMRYLTFVHMVLQ-RNAAGQNLHVDDLAVISTSAVKESL        124
Brassica         ERQAEALALPLGMSF----------AAFANLVLQRKSADGQNVYVDDLAGISASAVKESL        155
Arabidopsis      ERQAEKLAKPLGFSL----------AAFANMVIARKNAAGQNVYVDDLVEIFATLVEESL        158
Capsella         ERQAEALALPLGMSF----------AAFAILVLQRKNAAGQNVNIDDLAGISATAVTESF        155
Medicago         GNNVAAIGFPLGMSV----------AAVMAQVLYRRDAAAEGMSPDHLSSMCTSAIQESL        116
Gylcine          ----------------------------------------------MCSSAIKESL         10
Ricinus          DSQMEAVALPLGMSF----------AAIIAKFMETKDAAGDKMSVDRLTKICASAVRESL        129
                                                                : :: : **:

Eutrema          ANVYGDKLGSFATNFERSFKSTLSILKLTNEYTFPQHSNNNNVA----------TSTIDE        174
Brassica         ANVYGDKLGSFATNFEKSFNSTLKILKLINESTSPHRLDNTNVA-----SCDLDRSTLDG        210
Arabidopsis      ANVYGNKLGSFATNFEQTFSSTLKILKLTNECANPHQSNNNDGG-----SCNLDRSTIDG        213
Capsella         ANVFGDKLGSFSTNFEKSFNSTLKILNLINESANPPQLKNNDVG-----SCNLDRSTIDG        210
Medicago         SSVFGDKLDGLTRNFEQSFDSTLSTLRLIYESPASNEGNKLNNMRLEIPSSKFIRGD---        173
Gylcine          ASVFGDKLDGLTRNFEQSFCSTLSTLQLIYESSKSNEGNKLNNTKMEIMSSKLTLNKEE-         69
Ricinus          ANVFGDKFDFFARNFEKSFGSTLSTLRLINEASLNRETCHLSQVDVENSASDLNLNNGIR        189
                 :.*:*:*:  :: ***::* ***. *.*  *     .  . .

Eutrema          CSDTELSAM-------------------ETSS----------------------ATSAYE        193
Brassica         CSDSELSAK-------------------ETSS----------------------TTSAYE        229
Arabidopsis      CSDTELFER-------------------ETSS----------------------ATSAYE        232
Capsella         YSDTELYAK-------------------ETSS----------------------AMSAYE        229
Medicago         CSRDIAIEE---------------DQPGLLS-HAHAE-------------IV---DQSNR        201
Gylcine          CSGDIVTEVGHSRHAEIQDQSISHDSPEEVRDNFHIDSIGRDSPEEGRDNIHNNSVSLDS        129
Ricinus          CMSCI-KDC----HSE---------TP--LTS-----------------SLFQNQMRMDE        216


Eutrema          VIQGER----------------------------SEM-----SLTTFERRRLEEKARANDL        221
Brassica         AIQGGTRTTTTSSMNEIVLHEETRQLSTVSRTSSAM-----PLTTFEKALEEQA-RANDL        283
Arabidopsis      VMQGSA--TATSLMNELALFEETLQLSCVPPRSSAM-----ALTTDERFLKEQT-RANDL        284
Capsella         EIQGSA--TATSLMNELVLHEETRQLSCAPRRSSAM-----SLTTAERFFGEQVQEANDL        282
Medicago         REEVGDNFHMESVTRDVALHGQSNQMVCFAPPSSGALVNNPVISTFEKSIMEQC-RSNDL        260
Gylcine          PEESRDNFHMGSVSRDLTLYGQSNQMVSFSQISFGS-VNNPMVSIFEKSIMEQC-RSNDL        187
Ricinus          DVE----DSSPSMNQELVLHGQINQQVCTP-CSSGS-IIN-QLSTVEKSIMEHA-RSNDL        268
                   :                                        ::  *:   *.  .:***

Eutrema          KTMEIGLKMTK-------IAISSESNNLEKAKLEMAASKAAFKDEKFKTELEDSRRDDFV        274
Brassica         KSMEIGLTMRKLRLKETELALSYESNNLGKSKLEMNVSKAAFRAEKFKTEQEDSKKADMV        343
Arabidopsis      KTVEIGLQIRELRCKETALGLKFESNNLGKAALELDVSKAAFRAEKFKTELEDTRKEEMV        344
Capsella         KKVELGLKKIELSLKESELGLRYDSNNLGKRKLEIGVEQAAFQADKFKTKLEDTRKAEMV        342
Medicago         KTVELGLTMKKMKLKETELALRYDLNDLQRSKLAVDISKTSFKAEKFKTQLEDTRYGELN        320
Gylcine          KTLELGLKMKELKLKEDELALNLDSNNLNRSKLVMGESKQSFKEEKFKTQLEDTRHGELK        247
Ricinus          KTFKIGLTMRKLKLKEAALALNLDANHLERSKLAMGMSKASFKIEKFKTQLEDTRHAELL        328
                 *..::**   :       :.:  : *.* :  * :  .: :*: :****: **::  ::

Eutrema          VKTMDWLGVSVFIMVACMIYGASVFSQEKIKGATSICEPSKEKSYLWWGFSWLN--TESI        332
Brassica         HKIMDWLVLSVFTMLASMLYGAYVFSHQRITEAASICEPSEEKTSSWWVPKQVSSINAEF        403
Arabidopsis      TRIMDWLLVSVFSMLASMVLGVYNFSIKRIEDATSVCDQSEEKSSSWWVPKQVSSINSGF        404
Capsella         TKIMDCLVVSVFSMLASMLIGAYNFSQKRIADATSICEKPEEKSSSWWVPKQVSNLNSDF        402
Medicago         KKCIDCLIAGLFIMSSSLFYGAYVYNFERINEAAESCTPEEASIIFFYSLFQFYS-----        375
Gylcine          KKCIDCLITGLLIMSSSLLYGAYVYSYERITKATESCTPSTQESSSWWTPKSMVLFNSKL        307
Ricinus          KKCIDCLVAGLFIMSASILYGAYVFSYQRITDATASCSPLIEDSKSWWIPRPFSSFNSGW        388
                  : :* *  .:: * :.:. *.  :. ::*  *:  *     .   ::    .

Eutrema          NILICRLGVWLK-YFGVLMILVFTYLIMSRS---SQTRPVTFIVVFLSIVCGLAGEFCVE        388
Brassica         NILLCRLRVWVQSFFGVLMILVITYFIIQS-GTAKQTMPVTFIVIFLGVVCGLPGKFCVD        462
Arabidopsis      NTFICRVRVWVQIFFGVLMIIVFTYFLNKRSSGTKQTMPISFIVLFLGIFCGVSGKLCVD        464
Capsella         SAMICRIRVWVQMCFGFVMICVFIYFMMQRSSGRTQTMPISFLLIFLGIVCGLPVKVCVD        462
Medicago         ------------M-----LQAIIIFLLLQDENN-------------------------        391
Gylcine          HILWCQVQVMSRMAFGILMIFAVAYLLLQRSTTTSQTMPVTFILLMLGIFCGYCGKLCVE        367
Ricinus          HTLRCHVQVASRMLFGILIILAVAYLLLQRSATSRQTMPVTFILLLLGAACGFAGKLCVD        448
                                 :   . ::: .

Eutrema          TLGGNSDLWLYLWQVFCVLLFLANVFTDVVYGLIYGPVTVTQGTRLSRRDKNTWVPYWAR        448
Brassica         TLGGDGKLWLLLWETFCVLQFVANVFTLAFYRLMYGPVTT----QGGRC--DAMVPYWAR        516
Arabidopsis      TLGGDGKLWLIVWEVFCLLQFVANVFTLALYGLMFGPINVTQETRSNRC--NSMFPYWAR        522
```

```
Capsella        TLGGNGKLWLILWQVFCVFQFVANVCTLALYRLMYGPISVTEGTKRSRW--STIFPYWAR        520
Medicago        ------------------------------------------------------------        391
Gylcine         TLGGSGIVWLLYWEIMCLLHFLSLCWTSALFRILHGPVTPSQTM-----EENTILRYWIR        422
Ricinus         TLGGSGFHWLLYWETMCLLHFFSNVCISTLFCILHGPLTVSRRT-----NYNSLCPYWFR        503


Eutrema         RCFFFGLIRFVLPAITGLLAFATFGEWIDHFFG------GRGSKD---  487
Brassica        RSVFFALVVLALPVMNGLLPFATYGEWREHWATWW---AGPGSDD---  558
Arabidopsis     RSVVYVVILFVLPVINGLLPFATFGEWRDFAMYHL---HGGSDYA---  564
Capsella        RTVLYVVFMLALPAITGLLPFATVGEWRDLAMYNF---LGESASIPEV  565
Medicago        ------------------------------------------------  391
Gylcine         RVLFYAAMLVFLPLFCGLMPFASLGQWKEHFTLKGSDFNGSEW-----  465
Ricinus         RILFYAALLLVLPLCCGLLPFADPGEWKDHFFLLATNFLSITDD----  547
```

D.

```
Arabidopsis     MEA---------------------------------LLLPPSPEPQNQITNPANSKPNHQS        28
Vitis           MDTPLHSLQTLESSVCGSTEESSDADLVDTISVDSIISHANP--TVQFPPPTSAST----        54
Sorghum         MDGS--------------------------------HIAAASP--E--AGGA-SSSA----        20
Oryza           MDAA--------------------------------AASSSSS--SATMAAA-AASA----        22
Zea             ------------------------------------------------------------        0


Arabidopsis     GDVHKDETMMMKKKKDTNPSNLEKRKLKGKKKEIMDNDEASSSYCSTSSTSNSNSTKRVT        88
Vitis           --TTASKLINMKKK------------KKKKKRALVD-------VSAP-SSSSNSMYVRVG        92
Sorghum         --SSAS-----------------------------S-------YG---GSASGRSWLRKG        39
Oryza           --AEAS-----------------------------S-------SGSPASSRNARHRQRKG        44
Zea             ------------------------------------------------------------        0


Arabidopsis     R----VVHRLRNPMR-LGMARRSVGERQAEKLAKPLGFSLAAFANMVIARKNAAGQNVYV        143
Vitis           N----KRRNPR-----ILMGLNRRSESEAEAIALPLGMSIAAIVSQILERKDEAGERMSV        143
Sorghum         VHLRRRRRRV------AVRGGERAGGDDVQDLALPLGMSFAAVLAQVLNRCNGSEGRLRP        93
Oryza           VRLRMLRRRGRQPVEAERAPGDGGGGAVQEDLALPLGMSFAAVLAQVINTKNISGQRLHP        104
Zea             ---------------------------------------MSFAAVLAQVVNTKNRSGERLQP        23
                                                       :*:**.   ::   : :  .:


Arabidopsis     DDLVEIFATLVEESLANVYGNKLGSFATNFEQTFSSTLKILKLTNECANPHQSNNN----        199
Vitis           DHLSMICTMAVRESLANVFGDKFDCFVRNFEKSFGSTLRTLRLINESPLNKVGGHLSHFN        203
Sorghum         DVLSKMCTSAVKESLTNIYGDRFQSFMGNFEKAFGSTLRTIHLINETPVYEQD-MAQ-CS        151
Oryza           DFLSKICTSAVKESLTNIYGDSSNSFIKKFEKSFSSTFRTLHLVNEIPVNERSHIPE-CS        163
Zea             ALLSKICTSAVKESLRNIYGNKLDSFMRNFEKSFSSTLTTLHLVNEMPVYGQGPVPQ-CS        82
                  *  : :  *.*** *::*:   .*  :**::*.**:  ::* **


Arabidopsis     ---DGGSCNLDRSTIDGCS-----------DTELFERETSSATSAYEVMQGSATATSLMN        245
Vitis           VQNSGSDLNLPVPLNRGDTISSSSISDHHSSMGLHPVATLECLNTLEERQESIQTDSANR        263
Sorghum         YK-------------DGNSVPEIKLGGADSLPE-IKLGGADSQSQIHDVQTDMPSSSMNN        197
Oryza           FK-------------HDDSVAVDSLSS------------SDLQNQTNRIEHD-LVNTVES        197
Zea             SK-------------HEDSEASSKLST------------GGTQNLTREIKQE-RLNSVES        116
                                :                          .  .  : .    :


Arabidopsis     ELALFEET-LQLSCV-----PPRSSAMALTTDERFLKEQTRANDLKTVEIGLQIRELRCK        299
Vitis           ELTVHGQISQQLACVSSSSLGSVMNRSTVSTFEKSVVEQARSNDLKALELSLSMKKMKLK        323
Sorghum         QIVLHAGVNHQLVHLTRSRSNPEIDQHILSVFERSLNEQTRSNELKELEIGLTMRKLQLS        257
Oryza           QLVLFASDNQQLMHLRRSRSSPEADNRILNAID-------RSNELKEFEIGLTMRKLQLK        250
Zea             QLVLYAGGYQQMTRRAHSLSSPEADQRILNAFERSLKEQTRSNELKEFEISLSMRKLKLK        176
                ::.:.     *:            .   :.. :       *:*:** .*:.* ::::: .


Arabidopsis     ETALGLKFESNNLGKAALELDVSKAAFRAEKFKTELEDTRKEEMVTRIMDWLLVSVFSML        359
Vitis           ETQLALNSDSYLLERMKFSMGISKASFKAEKFKNQLEDTKHAELLRKCIDCLVAGLLIMS        383
Sorghum         RSQLALSSESHMLEKIKVSMGFEKASFKEKKFKTEMEDTKHAELFRKLIDMLLTAVVLMS        317
Oryza           QSQLALSSHSHMLEKIKLSFGFQKASFKGEKFKTHMQETRDAEILRTLIDFLVSAVIVMS        310
Zea             QSQLELNSYSHMLDKIKLSLGFQKASFQVEKFKTQMQDTRHAQIMRTLIDFLVSAVIIMS        236
                .: * *.  *  * :  ..: ..**:*: :***..:::*:. ::.   :* *: .:. *


Arabidopsis     ASMVLGVYNFSIKRIEDATSVCDQSEEKSSSWWVPKQVSSINSGFNTFICRVRVWVQIFF        419
Vitis           GSLVYGAYIYSFKRITEATETCTASHKEFKSWWIPKPMASLNSGFYILRCQVQVVSRMMF        443
Sorghum         VCFGYGTYIYSYQRITAATSACAAISREYTSWWMPSSVSAFNSGLLLFKCHVIAATRIFF        377
Oryza           ACFGYGTYIYSYQRITDVTSACSATSKGFKSWWMPNSVSNFSSGFLFLRCHVIAVTRMCF        370
Zea             VCFGYGTYTYSYQRITDITAACSATSKGSKSWWMPNSVSNFNTGLLFIRCHVIAATRMFF        296
                 .:  *.* :* :**   * .*    .  .***:*. :: :.:*:  : *:* .  :: *


Arabidopsis     GVLMIIVFTYFLNKRSSGTKQTMPISFIVLFLGIFCGVSGKLCVDTLGGDGKLWLIVWEV        479
```

```
Vitis            GIFMIVAIAYLLLQRSASSKQTMPVTFIILLLGVVCGVAGKFCVDTLGGSGYHWLLYWEA        503
Sorghum          GILMLLLIGWLIFQRSAMTGPNMPVTFNVMILGVLCGYFGSRCVDTLGGDGNVWLIYWEA        437
Oryza            GILMILAIAWLAFQRSSTTGSNMPITFNLILLGIICGFAGRFCTNTLGGDGNTWLMYWEV        430
Zea              GIVMILAVVWLAFQRSAVSVSSMPVTFNIILLGVICGYAGRFCTNTLGGDGNMWLICWEV        356
                 *:.*:: . ::  :**: :  .**::* :::**:.**  *  *.:****.*  **: **.

Arabidopsis      FCLLQFVANVFTLALYGLMFGPINVTQETRSNRCNSMFPYWARRSVVYVVILFVLPVING        539
Vitis            LCLLHFVSNICTSALFLLLHGPITVSQGAK---GNTMIPYWIRRTLFYAAILLFLPLLCG        560
Sorghum          LCSIHLLGNIWPSLLHHILYGPISVTHRTK----AVHLPYWTRRYAFYVLVTLILPCLAG        493
Oryza            LCSIHLLGNLFPSLLYHVLHGPISVSHREQ----VVWLPYWVRRCLFYGAVGLILPALTG        486
Zea              LCSVHLIGNCYPSVLYRVLHGPISVSHSKE----SVWFPYWIRRWMFYAVLGFFIPALNG        412
                 :* ::::.*     *. ::.***.*::  .       :*** **  .*  : :.:* : *

Arabidopsis      LLPFATFGEWRDFAMYHLHGGSDYA----  564
Vitis            LMPFASPGEWKDRFLLLLADSLSN--IDD  587
Sorghum          LLPFASLSDWKELAVQYLKSRFSGSNIED  522
Oryza            LLPFASLSDWKDHFVEEIKSIVIGDKIEA  515
Zea              LLPFASLSDWNNHFTKELKSIFIDERIEA  441
                 *:***: .:*.:     :
```

Figure 3.1 Multiple sequence alignment of CPR5 putative homologues

This figure shows the protein sequence alignments of CPR5-like proteins; **A)** Arabidopsis versus *Moss*, **B**) eudicot relatives, such as, *Capsella*, *Camelina*, *Brassica* and *Eutrema*, **C)** *Capsella*, *Brassica*, *Eutrema Medicago*, *Glycine* and *Ricinus* and **D)** monocot homologues. The protein sequences were extracted from the NCBI database (see Appendix 11.2 for accession numbers of the proteins) and were aligned using CLUSTAL omega (http://www. ebi.ac.uk/Tools/msa/ clustalo/)default settings.

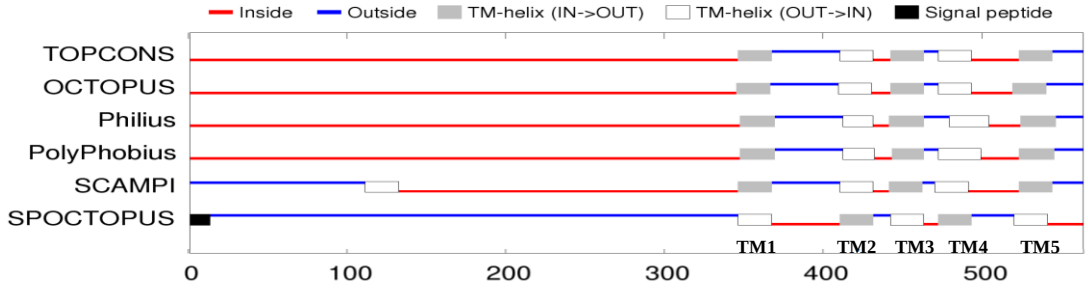
3.2.4 AtCPR5 carries two casein kinase phosphorylation sites

AtCPR5 is predicted to contain NLS (Jing et al., 2002, Yoshida et al., 2002). It is reported that a number of nuclear proteins carry casein kinase (CK) phosphorylation site(s) which either flank CK site or are prsent adjacent to the putative NLS clusters (Harreman et al., 2004). CK site get phosphorlated in order to facilitate NLS-mediated nuclear localisation (Harreman et al., 2004). Interestingly, the CPR5 protein was predicted to contain two casein kinase phosphorylation sites I (CKI) and site II (CKII) when CPR5 was analysed. The CKII site resides at 46-49 positions between the annoated NLS cluster-A and region-B, whereas, the CKI site locates at 71-81 positions downstream of the annotated NLS cluster-C in the CPR5 protein. Based on their positions, CKI and CKII phosphorylation sites essentially flank the annotated NLS clusters B and C. Additionally, the CKI and CKII phosphorylation site appears to be conserved in CPR5 dicot homologues compared to monocot homologues, which were predicted to contain only one CK1 phosphorylation site. Based on these studies, it can be concluded that *Arapidopsis* CPR5 is prdicted to contain CKI and CKII phosphorylations sites in the vicinity of puative NLS.

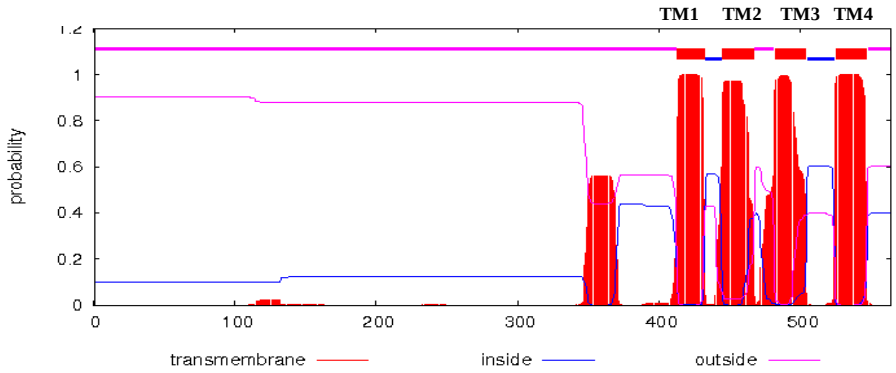
3.2.5 AtCPR5 carries multiple transcritpion or translation initiation sites

It is likely that a gene transcript is translated from multiple positions if the transcript has multiple in-frame start codons, transcription initiation supporting nucleotides (for example, purines at position -3, +3 and +4 of the putative and start codon), and RNA stem-loop formations (Kozak, 1991, Kozak, 1999, Chabregas et al., 2003). As shown in Figure 3.3, At*CPR5* possesses 4 in-frame start codons; the first at 1-3, second at 112-114, third at 187-189 and fourth at 292-294 nucleoties of *CPR5* cDNA. Similarly, *CPR5* mRNA is predicted to contain 4 stem-loop structures in the first 190 nucleotides of *CPR5* gene (Figure 3.3). Of these 4 stem-loops, the first stem-loop appears to be strong, as 50% of the nuceleotides associated in the formation of the first stem-loop are guanine (G) and cytisine (C) that make triple bonds between nucleotides. Additionally, a putatative stem-loop structure is present before each putative start codon (Figure 3.3). Similarly, purines (adenine and guanine; A/G) which are reported to be favourable for transcription initiation are also present at the expected positions (-3, +3 and +4) of each start codon of AtCPR5 (Figure 3.3). To coclude, these results show that At*CPR5* is annotated to contain multiple transcription initiation sites, RNA secondary structures (stem-loops) between the start codons, and transcription initiation optimal context (favouable residues) at the expected positions.

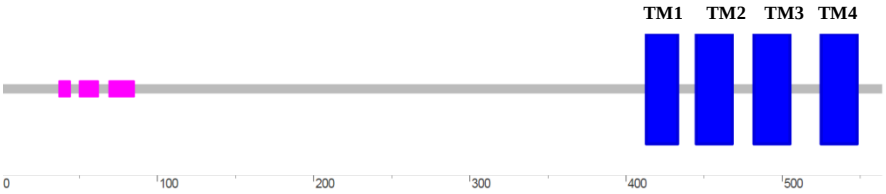
A.


Inside
Outside
TM-helix (IN->OUT)
TM-helix (OUT->IN)
Signal peptide
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
TM1
TM2
TM3
TM4
TM5
0
100
200
300
400
500

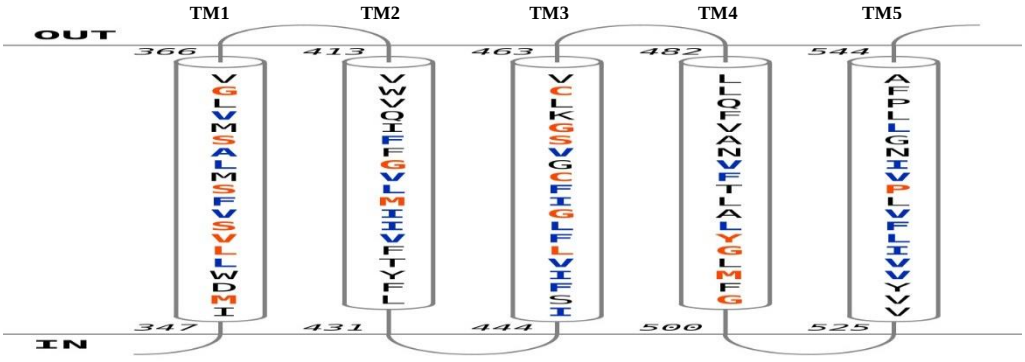
B.


TM1
TM2
TM3
TM4
probability
1.2
1
0.8
0.6
0.4
0.2
0
0
100
200
300
400
500
transmembrane
inside
outside

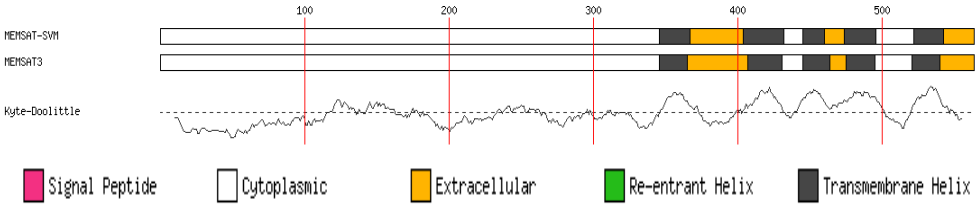
C.


TM1
TM2
TM3
TM4
0
100
200
300
400
500

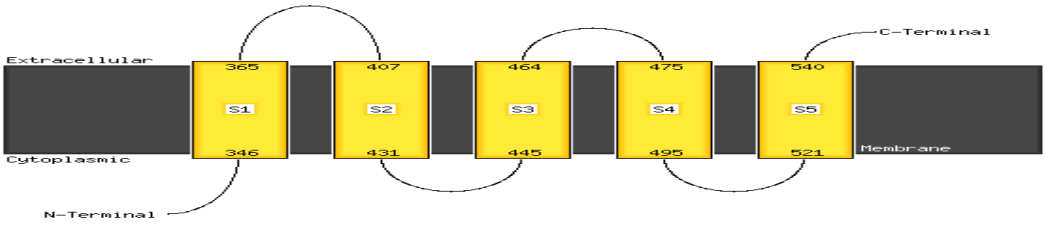
D.


TM1
TM2
TM3
TM4
TM5
OUT
IN
366
413
463
482
544
347
431
444
500
525

E.


100
200
300
400
500
MEMSAT-SVM
MEMSAT3
Kyte-Doolittle
Signal Peptide
Cytoplasmic
Extracellular
Re-entrant Helix
Transmembrane Helix

F.



G.

```
                                                    TM1
O. sativa           FKGEKFKTHMQETRDAEILRTLIDFLVSAVIVMSACFGYGTYIYSYQRITDVTSACSATS 336
Z. mays             FQVEKFKTQMQDTRHAQIMRTLIDFLVSAVIIMSVCFGYGTYTYSYQRITDITAACSATS 262
S. bicolor          FKEKKFKTEMEDTKHAELFRKLIDMLLTAVVLMSVCFGYGTYIYSYQRITAATSACAAIS 343
A. thaliana         FRAEKFKTELEDTRKEEMVTRIMDWLLVSVFSMLASMVLGVYNFSIKRIEDATSVCDQSE 385
P. patens           FKEAEYQDRKLTAAHAIFAKRCADEMVAGLCVMLCALFFGAWKYSHERLVDIVSTCQPTL 780
                    *:  ::: .   : .  :     * :: .:  *  .:  *.: :* :*:   .:.*

O. sativa           KGFK----------------SWWMP-------------NSVSNFSSGFLFLRCHVIAVTR 367
Z. mays             KGSK----------------SWWMP-------------NSVSNFNTGLLFIRCHVIAATR 293
S. bicolor          REYT----------------SWWMP-------------SSVSAFNSGLLLFKCHVIAATR 374
A. thaliana         EKSS----------------SWWVP-------------KQVSSINSGFNTFICRVRVWVQ 416
P. patens           YVSNPLVPLLDSVFIVLPHVTLWMEPSNSYVPGFNTVVNSLNHMSSRLHMFVCEVTVASR 840
                       .                 : *:             ..:. :.: :  : *.* .  :

O. sativa           MCFGILMILAIAWLAFQRSSTTGS-NMPITFNLILLGIICGFAGRFCTNTLGGDGNTWLM 426
Z. mays             MFFGIVMILAVVWLAFQRSAVSVS-SMPVTFNIILLGVICGYAGRFCTNTLGGDGNMWLI 352
S. bicolor     TM2  IFFGILMLLLIGWLIFQRSAMTGP-NMPVTFNVMILGVLCGYFGSRCVDTLGGDGNVWLI 433
A. thaliana         IFFGVLMIIVFTYFLNKRSSGTKQ-TMPISFIVLFLGIFCGVSGKLCVDTLGGDGKLWLI 475
P. patens           MLLGLGIISFVAAALLRKSVASNSQAMPATILIVVLGGICGFAGKLAVDSLGGSGYHWLL 900
                    : :*: ::  .     ::*  :    ** :: ::.** :**  *  ..::***.*  **:
                                                    TM3
O. sativa           YWEVLCSIHLLGNLFPSLLYHVLHG-PISVSHREQV----VWLPYWVRRCLFYGAVGLIL 481
Z. mays             CWEVLCSVHLIGNCYPSVLYRVLHG-PISVSHSKES----VWFPYWIRRWMFYAVLGFFI 407
S. bicolor     TM4  YWEALCSIHLLGNIWPSLLHHILYG-PISVTHRTKA----VHLPYWTRRYAFYVLVTLIL 488
A. thaliana         VWEVFCLLQFVANVFTLALYGLMFG-PINVTQETRSNRCNSMFPYWARRSVVYVVILFVL 534
P. patens           MWEAFCLVHTLGTCFTSTLYRILNGTPESVTGKWQWQRQTYFMKPWLRRFWFHVTMVLVL 960
                     **.:* :: :.. :.  *: :: * * .*:   .       :  * **  .:  : :.:

O. sativa           PALTGLLPFASLSDWKDHFVEEIKSIVIGDKIEA-------------- 515
Z. mays             PALNGLLPFASLSDWNNHFTKELKSIFIDERIEA-------------- 441
S. bicolor     TM5  PCLAGLLPFASLSDWKELAVQYLKSRFSGSNIED-------------- 522
A. thaliana         PVINGLLPFATFGEWRDFAMYHLHGGSDYA------------------ 564
P. patens           PILTGLIPFAHLREVVQLVAFNFWHLAYSPIESYLRKNHSQLTSVKEI 1008
                  * : **:*** : :  :     :
```

Figure 3.2 Annotation, position and number of TMs in CPR5 protein

This figure shows the total number, positions and orientation of the transmembrane regions (TMs). TM domains were predicted using (**A**) TOPCON, (**B**) TMHMM, (**C**) SMART, and (**D**) RHYTHM. For the determination of CPR5 orientation, PSIPred (E) and XtalPred) (F) were used. Amino acid sequences of CPR5 and CPR5-like proteins were analysed for the prediction of TM domains and the positions of annotated domains are highlighted in green text (G). http://topcons.cbr.su.se/pred/result/rst_0llBPs/; http://www.cbs.dtu.dk/services/ TMHMM-2.0/; http://smart.embl.de/smart/show_motifs.pl; http://proteinformatics.charite.de/ rhythm /index.php?site=helix; http://ffas.burnham.org/XtalPred-cgi/xtal.pl

3.2.6 AtCPR5 is annotated to contain intrinsically disordered regions

Many of the proteins which are shown to be involved in multiple pathways, are known to be intrinsically disordered proteins (IDP) (Sun et al., 2010a). Thus, it was assumed that AtCPR5 could be one of the IDPs or contain intrinsically disordered regions (IDRs) within it. The CPR5 protein was investigated through a number of *in silico* tools in order to find the presence of intrinsically disordered regions in CPR5. As shown in Figure 3.4, AtCPR5 is predicted to contain a number of disordered regions. Closer examination of the annotations shown in Figure 3.4 reveals that AtCPR5 essentially contains 3 or 4 stretches of intrinsically disordered regions. For example, the first IDR ranges from 1-33 amino acids of AtCPR5, the second from 37-58, and the third from 63-94 (Figure 3.4B and 3.4C). The fourth predicted stretch of disordered region is present in the middle part of the protein that ranges from 180-220 amino acids of AtCPR5 protein (Figure 3.4B and 3.4C). In conclusion, the AtCPR5 protein is predicted to have a number of disordered regions, thus AtCPR5 could be one of the IDPs.

3.2.7 AtCPR5 is also predicted to contain coiled-coil domains

A close investigation of the CPR5 protein sequence revealed that AtCPR5 contains a number of conserved leucine residues. Thus it was hypothesized that leucine residues of AtCPR5 could be part of coiled-coil domains in which leucine residues are ordered in the form of an exclusive arrangement called *heptad* (Zhang et al., 2008). As shown in Figure 3.5, COILS predicted 4 coiled-coil regions in the AtCPR5 protein sequence based on three windows; 1) window 14 that considers and compares 14 residues (two *heptad*s); window 21 that analyses 21 residues, and window 28, in which 28 residues of 4 *heptads* are analysed. For example, AtCPR5 is predicted to have leucine zippers of 2 (14 residues; represented with a peak in green color) 3 (21 residues; represented with blue peaks) and 4 (28 residues; represented with red peak) *heptad*s. To conclude, AtCPR5 is proposed to have three types of leucine zippers at four different positions in the protein.

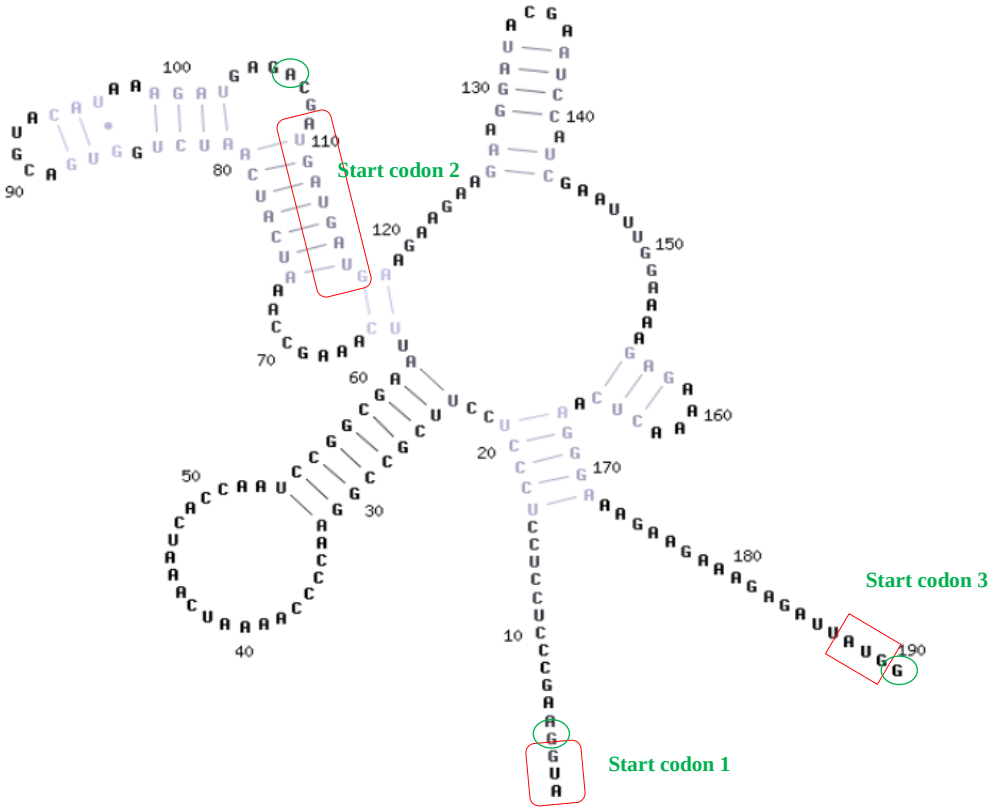


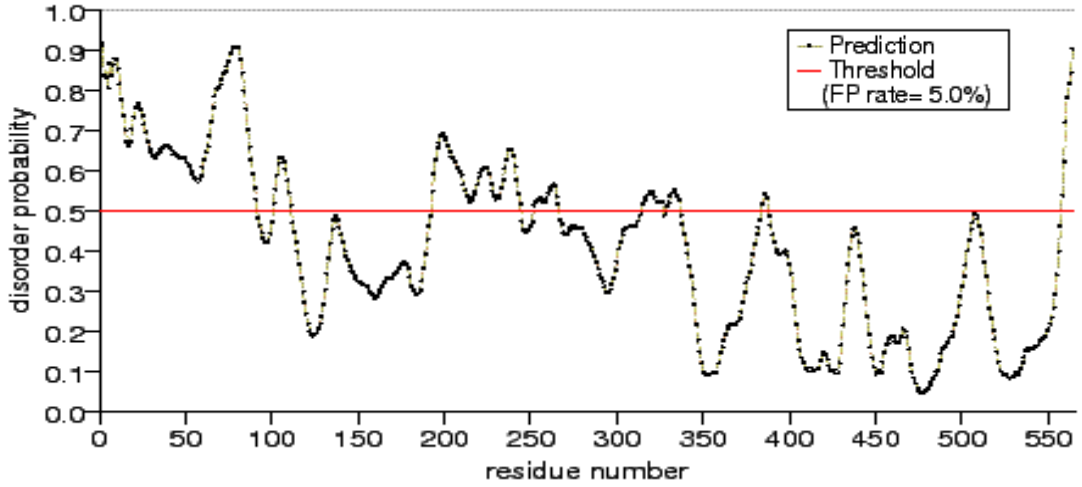
Figure 3.3 Positions of in-frame start codons and RNA stem-loop structures

This figure shows the positions of the predicted stem-loop structure formation in the *CPR5* coding sequence. The nucleotides highlighted in red boxes represent start codons (ATG), and green oval shapes represent the position of transcription favourable nucleotides in the vicinity of start codons.

3.2.8 CPR5 is predicted to have glycine zipper

In addition to leucine, glycine residues were also found to be conserved within AtCPR5. Glycine residues are known to be an integral part of TM domains and glycine zippers. As mentioned above, CPR5 is predicted to possess 4 or 5 transmembrane regions. Therefore, it was assumed that glycine residues could be part of the glycine zipper of the AtCPR5 TM domain. Analysis of CPR5 protein sequence showed that a typical glycine motif can be observed in the third (from N-terminus of CPR5) transmembrane region (452-463) of AtCPR5 (Figure 3.6). The other transmembrane regions also carry glycine residues, however, the typical pattern was found to be missing in them. Studies suggest that the residues that follows glycine in the typical glycine motif has frequently been a hydrophobic molecule, specifically, alanine (Kim et al., 2005). As seen in Figure 3.6, 2 out of the 3 glycine residues present in the third TM region of CPR5 are followed by hydrophobic residues, for instance, isoleucine follows glycine at position 452 and valine follows glycine at position 456. From these studies, it can be concluded that the annotated TM3 of AtCPR5 is predicted to carry the typical pattern of a glycine motif.

A.



B.

1	**MEALLLPPSP EPQNQITNPA NSKPNHQSGD VHKDETMMMK KKKDTNPSNL**	50
51	**EKRKLKGKKK EIMDNDEASS SYCSTSSTSN SNSTKRVTRV V**HRLRNPMRL	100
101	**GMARRSVGER Q**AEKLAKPLG FSLAAFANMV IARKNAAGQN VYVDDLVEIF	150
151	ATLVEESLAN VYGNKLGSFA TNFEQTFSST LKILKLTNEC AN**PHQSNNND**	200
201	**GGSCNLDRST IDGCSDTELF ERETSSATSA YEVMQGSATA TSLM**NELALF	250
251	E**ETLQLSCVP PRSSAM**ALTT DERFLKEQTR ANDLKTVEIG LQIRELRCKE	300
301	TALGLKFESN NLGK**AALELD VSKAAF**RA**EK FKTELE**DTRK EEMVTRIMDW	350
351	LLVSVFSMLA SMVLGVYNFS IKRIEDATSV CDQ**SEEK**SSS WWVPKQVSSI	400
401	NSGFNTFICR VRVWVQIFFG VLMIIVFTYF LNKRSSGTKQ TMPISFIVLF	450
451	LGIFCGVSGK LCVDTLGGDG KLWLIVWEVF CLLQFVANVF TLALYGLMFG	500
501	PINVTQETRS NRCNSMFPYW ARRSVVYVVI LFVLPVINGL LPFATFGEWR	550
551	DFAMYHL**HGG SDYA**	600

C.

D.

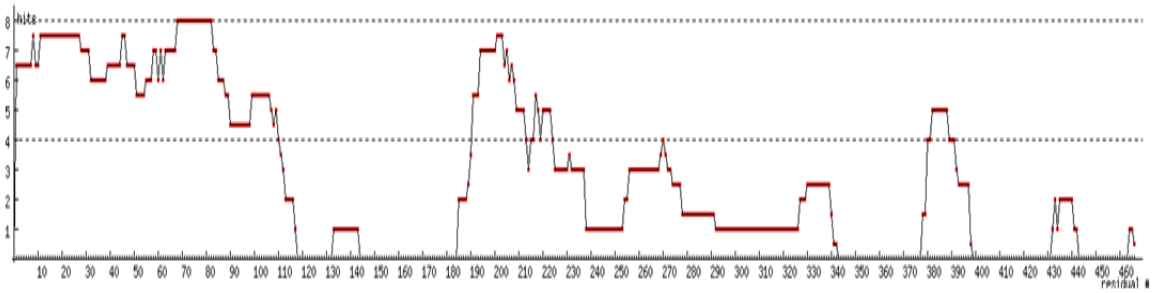


Figure 3.4 Position and number of Intrinsically Disordered Regions

These figures show the total number of disordered regions and their respective positions in the protein CPR5 annotated by MetPreDOS (**A**) IUPred (**B**) and DisPred (**C**). The bold letters in Figure 3.4B represents predicted disordered residues. Similarly, the blue rectangles in Figure 3.4C show amino acids annotated as disordered. Parts of the CPR5 proteins above threshold levels (middle dotted line in Figure 3.4A and 3.4D) represent intrinsically disordered regions of CPR5 protein.

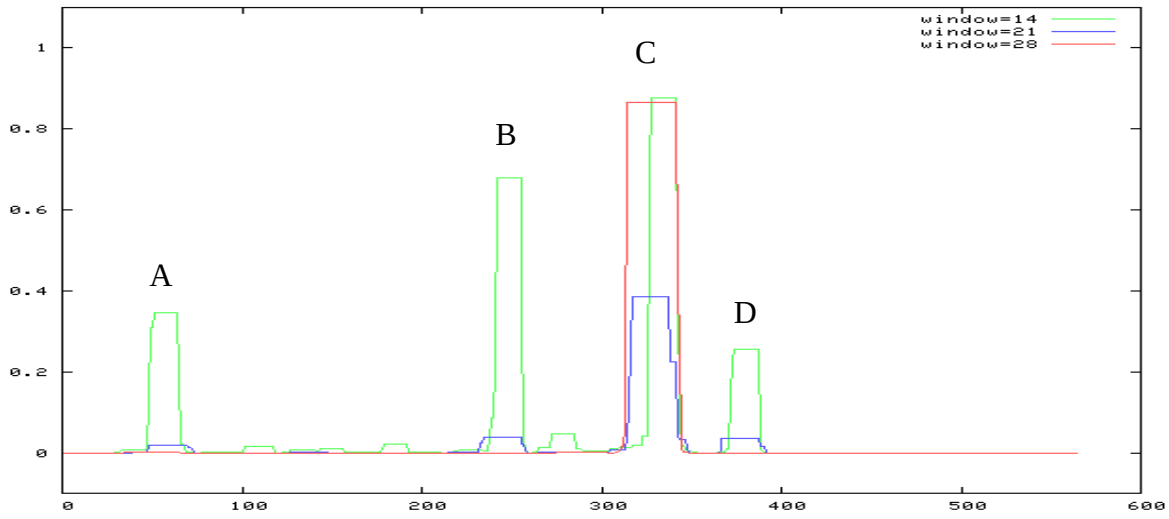


Figure 3.5 Position and number of annotated coiled-coil regions

This figure shows the position and number of coiled-coil or leucine zipper in the CPR5 protein. Window 14 takes 14 (7x2) amino acids, window 21 takes 21, and window 28 takes 28 amino acids into consideration and measures the possibility of any putative coiled-coil part of the protein. COILS was used to estimate the coiled-coil possibility in protein CPR5, which can be accessed at http://www.ch.embnet.org/software/COILS_form.html.

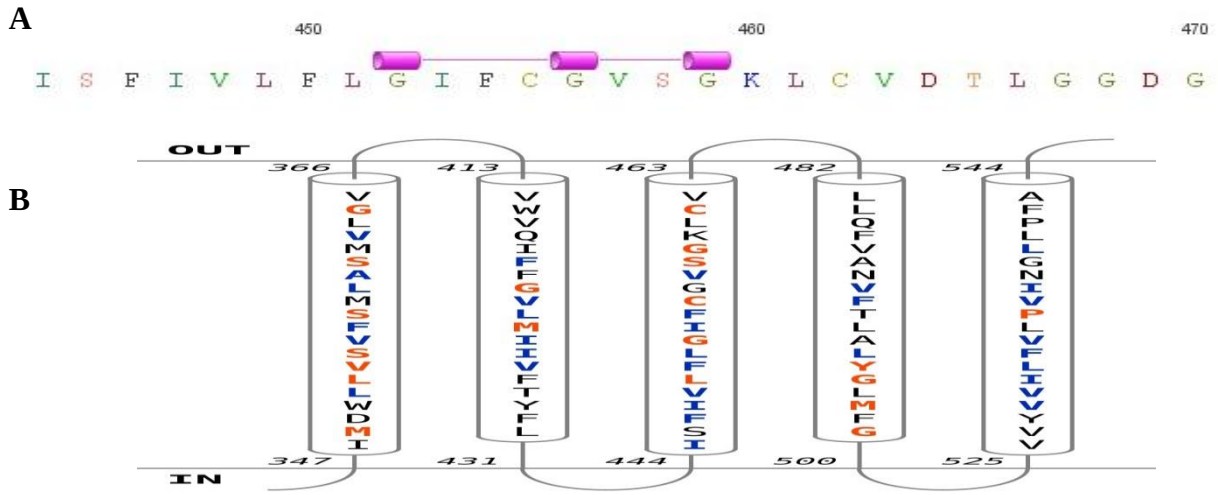


Figure 3.6 Position of glycine residues in CPR5 protein

Figure 3.6A shows the position of glycine amino acids of the predicted glycine zipper wherease, 3.6 B shows positions and amino acids of the predicted glycine zipper which could be involved in helix formation. Figures 3.6A and 3.6B were prepared using Geneious R6 (Biomatters, Limited) and RHYTHM (http://www.cbs.dtu.dk/services/TMHMM-2.0/), respectively.

3.3 DISCUSSION

3.3.1 AtCPR5 appears to be an important membrane protein

The current chapter essentially summarises and discusses the annotation of various putative structural motifs or sites predicted in AtCPR5. For example, the significance and need of the *CPR5* gene in plants is evident from the presence of CPR5-like proteins in the wide majority of plant genomes including, Moss. Additionally, low levels of identity among amino acids of CPR5-like proteins further show the level of diversity present among CPR5-like proteins from different plant species. The prediction of TM domains and glycine zipper reinforce the hypothesis that CPR5 is a membrane protein. The genetic analyses of At*CPR5* alleles, such as, *cpr5-2* (Bowling et al., 1997), *old1-1* (Jing et al., 2002) and *hys1* (Yoshida et al., 2002) show that positons of many mutations were in the putative TM regions of AtCPR5. Thus, loss-of-function phenotypes in *cpr5-2*, *old1-1* or *hys1* could be the consequence of mutation in transmembrane regions. A mutation in a membrane region could lead to improper folding and embedding in to the membrane and could eventually result in the change of orientation of the soluble part of the protein. For example, the soluble N-terminal half of the protein with five transmembrane domains and cytoplasmic orientation could be orientated into the extra-cellular space, perinuclear space or nucleoplasm due to improper folding and embedding, resulting in the loss of function of understudy protein.

Similarly, the prediction of three clusters in contrast to the previous bipartite NLS prediction (Jing et al., 2002, Yoshida et al., 2002), provides further evidence for AtCPR5 being a nuclear membrane protein. CKI and CKII phosphorylation sites further strengthen the existence of a functional NLS and AtCPR5 nuclear localisation. Lastly, the presence of putative start codons, stem-loop structure and transcription initiation supporting nucleotides reinforce the possibility of isoform(s) production from AtCPR5, which needs further investigation.

In summary, *in silico* studies reinforce that AtCPR5 is an important membrane protein. Additionally, the latest bioinformatics tools allowed for the carrying out of extensive and thorough analyses of CPR5, which provided a strong basis for the creation of various AtCPR5 transgenic lines and functional characterization, in order to confirm the role or implication of predicted motifs in numerous biological pathways.

Chapter 4 AtCPR5 may be one of the Intrinsically Disordered Proteins

4.1 INTRODUCTION

Several proteins adopt a well-defined structure, yet there is a large list of proteins that acquire no globular (secondary or tertiary) shape (Uversky and Dunker, 2010, Tompa, 2012, Forman-Kay and Mittag, 2013). Proteins, which can fold into a well-defined structure are classified as structured proteins (Jirgensons, 1966, Murzin et al., 1995, Orengo et al., 1997). On the other hand, proteins or segments of proteins that fail to form a definite structure are referred to as intrinsically disordered proteins (Dunker et al., 2013, van der Lee et al., 2014). Proteins, which carry part of disordered regions are known as intrinsically disordered regions (IDR) or intrinsically disordered proteins (IDPs) (Reed et al., 2015, van der Lee et al., 2014).

In addition to the alpha-helices, beta-strand and transmembrane domains (hydrophobic regions), disordered regions are also important parts of a protein (Dunker (Dunker et al., 2001, Huang and Sarai, 2012). In the protein database, thousands of proteins are reported to have disordered regions (Oldfield and Dunker, 2014, Peng and Kurgan, 2015). The majority of species contain IDRs or IDPs. However, they are more prevalent in eukaryotes, for instance, 25-30% of eukaryotic proteins are classified as intrinsically disordered regions or proteins (Dunker et al., 2000). Similarly, 44% of human proteins have been found to contain disordered regions in them (Oates et al., 2013). Surprisingly, 70% of the proteins known to be involved in signalling are reported to be intrinsically disordered regions (Oldfield et al., 2005).

Intrinsically disordered regions or proteins are difficult to classify into groups (Vucetic et al., 2003). However, some IDPs can be induced to fold upon interaction with protein partners, while the conformation of other IDPs remains unchanged upon binding with partner(s). Thus IDPs can be divided into foldable and non-foldable (Bardwell and Jakob, 2012, Yegambaram et al., 2013, Oldfield and Dunker, 2014). Similarly, the disordered regions of IDPs vary in length and thus based on their length, IDPs can also be divided into two groups 1) long disordered regions (LDR) and 2) short disordered regions (SDR). Short disordered regions of IDPs generally act as loops or coils, whereas, long disordered regions render the interaction interface(s) for other proteins to bind and interact with IDPs (Sun et al., 2014, Dunker et al.,

2015). Of human proteins, ~ 45% of proteins contain IDRs and the majority of these IDRs are 30, or less than 30, residues (Oates et al., 2013).

Intrinsically disordered regions or proteins act as a hub for the signalling of various proteins (Oldfield and Dunker, 2014), which enables IDPs to bind to multiple partners simultaneously (Kriwacki et al., 1996, Dunker et al., 1998, Hasty and Collins, 2001) either by having more than one binding site (Uversky and Dunker, 2010) or by altering their shapes due to their flexible nature (Dunker et al., 2005, Xie et al., 2007, Oldfield et al., 2008, Hsu et al., 2013). Alternatively, IDRs/IDPs bind to a single binding partner by transiting from a disordered-to-ordered shape (Oldfield et al., 2008, Hsu et al., 2013). Intrinsically disordered regions or proteins carry small segments or motifs of residues, which essentially undergo disorder-to-order transition and mediate interaction(s) with respective partner(s). These motifs or segments have been named differently by different research groups, for example, molecular recognition features (MoRFs) (Mohan (Mohan et al., 2006, Vacic et al., 2007, Kotta-Loizou et al., 2013), disordered binding regions (Linding et al., 2003) or pre-structural motifs (Lee et al., 2014).

Being flexible and lacking a definite structure, intrinsically disordered proteins or regions are known to perform more than 40 types of functions (Dunker et al., 2015). IDRs serve as linker between the structured domains of the protein (Qian et al., 1992, Wright and Dyson, 1999, Buske et al., 2015), act as elastic entropic springs (Trombitas et al., 1998), elastomers, entropic bristles (Hoh, 1998) and native molten globules (Uversky et al., 1997, Ebert et al., 2008). IDRs contain many important sites for glycosylation, methylation, phosphorylation, posttranslational modification, protease digestion, RNA, DNA or protein binding, and nuclear localisation signals etc. (Frankel et al., 1987, Xie et al., 1998, Uversky et al., 2000, Bracken, 2001, Gao and Xu, 2012, Crick et al., 2013). They are also shown to facilitate the movement of molecules through narrow pores, and are implicated in cell division, cell signalling, gene regulation, alternative splicing and protein-protein interactions (Dunker et al., 2005, Dyson and Wright, 2005, Tompa et al., 2005, Radivojac et al., 2007, Dunker et al., 2015).

Keeping in view the typical patterns and characteristics of the already characterized IDPs and IDRs, the predicted intrinsically disordered regions (IDRs) present in AtCPR5 are compared with the typical IDRs/IDPs here. This chapter mainly provides a structural comparison of AtCPR5 and the typical IDR or IDP. Further to the *in silico* analyses carried out in the current study, the i*n vitro* functional characterisation of putative CPR5 IDRs will reveal the implication

of these regions in CPR5 functioning as discussed in Chapter 8. The ultimate objective of the current study was to identify the exclusive patterns of IDRs in the CPR5 IDRs to show that these patterns of CPR5 IDRs are consistent with typical IDRs, thus CPR5 could be one of the IDPs.

4.2 RESULTS

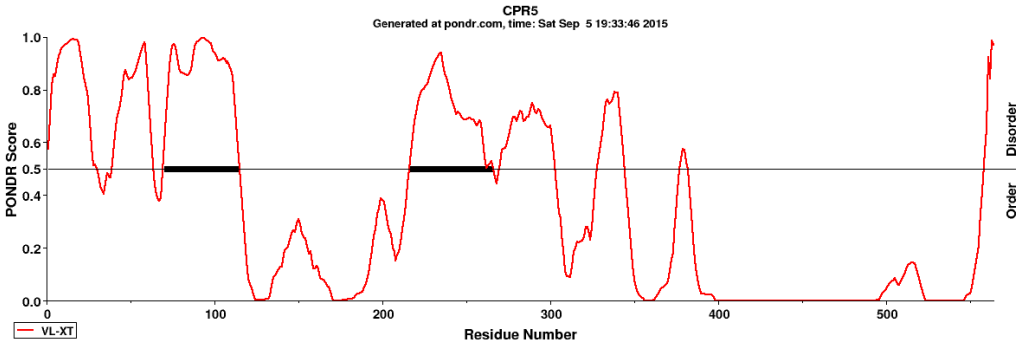
4.2.1 Like typical IDRs, CPR5 IDR is highly polymorphic in amino acid composition

As suggested by *in silico* studies, the CPR5 protein is annotated to contain intrinsically disordered regions (IDRs) at its N-terminus (Figure 4.1). Typically, disordered regions of proteins tolerate the accumulation of amino acid changes at a higher rate than structured parts of the protein. Therefore, IDRs of protein are polymorphic in amino acid composition when compared to its homologues (Light et al., 2013, Khan et al., 2015). Thus, AtCPR5 IDRs were hypothesised to be polymorphic when compared to the disordered regions of CPR5-like proteins from other species. To test, the amino acid sequence of AtCPR5 and its homologues were aligned and compared using Geneious software. The amino acid sequence alignment of AtCPR5 and CPR5-like proteins from different plant species revealed that CPR5 IDRs are highly polymorphic in amino acid composition compared to the N-terminal sequence of CPR5 homologues. Despite polymorphism, there are certain amino acids within the N-termini of AtCPR5 and CPR5-like proteins, which are conserved, for example, serine residues (Figure 4.2). These conserved residues could act as a reference point for the start of homology between CPR5 and CPR5-like proteins since part of the CPR5 protein displayed higher homology after these serine residues. In summary, the highly polymorphic N-termini of CPR5-like proteins (putative IDRs) are consistent with previous studies, which indicate that CPR5 IDRs could be true intrinsically disordered regions.

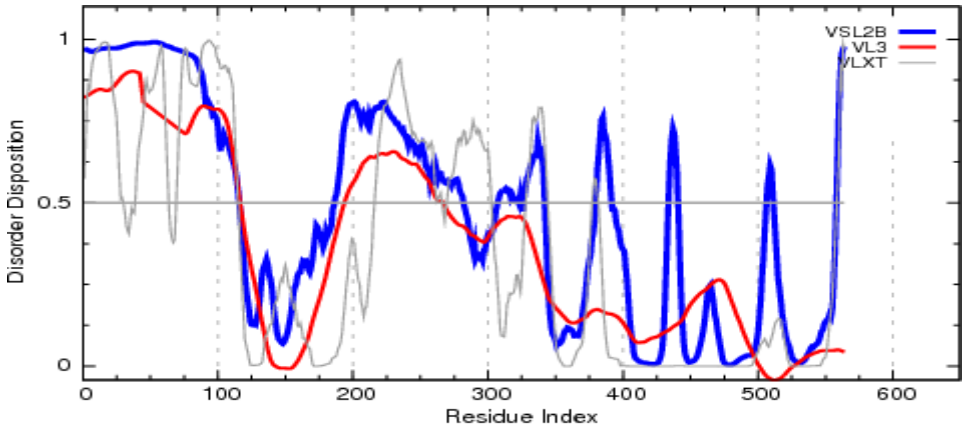
4.2.2 Amino acid composition of AtCPR5 IDRs is consistent with typical IDRs

Compositionally, typical IDRs are enriched in residues, such as, aspartic acid (D), glutamic acid (E), lysine (K), arginine (R), methionine (M), serine (S), glutamine (Q), and proline (P) (Theillet et al., 2013). Therefore, AtCPR5 IDRs were assumed to be enriched in with the aforementioned amino acids. When counted, ~ 70% of the amino acids present in AtCPR5 IDRs are shown to be from the aforementioned list of residues, compared to the presence of these residues (~ 25%) in the remaining (structured) portions of AtCPR5 (Figure 4.2). Additionally, these residues are shown to be present in the form of short repeats as suggested by (Theillet et al., 2013). In conclusion, these studies reinforce that CPR5 IDRs are abundant in typical residues, which are frequently found in IDRs.

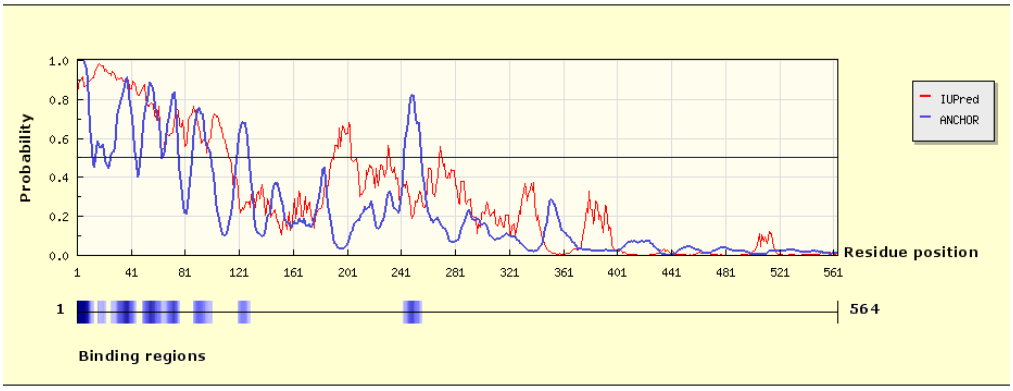
A.


CPR5
Generated at pondr.com, time: Sat Sep 5 19:33:46 2015
PONDR Score
Disorder
Order
Residue Number
VL-XT

B.


VSL2B
VL3
VLXT
Disorder Disposition
Residue Index

C.


IUPred
ANCHOR
Probability
Residue position
1
564
Binding regions

D.

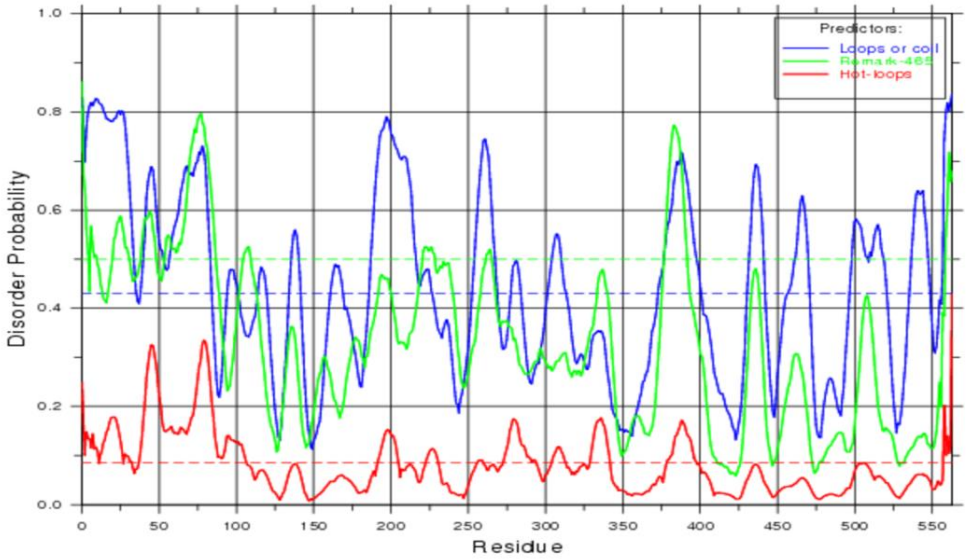


Figure 4.1 Position of disordered and binding regions in AtCPR5

These figures show the positions of disordered and ordered regions, molecular recognition features and disordered binding regions of CPR5 analysed by PONDER (A), MetaPrDOS (B), ANCHOR (C), and DisEMBL (D). To achieve these Figures, the CPR5 protein sequence was analysed using the online available FoldIndex web server.

Disordered Binding Region (ANCHOR)					**Molecular Recognition Features (MoRFs)**	
	From	**To**	**Length**			**Residues annotated**
1	1	12	12		**MoRF1**	MEA
2	16	21	6		**MoRF2**	PEP
3	26	44	19		**MoRF3**	HKDET
4	49	63	15		**MoRF4**	KK
5	65	76	12		**MoRF5**	DEA
6	87	100	14		**MoRF6**	HRLRL
7	12	128	9			
8	243	256	14			

Table 4.1 Positions of predicted MoRFs and binding regions in CPR5

This table shows residues and their positions, which were annotated as MoRFs and disordered binding regions when analysed by MoRF and ANCHOR predictors respectively. To achieve this table, the CPR5 protein sequence was uploaded on the online available MoRFs and ANCHOR web servers.

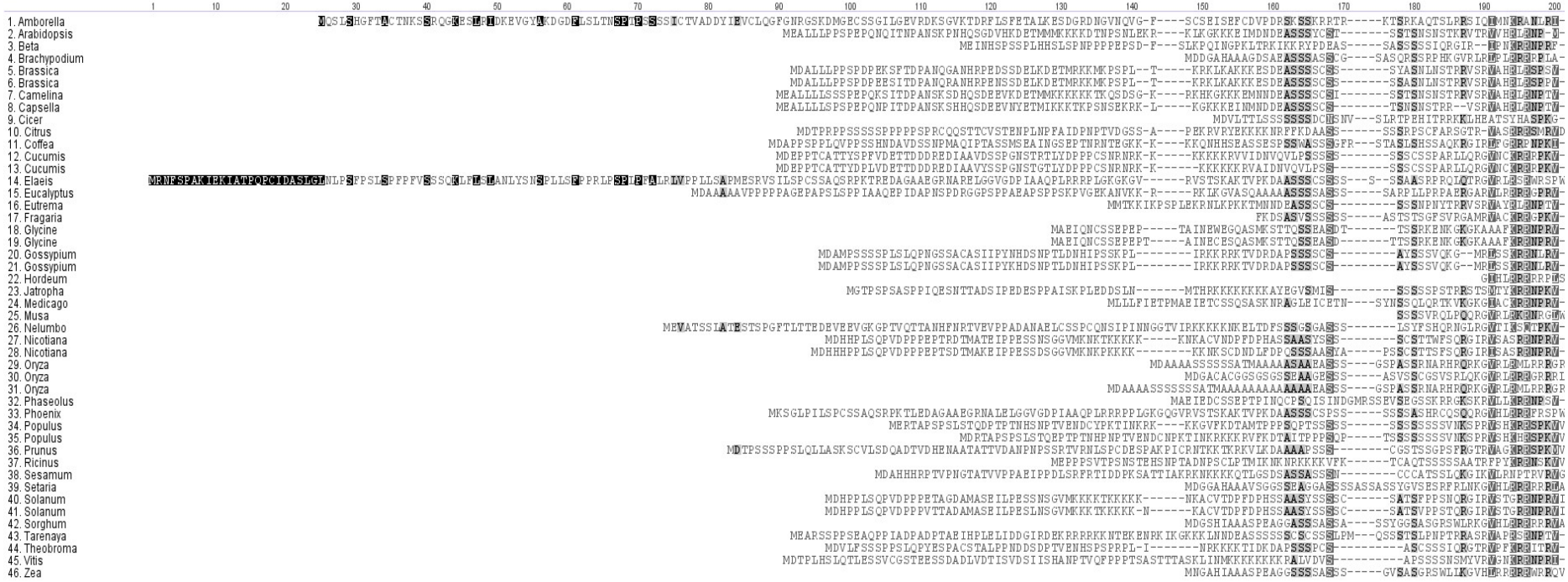


Figure 4.2 Presence of polymorphism in the N-terminus of CPR5 and CPR5-like proteins sequences

This figure shows the level of polymorphism present in the N-terminal regions of Arabidopsis CPR5 and CPR5-like proteins from different plant species. Protein sequences were extracted from the NCBI genome database (see Appendix 11.2 for accession numbers of the proteins) and were aligned using Geneious R6.

4.2.3 IDRs of CPR5 and CPR5-like proteins appear to have INDELs

INDELs have been frequently observed in typical intrinsically disordered regions of intrinsically disordered proteins (Light et al., 2013). Therefore, CPR5 IDRs were expected to have INDELS when compared to their homologues. To test, the amino acid sequences of CPR5-like proteins were compared using Geneious, in order to find out insertion(s) or deletion(s) of residues in AtCPR5 and its homologues. As shown in Figure 4.2, the highly polymorphic N-terminal regions of CPR5-like proteins vary in length. Compared to AtCPR5, the N-terminal regions of some of the CPR5-like protein sequences are shorter, whereas, others are longer than CPR5. For example, the N-terminal region of the CPR5-like proteins from *Abmorella, Elaeis*, and Eucalyptus are longer than the AtCPR5 N-terminus polymorphic region. Whereas, the N-terminal region of the protein sequence from *Beta*, *Brachypodium*, *Cicer*, *Eutrema*, *Fragaria*, *Hordeum*, *Medicago*, *Musa* etc. are shorter than AtCPR5. Similarly, a significant number of CPR5-like proteins, such as, *Brassica*, *Camelina*, *Capsella*, *Citrus*, *Coffea*, *Cucmis*, and *Gossypium* are shown to have N-termini indifferent in length from AtCPR5 (Figure 4.2. To conclude, the length of N-termini of CPR5-like proteins from different species vary, and the gaps or additional residues in AtCPR5 IDRs could be a consequence of INDELs (Light et al., 2013).

4.2.4 AtCPR5 IDRs are annotated to contain MoRFs and disordered binding regions

IDRs or IDPs are known to contain MoRFs or disordered binding regions (Dunker *et al.*, 2015). Therefore, the AtCPR5 IDR was hypothesised to contain MoRFs or disordered binding regions. In order to confirm the presence of such motifs within disordered region, AtCPR5 IDR sequence was further examined using MoRFPred, DisEMBL and ANCHOR predictors. The results show that like typical IDPs, the AtCPR5 IDR is predicted to contain molecular recognition features (MoRFs) or disordered binding regions. For examples, the AtCPR5 protein sequence is predicted to contain 6 regions as putative MoRFs using MoRFPred whereas 8 as disordered binding regions using ANCHOR (Table 4.1). In summary, these results show that like typical IDRs, AtCPR5 IDRs contain also MoRFs and/or disordered binding regions.

4.2.5 AtCPR5 disordered regions annotate to be unfolding and flexible regions

Structurally, intrinsically disordered regions (IDRs) do not fold to form secondary or tertiary structures (Ptitsyn, 1995, Uversky and Ptitsyn, 1996, Dunker and Kriwacki, 2011). Thus, it was hypothesised that CPR5 IDRs would not fold. To test, the AtCPR5 protein sequence was analysed by the FoldIndex *in silico* tool. As shown in Figure 4.3, the regions of the AtCPR5 annotated to be IDRs, are predicted to be un-foldable, in contrast to the remaining parts of the protein, which are proposed to be foldable. In conclusion, AtCPR5 IDRs show resistance to folding.

Taken together, the results shown in this chapter show that AtCPR5 IDRs possess a majority of the defining features of typical IDRs. Thus, the N-terminus of theAtCPR5 protein could act as an intrinsically disordered region and, in turn, allow CPR5 to perform multiple functions.

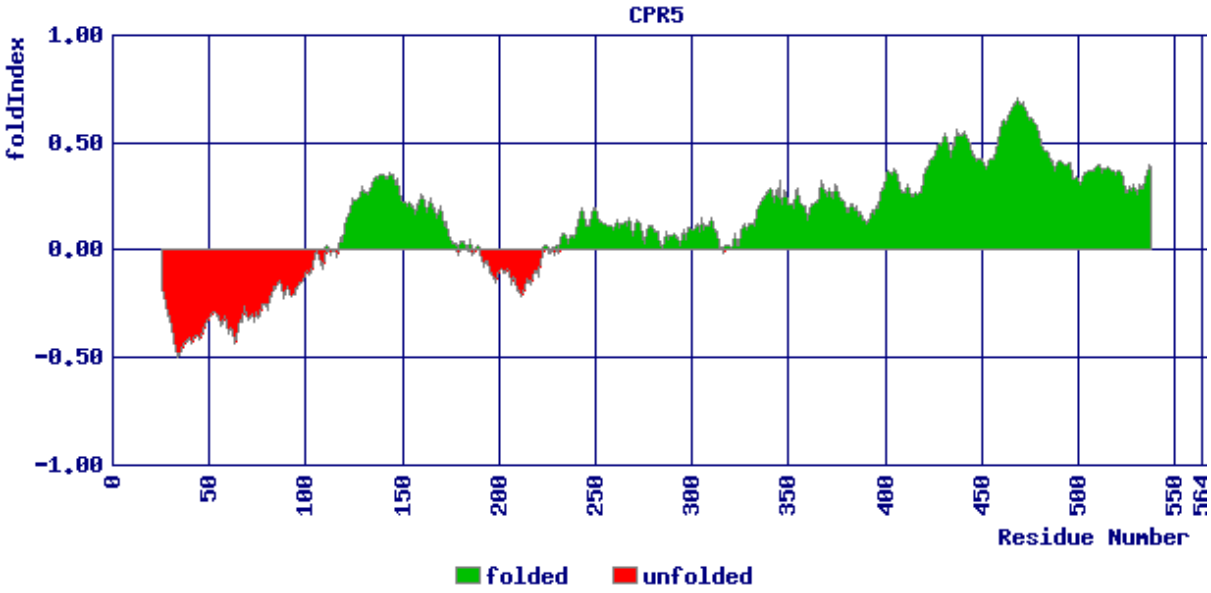


Figure 4.3 Folding and unfolding propensities of regions of CPR5

This figure shows the propensities of parts of CPR5 to be folded or unfolded during secondary structure formation of CPR5. To achieve this figure, CPR5 protein sequence was provided to the online available FoldIndex web server.

4.3 DISCUSSION

4.3.1 AtCPR5 appears to be an intrinsically disordered protein

As shown in the results, the amino acid composition of AtCPR5 IDR and CPR5-like proteins is highly different. The presence of polymorphism in the N-terminal disordered regions of AtCPR5 and CPR5-like proteins is in line with typical IDRs or IDPs (Daughdrill et al., 2007, Nilsson et al., 2011). Additionally, AtCPR5 IDRs are enriched in polar and charged residues, such as, aspartic acid (D), glutamic acid (E), lysine (K), arginine (R), methionine (M), serine (S), and glutamine (Q), compared to structured regions, which are involved in the formation of disordered regions (Andersen, 2011, Theillet et al., 2013). In addition to these residues, tryptophan (W), tyrosine (Y), phenylalanine (F), leucine (L), and proline (P) have also been observed in IDRs (Campen et al., 2008, Brown et al., 2010, Theillet et al., 2013). Moreover, the aforementioned residues of AtCPR5 IDRs are present in a series of short repeats (MELLL, PPSPEP, MMM, KKKK and SSS) as proposed in several studies (Sun et al., 2010a, Andersen, 2011, Theillet et al., 2013). Proline (P) residues of disordered regions generally flank pre-structured motifs (motifs that transiently change from disordered-to-order form such as MoRFs) and are involved in the formation and stability of helices in the disordered regions of IDPs (Lee et al., 2014). Additionally, the number of proline residues in IDRs is 1.4 times higher than in the structured regions (Theillet et al., (2013), which is also consistent with the proline content of AtCPR5 IDRs, in which proline content is 1.5 times higher than in the structured areas. In addition to proline content, the AtCPR5 IDR also contains a number of lysine (K), arginine (R) and serine (S) residues. Similar to typical IDRs or IDPs, the variable length of IDRs from AtCPR5 and its homologues is also consistent and indicative of IDP for AtCPR5. In summary, AtCPR5 is shown to possess many of the defining features of typical IDRs or IDPs, which strongly suggests AtCPR5 to be an IDP.

4.3.2 Having IDP characteristics, AtCPR5 could exert pleiotropic effects

4.3.2.1 AtCPR5 could serve as a hub for multiple interactions and signalling

Molecular recognition features (MoRFs), binding sites or pre-structural motifs (preSM) interact or bind with proteins, RNAs or DNAs (Theillet et al., 2013; Lee et al., 2014). As

mentioned in the results and shown in Table 4.1, AtCPR5 is predicted to contain 6 MoRFs and 8 disordered binding regions, which could facilitate AtCPR5 IDR transit from the disordered-to-ordered form, so that AtCPR5 could gain the desired shape in order to bind with their binding partner(s). Thus, the presence of binding regions or MoRFs within the disordered region(s) of the AtCPR5 protein indicate that AtCPR5 IDRs are able to render different interacting interfaces to interact with multiple partners. The structural flexibility and plasticity allows intrinsically disordered proteins or regions to interact with a broad range of partners, such as, proteins, membranes, nucleic acid and small molecules (Tompa and Csermely, 2004, Uversky et al., 2005, Oldfield et al., 2008). For instance, the p53 protein, which has intrinsically disordered regions, MoRFs and binding regions, regulates more than 150 genes and interacts with over 100 other proteins or peptides (Zhao et al., 2000). The type of intrinsically disordered proteins provides a network of connections or signalling hub for interacting partners (Rual et al., 2005, Stelzl et al., 2005). For example, DELLA and GRAS, two plant-specific intrinsically disordered proteins, which also carry disordered regions in their N-terminus, have been shown to be implicated in multiple interactions (Sun et al., 2014, Sun et al., 2010a, Sun et al., 2010b). Thus, AtCPR5 IDRs could interact with multiple interacting partners in a DELLA or GRAS fashion (Sun et al., 2014).

In conclusion, these studies suggest that AtCPR5 could act as a signalling hub, which, in turn, allows AtCPR5 to interact and take part in various cellular pathways, such as, cell death regulation, senescence, oxidative stress, cell division, cell development and cell expansion (Dunker et al., 2002, Ward et al., 2004, Xie et al., 2007, Jing and Dijkwel, 2008).

4.3.2.2 AtCPR5 IDRs could facilitate AtCPR5 in yielding protein isoforms

Intrinsically disordered regions play an important role in helping the protein to produce isoforms (Buljan et al., 2013, Buljan et al., 2012, Light and Elofsson, 2013), though the exact mechanism of how IDRs facilitate alternative splicing or alternative start codon usage is still obscure. Studies document that the majority of the proteins, which yield variants, are enriched in disordered regions (Haynes and Iakoucheva, 2006, Romero et al., 2006, Van Roey et al., 2012, Weatheritt et al., 2012, Weatheritt and Gibson, 2012, Hsu et al., 2013, Buljan et al., 2013, Niklas et al., 2015). Analyses of AtCPR5 to assess its ability to produce isoforms have

indicated that *AtCPR5* carries putative alternative start codons and RNA stem-loops. Therefore, it could be possible that AtCPR5 could result in protein variants.

Taken together, the discussion sections of Chapter 4 summarise that AtCPR5 IDRs possess characteristics of typical IDRs and could be involved in multiple functions and localisations. Further *in vitro* analyses and characterisation (Chapter 8) of AtCPR5 intrinsic disordered regions could help to understand its roles in the proposed pathways.

Chapter 5 *In silico* characterisation and purification of AtCPR5 protein for structural studies

5.1 INTRODUCTION

In the current study, attempts were made to purify the AtCPR5 protein in order to solve its atomic structure. AtCPR5 is a membrane protein (Kirik et al., 2001) and despite advancements in structural biology, the purification of membrane proteins has proven inherently hard (Baker, 2010, Zhao et al., 2013). Among many factors that could potentially reduce the probability of a transmembrane protein to be expressed, purified and crystallized, the first and most impeding one is the presence of hydrophobic or transmembrane (TM) domains (Carpenter et al., 2008a). While embedded in the membrane, the isolation and purification of membrane proteins has been shown to be very difficult (Caffrey, 2003, Bartesaghi and Subramaniam, 2009, Carpenter et al., 2008b, Baker, 2010). In addition to TM domains, the second undesirable feature for the expression and crystallization of proteins is the presence of disordered region(s), which is proposed to produce inclusion bodies, affect the crystallization ability of proteins and yield poor quality crystals (Dunker et al., 2001, Slabinski et al., 2007).

With the advent of bioinformatics, powerful and high-throughput computer-based tools have enabled structural biologists to analyse and evaluate proteins of interest in order to predict potential challenges for the expression and crystallization of proteins, and hence modify undesirable characteristics, such as, TM domains and IDRs (Zhao et al., 2013). There are a number of online bioinformatics tools, for example, XtalPred (Slabinski et al., 2007), CRSYTALP (Kurgan et al., 2009), ParCrys (Overton et al., 2008), SECRET (Smialowski et al., 2006)), XANNpred (Overton et al., 2011) and JPred (Drozdetskiy et al., 2015). These tools analyse and predict crystallizability, putative structure, undesirable residues or regions of a protein for structural studies, and ultimately propose substitutions in the undesirable region(s) of protein to improve its purification and crystallization.

A variety of purification methods are available to date to purify protein of interest from crude biological extracts (Zhao et al., 2013). Ion-exchange chromatography (IEC) has been used

since 1850 for the isolation of a variety of analytes, including proteins (Fritz and GJerde, 2009, Cummins et al., 2011). IEC facilitates the separation of peptides, proteins and nucleic acids based on their charge (Okada, 1998, Levison, 2003, Stanton, 2004, Ishihara et al., 2007, Bruch et al., 2009). However, this method has not been shown to be ideal for many proteins (Moustafa and Morsi, 2013). Protein purification using different types of affinity tags, such as, hexa-histidine (6xHis), FLAG, Strep-II, HA-tag, c-Myc-tag streptavidin-binding peptide (SBP), calmodulin-binding peptide (CBP), glutathione S-transferase (GST), and maltose-binding protein (MBP) have shown to be economic and easy methods (Zhao et al., 2013). However, these tags and in particular the bigger tags, such as, GST and MBP, were reported to have impacts on protein structure and folding of the tagged proteins, resulting in alteration to the original structure (Waugh, 2005, Arnau et al., 2006, Zhao et al., 2013).

The results shown in the current chapter covers the research efforts carried out to find a version of the CPR5 protein ideal for expression, purification and crystallization. To achieve this, a set of 24 CPR5 protein versions were designed, tagged and expressed in *E. coli* in order to resolve the atomic structure of AtCPR5, using x-ray crystallography. CPR5 is predicted to be carrying 4 or 5 transmembrane domains and disordered region(s). Therefore, parts of the AtCPR5 protein encoding TM domains and disordered regions were truncated in order to improve protein expression, isolation and crystallization of CPR5 recombinant proteins.

5.2 RESULTS

5.2.1 Eliminations of predicted transmembrane domains

As mentioned in Chapter 3, AtCPR5 contains 4 or 5 putative TM domains, the positions of which are shown to be conserved among CPR5-like protein sequences (Figure 3.2G). Additionally, β-strands are predicted between TM domains. Thus, it was hypothesised that constructs designed to eliminate TM domains include the deletion of only putative TM-encoding regions of CPR5, and β-sheets or other regions predicted to be important for CPR5 structure are present in some of the constructs. To delete TM regions and avoid the elimination of any other crucial structural part(s) of CPR5, the AtCPR5 protein sequence was analysed by a number of online *in silico* tools using their default settings. As shown by the results summarised in Table 5.1, 4 or 5 TM domains are predicted in AtCPR5, and the positions of the TM domains as predicted by different tools, appear to be same (section 3.2.5). The boundaries of the last 4 TM domains were predicted by all 6 tools in contrast to the boundary of the first TM domain which was annotated by only 3 programs. Based on information given in Table 5.1, two positions were selected for the elimination of the last four transmembrane (TM) domains. Thus, construct B (Figure 5.1) contains the first 413 amino acids of the AtCPR5 protein and the β-sheet at the end (Figure 5.1). In addition to construct B, another construct (construct C) was also designed in the same area. However, construct C excludes the β-sheet residues (Figure 5.1). In addition, the last 240 amino acids (324-564) from the C-terminus of AtCPR5 were removed in order to exclude all 5 predicted TM domains in construct D (Figure 5.1). Keeping in view the presence of a predicted helix downstream of the first putative TM domain, construct E, which includes the removal of the last 246 amino acids (318-564 aa) was prepared (Figure 5.1). In conclusion, set of constructs (C-E) include the deletion of the putative TM domains as well as β-sheet and helices, whereas, the putative β-sheet and helices are present in construct B.

5.2.2 Elimination of predictied AtCPR5 IDRs

In addition to TM domains, intrinsically disordered regions (IDRs) also impede protein purification and crystallization (Dunker et al., 2001; Slabinski et al., 2007). Since AtCPR5 is predicted to contain IDR, therefore the elimination of IDR was considered indispensable in

order to increase CPR5 expression and crystallization feasibility. Keeping in view the positions of the predicted disordered regions, (Chapter 4), AtCPR5 IDRs were removed in parts. In addition to the eliminations of the C-terminal parts, constructs F and G contain deletions of the first 62 and 91 amino acids, respectively (Figure 5.2). in contrast to F and H, construct H contains a deletion of 91 amino acids from the N-terminus and 222 amino acids from the C-terminus of AtCPR5 in order to exclude both IDR and TM domains. In addition to all of the above discussed constructs (B-G), a full length CPR5 protein (Figure 5.2) was also selected for expression and crystallization. In conclusion, constructs F, G and H are expected to have no IDR as well as TM encoding parts of CPR5 protein.

5.2.3 Eliminations of TM and IDR show no improvement in crystallizability of CPR5 truncated proteins

As mentioned above, TM domains and IDR could potentially impede AtCPR5 protein expression and purification from *E. coli* and eventually crystallization. Therefore, based on *in silic*o information, regions encoding both domains were deleted and it was hypothesised that these eliminations will improve AtCPR5 expression and crystallizability. In order to check the degree of improvement in AtCPR5 expression and crystallizability, the amino acid sequence of all of the constructs were analysed by XtalPred *in silic*o tool. As shown in Table 5.2, the full-length CPR5 protein was predicted to be highly difficult or the least promising for crystallization when analysed through XtalPred. Similarly, none of the truncations in construct B, C, D, or E showed any improvement in their crystallizability (Table 5.2). However, the probability of being crystallized of construct F, G and H, was shown to be slightly improved due to IDR truncations. In summary, constructs F, G and H are proposed to show better crystallizability.

5.2.4 *CPR5* constructs were cloned in pET32 expression vector

To amplify *CPR5* cDNA, total RNA was extracted from 3-week old leaf tissues of *Arabidopsis thaliana* ecotype *Landsberg erecta*. *CPR5* cDNA amplified from the RNA pool using a set of cDNA specific primers, showed the expected size (1.695 kb) on 1% agarose gel (Figure 5.2, lane 2). Whereas, no amplification was observed (Figure 5.2; lane= 1) using an intergenic set of primers. The introduction of mutation(s) in the amplified *CPR5* cDNA was confirmed by

Sanger sequencing, which confirmed that *CPR5* cDNA contains no additions, deletions or substitutions while being amplified (Data not shown). All the constructs shown in Figure 5.1 were amplified using construct-specific primers (Appendix 11.1) and subsequently were cloned into a pET32 expression vector under the T7 inducible promoter and T7 terminator, as mentioned in materials and methods (Sections 2.3 and 2.4).

5.2.5 Integration and orientation of *CPR5* transgenes were confirmed via PCR and restriction analyses

Following successful transformations and ligations of designed transgenes in pET32 (Figure 2.3), it was hypothesised that fragments were integrated in the desired position and orientation. To confirm the proper integration and orientation of the *CPR5* constructs (Figure 5.1) in pET32, each construct of interest was amplified from a pET32 DNA plasmid carrying respective construct, using vector-specific sense (T7 promoter based), and construct-specific antisense primers. As shown in Figure 5.4, migration of the amplified fragments on the 1% agarose gel confirmed expected sizes and the integration of all the designed versions of *CPR5* (Figure 5.1) in the right orientation in pET32 (Figure 5.4; data is shown for only few constructs). In addition to genotyping, the integration and orientation of cloned constructs was also verified by restriction analyses (Figure 5.5). The presence of the EcoR1 restriction site in the *CPR5* gene at positions 60-65 bp (Figure 6.3A) and also in the multiple cloning sites (MCS) of the pET32 vector (Figure 2.3) confirmed the expected pattern and number of DNA bands on the agarose gel for each transgene (Figure 5.5; data is shown for only few constructs). As shown in Figure 5.1, the first 186 nucleotides from the 5`end of the *CPR5* gene were deleted in the F-H constructs, which resulted in the deletion of the EcoRI restriction site as well. Thus, digestion of the pET32 carrying constructs (F-H) with EcoRI, showed a single linear band on the gel (data not shown). To conclude, the amplification of constructs through PCR and restriction digestion analyses of the pET32 vector verified the integration of *CPR5* constructs in the desired orientation in pET32.

Program	TM1	TM2	TM3	TM4	TM5
XtalPred	351-370	413-432	444-466	482-504	525-546
TOPCONS	347-368	411-432	443-464	473-494	542-545
PSIPred	x	413-433	-	-	-
TMHMM	x	413-432	445-467	482-504	525-547
SMART	x	413-432	445-467	482-504	525-547
RHYTHM	347-366	413-431	444-463	482-500	525-544

Table 5.1 Predicted position of TM domains in CPR5 protein

The amino acid sequence of the CPR5 protein was analysed by 6 computer programs in order to make a consensus for the boundary of each transmembrane domain. Numbers shown in the table represent the predicted positions of the CPR5 TM domains, which were annotated by the said computer programs. X = No TM annotated. TM = transmembrane domain.

	CPR5	CPR5_411	CPR5_405	CPR5_324	CPR5_318	CPR5_ 63-324	CPR5_ 98-324	CPR5_ 98-342
EP	5	5	5	5	5	3	3	3
RF	11	11	11	11	11	7	7	7

Table 5.2 XtalPred probabilities for CPR5 recombinant proteins to be crystallized

This table shows the probability of CPR5 recombinant proteins to be crystallized. The numbers in the first row represent the number of CPR5 amino acids analysed by XtalPred e.g. CPR5_411 means the first 411 aa of CPR5 were analysed. Similarly, 63-324 means that amino acids from 63 to 324 were anlaysed. XtalPred calculates crystallizability using two methods; EP = Expert pool crystallizability; RF = Random Forest crystallizability. For EP scores, 1 = promising for crystallization, whereas 5 = difficult; and for RF, 1 = promising for crystallization, whereas 11 = difficult. The numbers after CPR5 in the top row represent the number of amino acids of CPR5 analyzed.

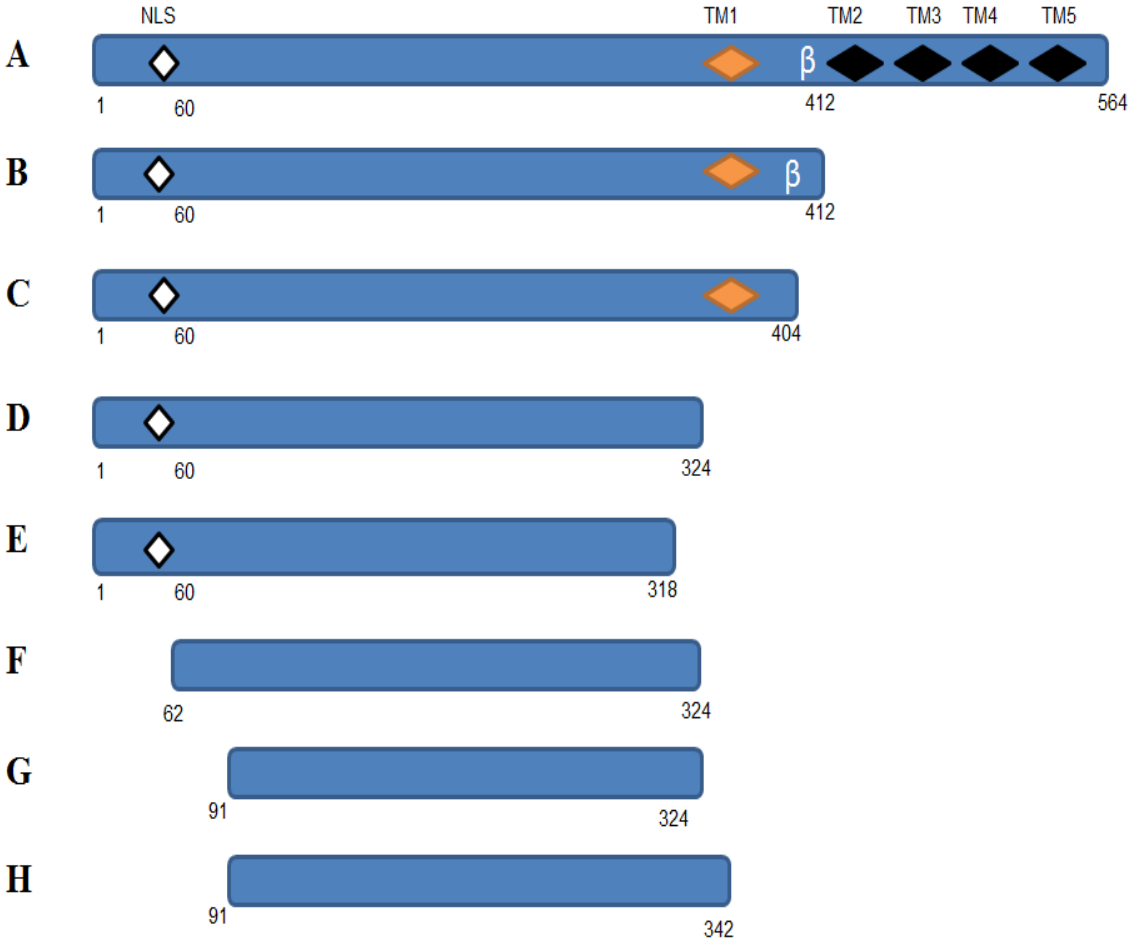


Figure 5.1 Positions of various selected truncations in CPR5 protein

This figure represents the various constructs designed to express the CPR5 protein for crystallization. The letters A-H represent construct names, whereas the numbers represent amino acid numbers (the start and the end) of the CPR5 protein. NLS = nuclear localisation signal; TM = transmembrane domain.

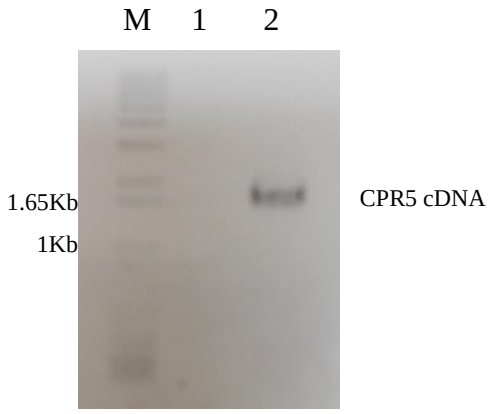


Figure 5.2 Migration of *CPR5* cDNA amplified fragment

This figure shows the migration of double stranded *CPR5* DNA amplified using cDNA-specific set of primers. M = DNA standard (1 Kb plus), 1= negative control and 2= *CPR5* cDNA. The total RNA and *CPR5* amplified fragment, both were migrated on 1 % agarose gel.

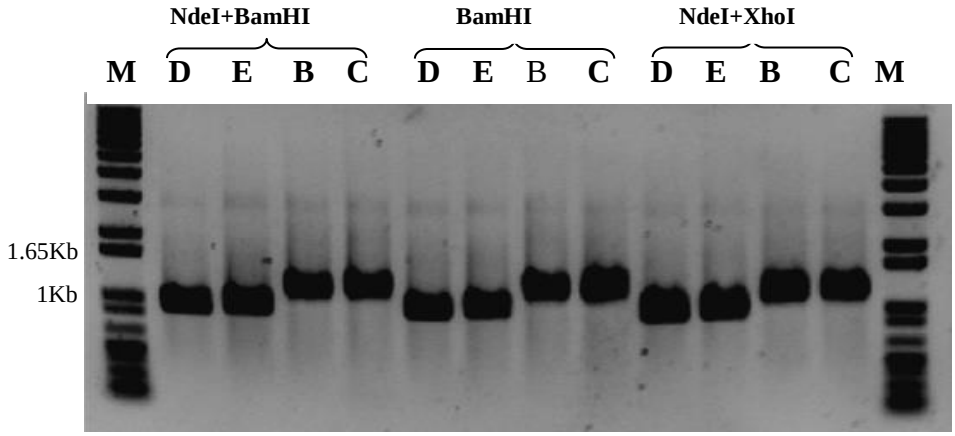


Figure 5.3 Amplification of *CPR5* constructs

This figure shows the migration of amplified fragments of *CPR5* constructs (Figure 5.1) from *CPR5* cDNA. NdeI + BamHI, BamHI & NdeI + XhoI (mentioned at the top) display the names of restriction sites added through primers. All constructs (Figure 5.1) were amplified however only a set of constructs (B, C, D, & E) is shown here. The size of fragments are as follows; construct B = 1233 bp; construct C = 1212 bp; construct D = 972 bp, construct E = 954 bp.

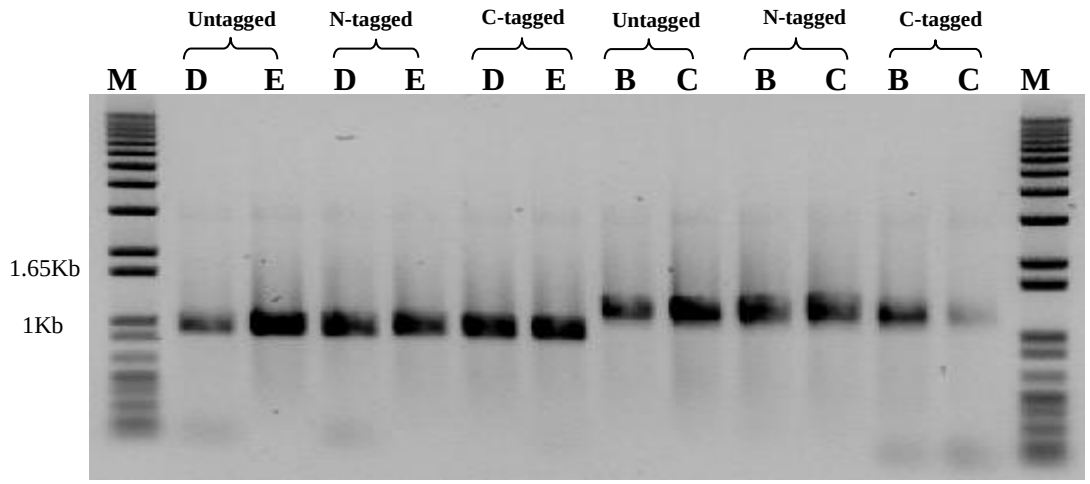


Figure 5.4 Confirmation of integration of constructs in pET32 by PCR

This figure represents the amplification of sets of *CPR5* constructs from the isolated pET32. The products of each construct were amplified through PCR using a sense primer from pET32 (T_7 promoter; Figure 2.3) and antisense primer from within the construct (construct-specific). The amplified fragments were migrated on 1% Agarose gel. Expected sizes are as follows: B = 972; C = 954; D = 1233 & E = 1212 bp. M = DNA marker, and the numbers on the left side of the gel represent the size of the corresponding bands from DNA standard. Constructs F-H were also confirmed in a similar way (data not shown).

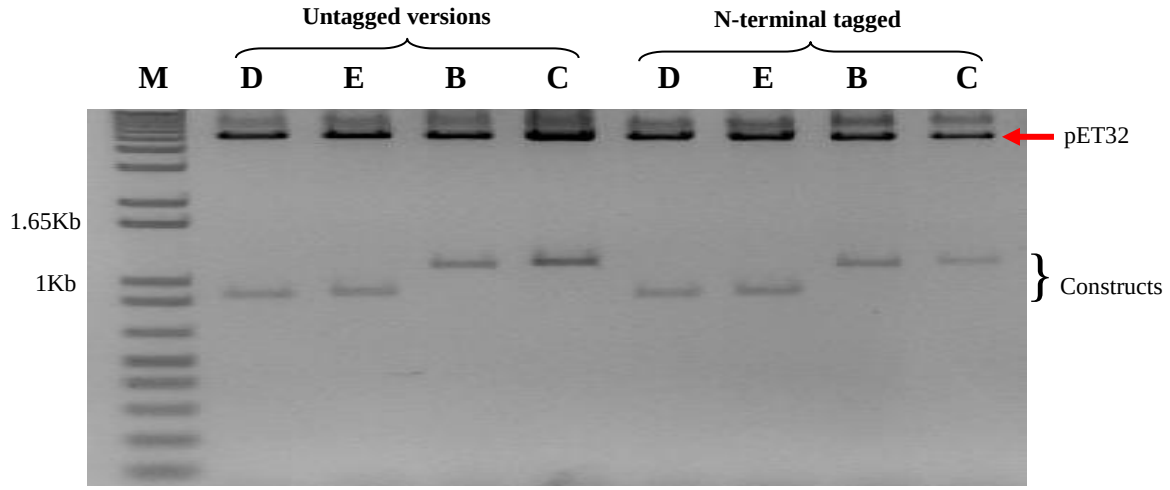


Figure 5.5 Confirmation of integration of constructs in pET32 by restriction analyses

This figure shows the fragmentation pattern of the pET32 vector carrying constructs when digested with EcoR1. Expected sizes are as follows: B = 1187; C =1166; D = 926; E = 908 bp. M= DNA marker, and the numbers on the side of the gel represent the size of the corresponding bands from DNA standard. Similarly, the proper integration of the remaining constructs was also confirmed through restriction analyses.

5.2.6 AtCPR5 recombinant proteins were soluble in Rosetta-gami cells

Following confirmation of proper integration and orientation of constructs in pET32, the next step includes the confirmation of expression and solubilisation of recombinant proteins. Based on the exclusions of disordered and TM regions, CPR5 recombinant proteins were expected to be expressible in the heterologous bacterial expression system and soluble in purification buffers. To test hypothesis, CPR5 transgenes (pET32$^{+constructs}$) were transformed and expressed in *E. coli* Rosetta-gami cells. After the successful transformation of constructs in pET32, *CPR5* transgenes (B-H) were evaluated for their expression and solubility in the heterologous bacterial expression system. In the heterologous expression system, many of the genes have been reported to interfere with the survival of the host organism when such heterologous genes are expressed (Saida et al., 2006, Rosano and Ceccarelli, 2014). In order to check the expression and solubility of *CPR5* transgenes, pET32 carrying constructs of interest, was transformed and expressed in *E. coli* Rosetta-gami cells. A significant level of CPR5 recombinant proteins were found in the soluble fraction as compared to the amount of CPR5 recombinant proteins in the pellet fraction when migrated on SDS-PAGE gel (Figure 5.6). In conclusion, CPR5 recombinant proteins show a significant level of expression and solubility.

5.2.7 Purification of His-tagged AtCPR5 recombinant proteins

5.2.7.1 CPR5 recombinant proteins with C-terminal His-tag were failed to purify

Proteins fused with tags are easier to purify than untagged proteins (Crivat and Taraska, 2012). Thus, CPR5 recombinant proteins fused to His-tag were expected to make their purification easier and faster. To test, each set of constructs was fused with 6x His tag at the N-terminus and C-terminus separately, and 24 transgenes were developed. For example, a set of construct B had the transgene containing no tag, and was fused with the N-terminal and C-terminal tag. In the first attempt, CPR5 recombinant proteins (B and C) containing His-tag at their C-terminus were processed for expression and purification. CPR5 recombinant proteins B and C both showed a significant amount of recombinant protein in the soluble fraction compared to the pellet fraction (Figure 5.6). However, the recombinant proteins B and C failed to be purified or eluted through Ni^{+} column chromatography, as described in materials and methods (Section 2.9.3). Recombinant proteins D and E, in which all five TM domains were excluded, were also

not observed in the final elution fraction (data not shown). In a similar fashion, the expression and solubility of recombinant proteins F, G and H were also checked and these recombinant proteins were also unable to be purified (data not shown). Upon observing no desired proteins in the final elution fractions, the majority of the SDS-PAGE gels on which elution fractions were migrated, were not photographed. In conclusion, the purification of CPR5 recombinant proteins using a small 6x His tag was unsuccessful.

5.2.7.2 Recombinant proteins with N-terminal His-tag also showed no improvement in purification

As CPR5 recombinant proteins B and C contained a putative hydrophobic region (TM) in their C-terminus immediate before the His-tag, therefore, these recombinant proteins could have been misfolded and the His-tag was buried within the hydrophobic region, and hence resulted in His-tag inaccessibility to bind onto the column. Thus, it was assumed that the fusion of His-tag at the N-terminus of the protein would facilitate the purification of CPR5 recombinant proteins. To purify, again transgenes B and C (Figure 5.1) were firstly chosen to be expressed and purified. These recombinant versions (B and C) showed significant levels of solubility (Figure 5.6). However, like the C-terminal tagged versions, these N-terminally tagged recombinant proteins also failed to be purified using the Ni^{+} column chromatography (data not shown). Despite being soluble, purification of the smaller N-terminally tagged versions D, E, F, G and H also failed to be purified (data not shown). In conclusion, the purification of the CPR5 recombinant proteins carrying the N-terminal His-tag was also found to be unsuccessful.

5.2.7.3 Recombinant CPR5 protein fails to bind or binds too strongly onto the column

The unsuccessful attempts for the purification of the CPR5 recombinant versions carrying His-tag at both the N-and-C-termini, raised a few questions; (1) does the purification system (protocol) work well? (2) Where does the protein disappear to despite being present in the supernatant soluble fraction? and (3) Are CPR5 recombinant proteins able to bind onto the Ni^{+} column loosely or strongly? To find out the answers of these questions, firstly bacterial protein fused with His-tag (protocol control) was purified using the working protocol, and this protein was then successfully purified, which confirmed that there is no problem with the purification protocol and handling. Secondly, the crude protein extract was incubated for 1 hr at 4^{0}C to allow the protein to bind with Ni^{+} ions. Subsequently, the sample was processed for purification

and all the elution fractions (washes) were migrated on SDS-PAGE gel. Surprisingly, a significant amount of protein was observed in the first elution fraction when the protein was washed with elution buffer without imidazole to remove non-specific bacterial proteins (data not shown). This observation indicated that it is highly likely that the majority of the protein could not have been able to bind or loosely bind onto the column and hence failed to be purified.

In the next step, after taking the supernatant, the leftover resin in the final step of protein purification (in which recombinant proteins were washed with elution buffer containing 0.1 or 0.5 M imidazole) was incubated with SDS, boiled after adding loading dye and eventually migrated on the SDS-PAGE gel along with the samples from the other elution fractions. The migration of samples from the resins exhibited the presence of a significant amount of protein, indicating that recombinant proteins could have tightly adhered to the resin particles and require an elution buffer with a stronger concentration of imidazole. In order to see the effect of a higher concentration of imidazole, the recombinant proteins were washed with elution buffer containing a gradient range of imidazole (0.1, 0.2, .25, 0.3, 0.35, 0.4, 0.45, and 0.5 M). However, the CPR5 recombinants proteins were unable to be purified. An effort to elute the recombinant proteins with elution buffer containing 1.0 and 1.5 M imidazole also failed to displace the recombinant proteins from the resin (data not shown).

5.2.7.4 Cobalt and Nickel ions show higher binding affinity than copper and magnesium ions

A number of metal ions are used in order to purify His-tagged proteins. It was hypothesised that CPR5 recombinant protein may have a different binding affinities with either of the commonly used metal ions which would be helpful in the selection of a right metal ion for CPR5 purification. To test which metal ion shows a higher binding specificity for CPR5 recombinant protein, 4 types of metal ions (Ni^+, Co^+, Cu^+ and Mg^+) were tried to purify CPR5 recombinant protein B having His-tag at the N or C-terminus (Figure 5.7). The purifications using different metal ions (Ni^+, Co^+, Cu^+ and Mg^+) at different concentrations of imidazole (0.1-1.0 M) displayed similar binding specificities for Ni^+ and Co^+.

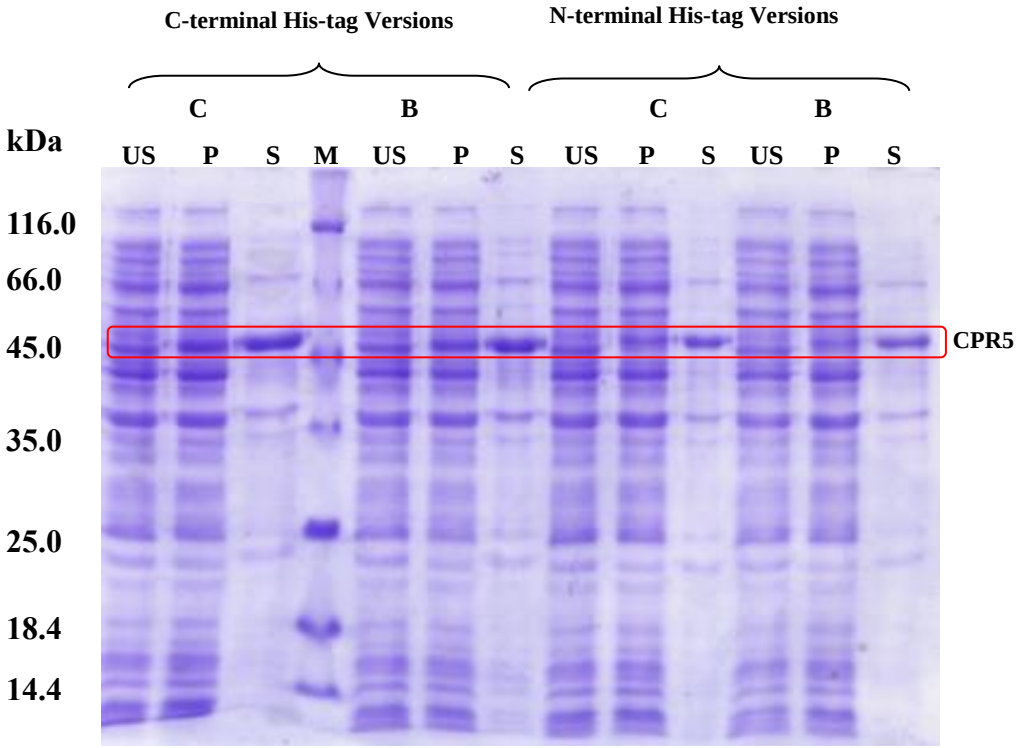


Figure 5.6 Confirmation of expression and solubility of CPR5 tagged recombinant proteins

This figure shows the amount of CPR5 recombinant proteins (B and C transgenes) in the un-induced supernatant (US), pellet (P) and supernatant (S) fractions after being induced with IPTG. The red highlighted rectangular box represents the expected size range for CPR5 recombinant proteins, B and C versions. Supernatants and pellet samples were also extracted from the pET32 empty vector (data not shown). The expression and solubility of the remaining CPR5 recombinant proteins were also confirmed in the same way. US = supernatant from un-induced sample, S = Supernatant from induced sample; P = Pellet dissolved in 8 M urea. M= protein marker and the number shown on the left side of the figure indicate the sizes of the bands of protein marker.

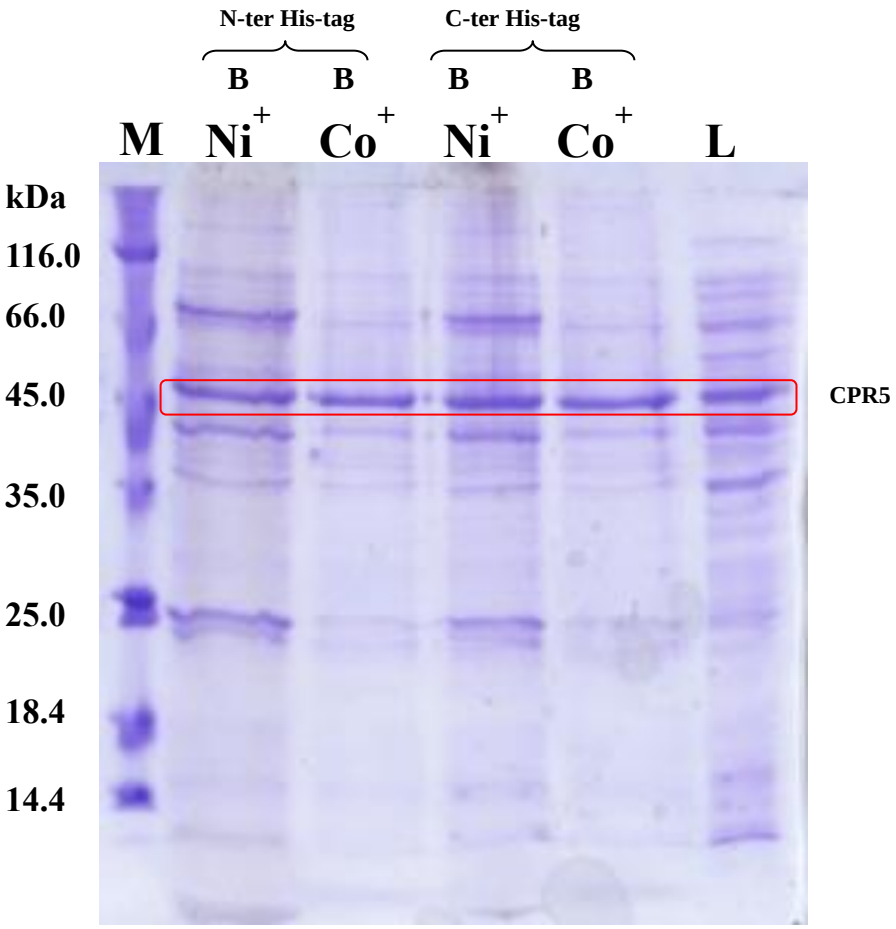


Figure 5.7 Confirmation of binding affinity of CPR5 recombinant protein B

This figure shows the binding affinity of various metal ions with CPR5 recombinant protein B. The red rectangular box shows the expected size of recombinant protein B. Cell lysate from recombinant protein B was incubated with resin pre-charged with the said metal ion for 1 hour. After 1 hour of incubation, the mixture was centrifuged and the supernatant was transferred to a tube. However, the leftover resin (pellet) was incubated with SDS, boiled and migrated on SDS-PAGE gel. M= protein marker and the number shown on the left side of the figure indicate the sizes of the bands of protein marker.

However, protein yields with Ni^+ and Co^+ charged resins were found to be slightly higher compared to the amount of proteins in the incubated resins containing Cu^+ and Mg^+ (Figure 5.7). However, the purification of proteins with either Ni^+ or Co^+ was unsuccessful.

5.2.8 Purification of untagged versions using ion exchange chromatography

Generally, the purification of proteins fused with tags is easier than conventional methods (Wombacher and Cornish, 2011, Crivat and Taraska, 2012). However, in the current study, this method was proven to be unsuccessful for CPR5 recombinant proteins. Thus, it was assumed that the purification of CPR5 recombinant proteins using conventional methods might be useful. Based on amino acid composition and characteristics (instability index and isoelectric point) of each version of CPR5 recombinant proteins, firstly, the longest untagged recombinant protein version (construct B) was selected for purification using ion-exchange chromatography.

Recombinant untagged protein B displayed a substantial level of solubility in the supernatant fraction (data not shown). As shown in Figure 5.8, lysate (L), or the supernatant fraction, is shown to contain two strong bands (named as upper and lower) in the expected size range of the recombinant B protein. In the first attempt, purification of CPR5 untagged recombinant proteins (B-E) through the cation-exchange chromatography protocol, as described in materials and methods (Section 2.9.4), was proven to be unsuccessful. However, purification of the recombinant protein B using the anion-exchange chromatography protocol, was shown to be successful (Figure 5.8). As shown in Figure 5.8, the red rectangle (on the right side of the images) indicates the expected CPR5 recombinant protein obtained as a result of anion-exchange chromatography. In comparison, CPR5 purified bands align with the lower band of two thick bands present in lane 2 (wash) (Figure 5.8). Thus, the presence of two distinct bands in lane 2 and the difference in the position of purified putative CPR5 recombinant protein bands (present in lanes 5, 7 & 9), raised the question as to which band(s) contained recombinant CPR5 protein.

5.2.9 Identification of CPR5 protein by Mass Spectrometry

As shown in Figure 5.8 and mentioned in the above section, there are two strong bands in lane 2, and the purified putative CPR5 recombinant protein is also lower than its expected size.

Therefore, identification of these bands by mass-spectrometry was deemed necessary. Thus, all five bands were analysed using mass spectrometry. For this purpose, both bands from lane 1 (upper and lower bands), lane 2 (wash), and all three bands from lanes 5, 7 and 9 were extracted from the gel and digested with trypsin using the in-gel digestion protocol as mentioned in materials and methods (Section 2.10). The fragmented peptides detected by the mass spectrometer displayed > 39-42% similarity with HYS1 (CPR5) fragments present in the database, thus, all these bands were identified as CPR5 (Figure 5.9). In conclusion, the mass spectrometer confirmed the presence of CPR5 proteins in all of the extracted protein bands.

5.2.10 CPR5 protein was failed to be purified using size exclusion chromatography

Following identification of CPR5 by the mass spectrometer, all the samples representing the abovementioned recombinant versions of CPR5 (Figure 5.8) are shown to still have bacterial proteins despite being purified through anion exchange chromatography. Thus, it was assumed that size exclusion chromatography could help to separate CPR5 recombinant protein. Therefore, all the samples representing identified CPR5 bands were further purified using size exclusion. Based on the expected size, the CPR5 protein was not observed at the expected position in the SEC chromatograph (data not shown). Additionally, none of the separated fractions from the size exclusion chromatography showed any protein when migrated on SDS-PAGE gel (data not shown). In summary, the purification of recombinant protein using size exclusion chromatography was also not successful.

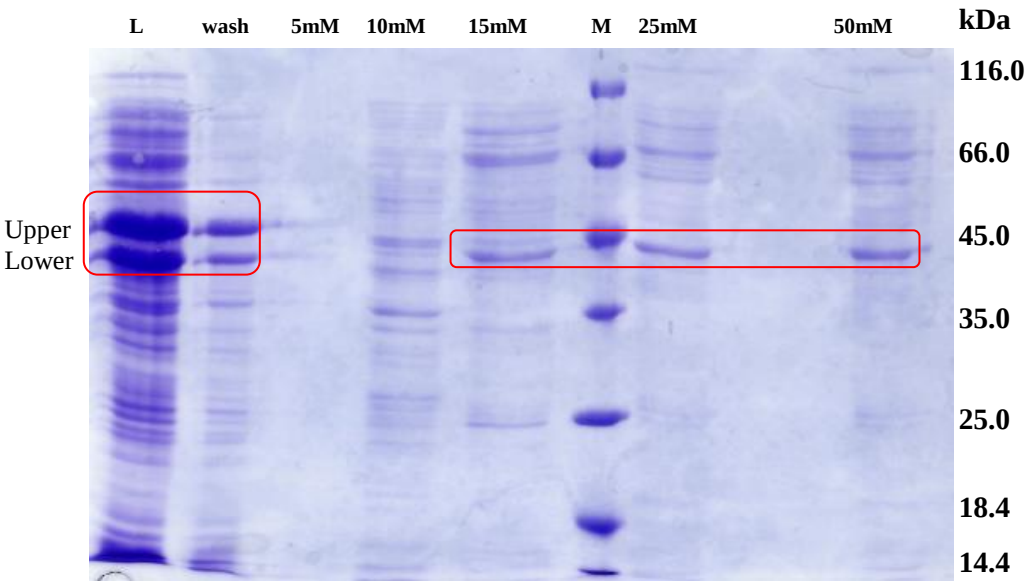


Figure 5.8 Confirmation of expression and purification of untagged version

The lysate fraction (lane: L) of the gel shows the level of expression and presence of the CPR5 recombinant protein B (enclosed in the smaller red rectangle). The longer red rectangular box highlights the amount of CPR5 version B purified using anion exchange chromatography. The number (5-50 mM) represents the gradient concentration of Tris-HCl buffer (pH= 8.0), whereas the lane wash represents the first supernatant collected after washing the protein with loading buffer to remove non-specific bounded proteins. M= protein marker and the number shown on the right side of the figure indicate the sizes of the bands of protein marker.

```
  1 MEALLLPPSP EPQNQITNPA NSKPNHQSGD VHKDETMMMK KKKDTNPSNL
 51 EKRKLKGKKK EIMDNDEASS SYCSTSSTSN SNSTKRVTRV VHRLRNPMRL
101 GMARRSVGER QAEKLAKPLG FSLAAFANMV IARKNAAGQN VYVDDLVEIF
151 ATLVEESLAN VYGNKLGSFA TNFEQTFSST LKILKLTNEC ANPHQSNNND
201 GGSCNLDRST IDGCSDTELF ERETSSATSA YEVMQGSATA TSLMNELALF
251 EETLQLSCVP PRSSAMALTT DERFLKEQTR ANDLKTVEIG LQIRELRCKE
301 TALGLKFESN NLGKAALELD VSKAAFRAEK FKTELEDTRK EEMVTRIMDW
351 LLVSVFSMLA SMVLGVYNFS IKRIEDATSV CDQSEEKSSS WWVPKQVSSI
401 NSGFNTFICR VRVWVQIFFG VLMIIVFTYF LNKRSSGTKQ TMPISFIVLF
451 LGIFCGVSGK LCVDTLGGDG KLWLIVWEVF CLLQFVANVF TLALYGLMFG
501 PINVTQETRS NRCNSMFPYW ARRSVVYVVI LFVLPVINGL LPFATFGEWR
551 DFAMYHLHGG SDYA
```

Figure 5.9 CPR5 identification by Mass-spectrometry

This figure shows the positions of the peptides fragmented and detected by mass-spectrometry. Samples were prepared using in-gel digestion protocol as described in materials and methods (Section 2.10) and submitted in the mass-spec facility operated by A/Prof. Gillian Norris, Massey University, Palmerston North, New Zealand. The sequence of obtained peptides was aligned with the amino acid sequence of CPR5. Red highlights the amino acids, which show residues identified by Mass-spec. As seen, none of the peptides from the deleted region (411-564) of the CPR5 protein was detected by mass-spec.

5.3 DISCUSSION

5.3.1 Potential consequences of transmembrane domains eliminations

Hydrophobic regions of proteins usually reside inside the membrane or bury inside the protein during folding and make the inner core of the protein (Chothia, 1976, Dill, 1990, Honig and Yang, 1995). The presence of hydrophobic or transmembrane regions in a protein has been suggested undesirable for purification and crystallization studies in many ways. For example, hydrophobic regions promote proteins to produce inclusion bodies (Christendat et al., 2000), which effects protein solubility (Slabinski et al., 2007). Transmembrane domains make the isolation of proteins difficult while embedded in a membrane (Caffrey, 2003, Bartesaghi and Subramaniam, 2009, Carpenter et al., 2008a, Carpenter et al., 2008b, Baker, 2010). Transmembrane proteins could fail to produce crystals or produce low quality diffracting crystals (Slabinski *et al.*, 2007). Keeping in view these results, transmembrane regions present in AtCPR5 were eliminated in order to increase purification and crystallization feasibility.

However, the elimination of TM region(s) posed some challenges, the first of which was the determination of the total number of transmembrane regions. As shown in results, the total number of TM domains was selected by making consensus from *in silico* results. Assuming that CPR5 has five TM domains, the purification of CPR5 recombinant proteins B and C could have failed due to the presence of the first putative TM at the end of these two constructs. Having a TM domain, recombinant proteins B and C either may embed in the membrane or fold in order to bury hydrophobic residues, which (first TM) could result in improper protein folding (Vecino et al., 2012). Both cases (membrane embedding or folding) may result in the inaccessibility of His-tag to bind onto the column. Moreover, the first predicted TM region comprises of hydrophobic and hydrophilic residues, which could tend to aggregate due to the presence of hydrophobic residues and result in protein insolubility and/or a poor quality crystal lattice (Slabinski et al., 2007). On the other hand, assuming that CPR5 has four TM domains and the first TM region is an important structural part of the protein, which is essential for CPR5 proper folding, the removal of all five TM domains in D and E proteins could result in the loss of true structure of the CPR5 protein. Therefore, four constructs were developed in order to include and exclude the first TM domain and see any differences in structure.

Similarly, the improper elimination of TM domains without the consideration of the presence of other structural sheets or helices in the vicinity of each TM domain, could also result in improper protein structure. For example, the deletion of a β-sheet in construct C and D or deletion of a helix or coiled-coil region in construct E could result in improper crystallization and structure of the recombinant proteins. Therefore, the inclusion of a β-sheet, helix or coiled-coil regions in constructs B and D respectively, and their exclusions in constructs C and E, was believed to compensate structural losses resulted by the deletion of the β-sheet and coiled-coil domain in these recombinant proteins. In conclusion, a number of *CPR5* constructs were designed in order to achieve the best possible structure of the protein.

5.3.2 Putative impacts of IDRs eliminations on CPR5 structure

In addition to the alpha-helices, beta-strand and transmembrane domains (hydrophobic regions), disordered regions are also important parts of the protein, and are shown to be associated in the signal transduction of many pathways (Dunker et al., 2001). During structural studies, thousands of proteins have been reported to have disordered regions (Romero et al., 1998, Romero et al., 2001, Vucetic et al., 2003). Intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) of proteins are not ideal for protein purification and crystallographic studies. IDPs or IDRs are flexible in nature and unable to form a definite 3-D structure, hence failing to yield informative crystal lattices (Dunker et al., 2001). X-ray crystallography is not a perfect method for IDPs as some of the IDPs undergo a transition of conformation i.e. disordered-to-ordered or unfolded-to-folded, upon binding with their biological partner(s) (Edwards et al., 2000, Dyson and Wright, 2002, Dyson and Wright, 2005). Thus, it could be possible that AtCPR5, which has IDRs, may form poor crystals for x-ray crystallography and yield no structural information. For IDPs, NMR spectroscopy is suggested to be the ideal technique (Montelione et al., 2000, Smialowski et al., 2006) to resolve their structure. Alternatively, the removal of disordered regions may improve the crystallization of truncated proteins, however, their removal may not reflect the true structure (Edwards et al., 2000, Dunker et al., 2001). In summary, assuming that intrinsically disordered regions may affect protein purification and crystallization, and that their exclusion will improve protein expression and crystallization, CPR5 disordered regions were deleted in constructs F, G and H.

5.3.3 Purification of AtCPR5 protein is not trivial

There has been enormous development in the products and methodologies for recombinant protein production and purification in recent years (Pina et al., 2014). Affinity tags based chromatography has made the purification of proteins easier, more economic and high-yielding with fewer steps (Walsh, 2003, Carta and Jungbauer, 2010, Walter and Gottschalk, 2010, Milne, 2011). The current study discusses the purification strategies used to purify AtCPR5 recombinant proteins. However as mentioned in the results, CPR5 recombinant proteins were unable to be purified using the majority of available protein purification technologies.

The purification of proteins tagged with His-tag (6 or 10 histidine residues) has been shown to be successful for many recombinant proteins (Mason et al., 2001, Mason et al., 2002, Walls et al., 2011, Young et al., 2012). His-tag can be fused to either protein termini. However, the purification of a recombinant protein, nonglycosylated human serum transferrins, carrying a N-terminal His-tag resulted in a higher yield compared to the yield of the same protein with a C-terminal His-tag (Mason et al., 2001; Mason et al., 2002). In the current study, CPR5 recombinant proteins B and C, in which the His-tag was fused at their C-terminus, were firstly purified but their purification proved unsuccessful. Since construct B and C have a C-terminal His-tag, which also contains a hydrophobic region (first putative TM domain) immediately before the His-tag, it could be possible that the C-terminal His-tag was inaccessible due to packing or burying of the hydrophobic region. However, the purification of recombinant proteins D and E (having His-tag at their C-terminal, and where hydrophobic region encoding in the first TM domain has also truncated), also failed.

Similarly, the purification of CPR5 recombinant proteins (B, C, D and E) with N-terminal His-tag also failed, which prompted the consideration to delete the CPR5 disordered regions (IDRs). Intrinsically disordered proteins or proteins have longer disordered segments and have a shorter half-life (van der Lee et al., 2014). CPR5 IDRs could be degraded or cleaved off and result in the detachment of the His-tag, hence the protein was unable to bind onto the column. This hypothesis can be further supported by the Mass Spectrometry results, where none of the detected CPR5 fragmented peptides showed any similarity to the peptides belonging to the CPR5 IDRs regions. Thus, the recombinant proteins F, G and H were expected to be purified easily as proposed by XtalPred. However, the purification of the recombinant proteins F, G and

H was also unsuccessful. Although the purification of the untagged construct B through anion-exchange chromatography proved successful and the purified bands were identified as CPR5, the CPR5 recombinant protein was not able to be separated using size exclusion chromatography. Keeping in view the efforts and results, it can be concluded that purification of the CPR5 protein is not trivial.

For future experiments, it should be ensured that CPR5 recombinant protein remains present in the solution at each purification step, which could be probed using a CPR5 antibody. Similarly, protein insolubility represents an impeding factor (Warne et al., 2008, Kramer et al., 2012), so care should be taken that CPR5 recombinant proteins do not precipitate during purification. Moreover, the addition of lipids in solutions could help the recombinant protein to be soluble and stable during and after extraction from cells (Kurisu et al., 2003, Long et al., 2005, Cherezov and Caffrey, 2007, Hilf and Dutzler, 2008). Similarly, eukaryotic proteins could undergo posttranslational modifications, such as, glycosylation and disulphide bonding (Junge et al., 2008), or may be targeted to specific cellular location(s) or membrane(s) and require a native eukaryotic environment (Schnell and Hebert, 2003, Baker, 2010). Therefore, the expression and purification of CPR5 in other available expression systems, such as, yeast (Tornroth-Horsefield et al., 2006, Long et al., 2007, Molina et al., 2007, Pedersen et al., 2007), or insect cell lines (Hiroaki et al., 2006, Rasmussen et al., 2007, Jasti et al., 2007) could help CPR5 protein to be natively expressed and purified.

Taken altogether, the *in silico* studies represent an idea of the challenges associated with the extraction of CPR5 membrane protein. Despite failures, all of the efforts exerted to express and purify the CPR5 recombinant proteins would prove a great resource to consider different expression systems and purification methods, while planning to resolve the CPR5 structure in the future.

Chapter 6 Development of CPR5 mutants based on *in silico* information

6.1 INTRODUCTION

As shown in Chapter 5, the purification of CPR5 recombinant proteins failed, and hence the attempt to elucidate CPR5 function(s) based on the atomic structure of CPR5 was also unsuccessful. Nevertheless, the *in silico* analyses predicted a number of structural motifs, sites and regions in CPR5 (Chapter 3). These predictions provided the basis to formulate different hypotheses and mutate predicted motifs in order to confirm the role of predicted sites, motifs or regions in various pathways and CPR5 functions.

In vitro site-directed mutagenesis is an important technique, which enables researchers to insert, delete or change the nucleotides encoding the important structural motif(s) of a protein (Edelheit et al., 2009). The resulting mutants helped researchers to identify or isolate the function(s) of a gene or protein (Ling and Robinson, 1997, Antikainen and Martin, 2005, Giron and Salto, 2011). The development and functional characterisation of mutants have enormously contributed to the elucidation of functions of many genes or proteins (Sherman, 2005, Giron and Salto, 2011, Reva et al., 2011, Gäbler et al., 2013). Thus, the directed mutagenesis of genes of interest could help to elucidate the structure and function of proteins.

Based on the understanding derived from the *in silico* analyses described in Chapter 3, a set of 25 CPR5 mutants were designed to characterise the role of predicted motifs or regions. This chapter essentially describes the process of selection of the important sites, regions and motifs, and the positions of their truncations. It also includes the transformation of designed constructs i.e. complementation, and the selection of transgenic plants carrying transgenes. In the later part of this chapter, it is discussed as to how the homozygous and stable plant lines of each construct were developed and selected.

6.2 RESULTS

6.2.1 Putative IDRs were deleted to study their roles in CPR5 functions

As mentioned in Chapter 4, the first 276 bp (92 amino acids) of At*CPR5* are annotated as an intrinsically disordered region. To dissect the function(s) of AtCPR5 IDRs, the predicted IDR of CPR5 was further divided into three smaller sections (Figure 6.1A). Based on the positions of the in-frame start codons (methionine residues; 1, 112, 190 & 295 bp) of *CPR5* (Figure 6.1A), truncations in *CPR5* include the deletion of regions between the two start codons of *CPR5*. For example, the first 111 nucleotides were deleted for the first truncation, and the first 189 and 294 base pairs were removed in order to introduce second and third truncations. Compared to the wildtype gene sequence, the positions of the start codons (ATG) and the total length of the *SynCPR5* transgene (as shown in Figures 6.1A and B) are changed due to the addition of a stop codon, HindIII and Sal1 restriction sites in the *SynCPR5* B-transgene. The whole cassette (1.5 kb promotor + 1.710 kb *SynCPR5*) synthesis and introduction of mutations or deletions in the cassette was done by GeneScript.

As shown in Figure 6.1B, a stop codon present immediately after the first in-frame start codon was expected to allow translation initiation of the *Del37CPR5* transgene from the second translation initiation site (127-129 bp), resulting in the exclusion of the first 37 amino acids (Figure 6.1B). The restriction digestion of the *Del37CPR5* transgene with HindIII restriction enzyme is expected to remove 57 amino acids between two HindIII sites as illustrated in Figure 6.1C. Hence, the translation of *Del63CPR5* was expected to be initiated from the third in-frame methionine (205-207 bp of *SynCPR5;* Figure 6.1C). Alternatively, a coding sequence without the first 189 nucleotides (encoding the first 63 amino acids) of *SynCPR5* was fused to a *CPR5* promotor (Figure 6.1E). The restriction digestion of *SynCR5* B template i.e. *Del37CPR5* (Figure 6.1B) with Sal1 restriction enzyme was assumed to delete 80 amino acids of SynCPR5 between two Sal1 restriction sites, allowing translation initiation from the next position (295-297 bp of *SynCPR5*) in the *Del98CPR5* construct (Figure 6.1D). In summary, *Del37CPR5*, *Del63CPR5* and *Del98CPR5* were expected to be translated from their said positions and would allow the testing of function of AtCPR5 IDRs, if any.

6.2.2 Puttive start codons were mutated to study CPR5 isoforms

As shown in Chapter 3, the AtCPR5 N-terminus is predicted to have 3 putative transcription and/or translation start sites: a, b and c as depicted in Figure 6.2A. In order to test the role of predicted start codon sites (b and c) in CPR5 functions, the putative start codons at position b and c were mutated individually as well as collectively. Figure 6.2B represents an illustration of the *metCPR5b* transgene, in which the translation initiation site b is disputed i.e. all of the three methionine (M) residues present at positions 38-40 were converted into glutamine (Q) residues. Thus, the *metCPR5b* transgene would allow translation initiation from position a, yielding a full-length protein and a shorter protein if the putative position c has a role in translation initiation (Figure 6.2B). On the other hand, both of the predicted start codon sites, b & c (Figure 6.2A), were collectively mutated in the *metCPR5bc* transgene (Figure 6.2C), expecting *metCPR5bc* to be translated only from the first in-frame methionine and giving rise to a full-length CPR5 protein. No translation initiation was expected from putative positions b and c in the *metCPR5bc* transgene (Figure 6.2C). To conclude, *metCPR5b* and *metCPR5bc*, would allow the testing of roles of the putative start codon or translation initiation sites in CPR5 functions.

6.2.3 NLS clusters were mutated indiviually as well as collectively

As shown in Chapter 3, AtCPR5 is predicted to contain three regions or clusters (here indicated as regions A, B and C), which may act as a nuclear localisation signal (NLS) the CPR5 protein. It was hypothesised that either of the NLS-encoding clusters could act as the NLS. To test, the predicted NLS clusters were mutated individually as well as in different combinations as summarised in Table 6.1 by GeneScript. Since any of the predicted positively charged residues (lysine or arginine) could act as the NLS, thus, all of the predicted residues were substituted in series. The first class of NLS constructs includes those in which annotated NLS clusters were mutated individually (Table 6.1). Three different constructs were designed in order to disrupt cluster A of the predicted NLS (*nlsCPR5A*). In the first construct of *nlsCPR5A*, the middle two lysine residues were changed into glutamine (Q), whereas the last two lysine residues of the same cluster were mutated in the second construct. The conversions of all lysine residues of cluster A were expected to completely disrupt cluster A of the predicted NLS (Table 6.1). Similarly, substitutions of both lysine and arginine residues (KRK) of cluster B were expected

to mutate cluster B in *nlsCPR5B* plants (Table 6.1). In the second class of NLS constructs, predicted NLS clusters (A, B, C) were mutated in combinations, such as, regions A+B (KKK + KRK), regions B+C (KRK + KKK), and regions A+B+C (KKKK+ KRK + KKK in *nlsCPR5AB*, *nlsCPR5BC* and *nlsCPR5ABC* transgenes, respectively (Table 6.1). In summary, mutations in the *nlsCPR5* transgenes will disrupt putative NLS-encoding clusters and allow the testing of roles of these sites in CPR5 functions.

6.2.4 Casein kinase phosphorylation site 1 and 2 were deleted and mutated

In addition to the NLS, CPR5 is also predicted to contain casein kinase phosphorylation sites, the phosphorylation of which was assumed to facilitate NLS-mediated transportation of CPR5. In order to characterise function(s) of annotated casein kinase phosphorylation sites (CKI and CKII), 7 different constructs were designed. For instance, in order to disrupt the CKII site, putative residues, such as, threonine (T) and serine (A) present at positions 45 and 48 of the CPR5 protein respectively, were changed into alanine (A) and aspartic acid (D) individually (ckII1-1 and ckII1-2) as well as collectively (ckII1-3) as illustrated in Table 6.2. On the other hand, casein kinase phosphorylation site 1 (CKI) is predicted to be composed of 7 serine (S) and 2 threonine (T) residues as shown in Table 6.2. Of CKI deletion constructs, the first construct includes the exclusion of all of the seven serine and two threonine amino acids in the *DelckI-1* transgene (Table 6.2). In contrast to *DelckI-1*, fewer residues of the predicted CKI site were deleted in the *DelckI-2* transgene in order to have the least structural losses compared to *DelCKI-1*. It was assumed that the deletion of 9 residues in *DelckI-1* and 5 residues in *DelckI-2* would result in the loss or abnormal structure of CPR5. Thus, CKI site was also disrupted by mutating the CKI-encoding residues to have the least impact on CPR5 structure or folding in ckI-1 and ckI-1 recombinant protein compared to DelckI-1 and DelckI-2 proteins.

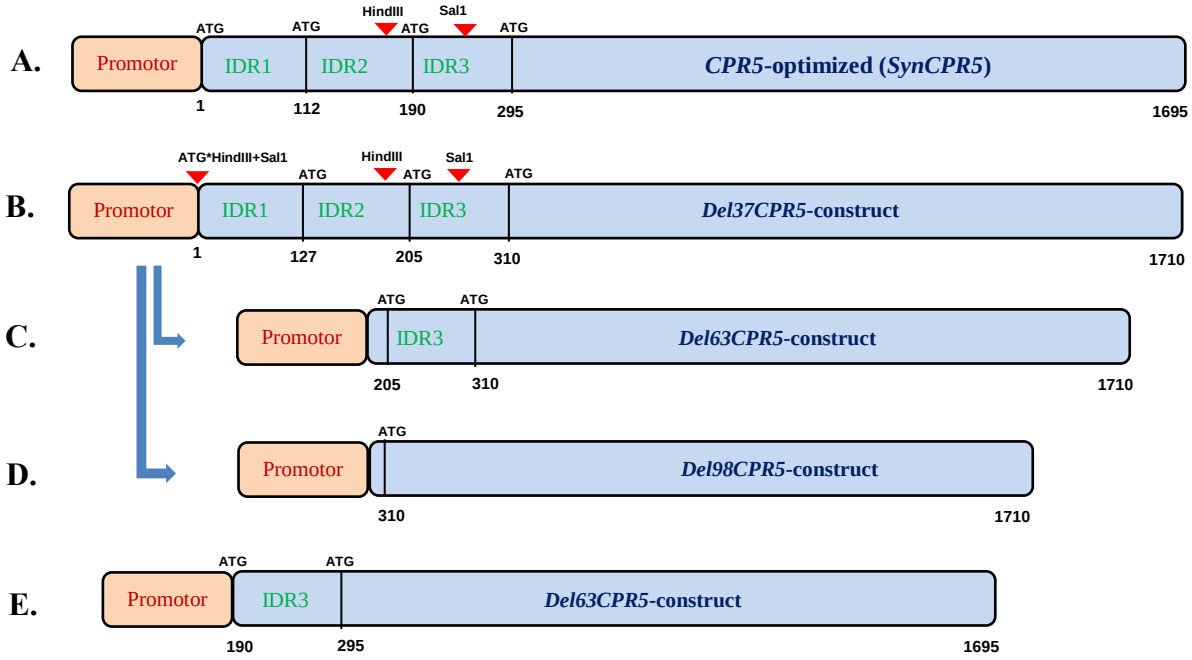


Figure 6.1 Construct designing of *Del37CPR5*, *Del63 CPR5* and *Del98CPR5* transgenes

This figure shows the strategy of how the N-terminus part of the *CPR5* gene was truncated to construct *Del37CPR5*, Del*63CPR5* and Del*98CPR5*. A stop codon (*) along with HindIII and Sal1 restriction sites was introduced immediately downstream of the first start codon (ATG) by GeneScript. The template was excised using HindIII sites and was subsequently self-ligated in order to develop *Del63CPR5*. Likewise, the excision of construct B with Sal1 restriction enzyme and subsequent self-ligation resulted in *Del98CPR5* construct. Promotor = *CPR5* native promotor synthesized and fused with the *SynCPR5* gene by GeneScript; ATG = start codon, IDR = intrinsically disordered region; numbers = position of nucleotides in *SynCPR5* and other transgenes, before and after the inclusion of a stop codon and said restrictions sties; downward arrows (red) = the position of introduced restriction sites, however, the arrows highlighted in blue indicate that the construct is derived from the template construct B. In construct 6.1E, the nucleotide sequence from position 190 bp (ATG)-1710 bp (stop codon) was directly fused with the promotor sequence.

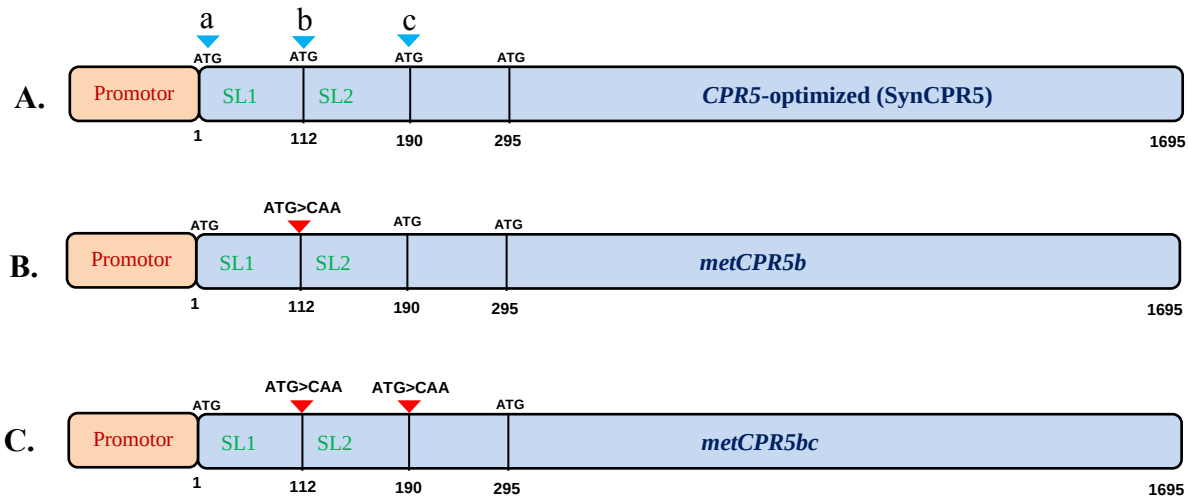


Figure 6.2 Designing of *metCPR5* transgenes

This figure shows the position of RNA secondary structures, such as, stem-loops (SL), and in-frame start codons (ATG) in the *CPR5* synthetic gene. In constructs (6.2B and C), the start codons encoding nucleotides (ATG; methionine) were substituted into CAA (glutamine). Letters a, b and c above the blue arrows in 6.1A indicate the predicted start codon or translation initiation sites as mentioned in Chapter 3.

Construct	Original sequence	Modified sequence	Position(s) disrupted
nlsCPR5A1-1	KKKK	KQQK	A individually
nlsCPR5A1-2	KKKK	KKQQ	A individually
nlsCPR5A1-3	KKKK	QQQQ	A individually
nlsCPR5B	KRK	QQQ	B individually
nlsCPR5AB	KKKK; KRK	QQQQ; QQQ	A & B together
nlsCPR5BC	KRK; KKK	QQQ; QQQ	B & C together
nlsCPR5ABC	KKKK; KRK; KKK	QQQQ; QQQ;QQQ	A, B & C together

Table 6.1 Mutations in putative NLS-encoding residues

Table 6.1 summarises the nature of mutations in the NLS encoding residues present in clusters A, B and C of the CPR5 annotated NLS during site-directed mutagenesis. The last column of the table (position disrupted) depicts the number of clusters mutated i.e. individually (*nlsCPR5A* or *nlsCPR5B*) or collectively (*nlsCPR5AB*, *nlsCPR5BC*, and *nlsCPR5ABC*). Since *nlsCPR5A1-1*, *nlsCPR5A1-2*, *nlsCPR5A1-3* all displayed similar phenotypes, thus *nlsCPR5A1-3* was renamed as *nlsCPR5A* and was used for further characterisation studies.

6.2.5 Putative leucine resiudes of coiled-coil domains were mutated into asparagine

CPR5 is annotated to contain a number of coiled-coil domains or leucine zippers (Section 3.2.7). Since leucine (L) residues of the *heptad* are conserved, thus it was hypothesised that the mutation of putative leucine residues will disrupt conserved *heptad* conformation of CPR5 coiled-coil domain(s). Therefore, leucine residues of the predicted *heptad* were substituted with asparagine (N) residues in order to disrupt coiled-coil *heptad* and to find their effect on CPR5 functioning. As shown in Table 6.3, leucine residues of the first annotated *heptad* at 146 and 153 positions of the CPR5 protein were mutated into asparagine (N) separately (*ccd1CPR5a* and *ccd1CPR5b*) as well as collectively (*ccd1CPR5c*). Similarly, leucine (at position 291) of the second predicted *heptad* was also substituted with asparagine (N) in the *ccd2CPR5* transgene. Additionally, the third transgene *ccd3CPR5* includes the conversion of leucine at 296, 305, and 312 into asparagine (Table 6.3) to disrupt the *heptad*. In conclusion, these constructs are assumed to form no coiled-coil domains and would allow the testing of the roles of selected leucine residues in CPR5 functions.

6.2.6 Putative glycine resiudes of glycine zipper were mutated alanine

In *cpr5-1* (Bowling et al., 1997) and *old1-1* (Jing et al., 2002) mutants, the glycine residues present at the 420 and 459 positions were found to be converted into aspartic acid (D) and serine (S), respectively. As shown in Section 3.2.8, the region of CPR5 containing glycine 459 is predicted to be a glycine motif. In order to test the role of glycine residues in the glycine motif, the glycine residues at the 420 (*Gln420CPR5*) and 459 (*Gln459CPR5*) positions of CPR5 were replaced with alanine (A) (Table 6.3). These genetic changes were expected to rescue *cpr5-2* compromised phenotypes if the predicted glycine residue(s) has no role in the annotated glycine motif. Otherwise, *GlnCPR5* transgenic plants will show *cpr5-2* like phenotypes if glycine residues are an essential or structural part of the putative glycine motif. Additionally, a closer view of the putative CPR5 glycine motifs revealed that there are two other glycine residues present at positions 452 and 456 in the vicinity of glycine 459 (Table 6.3). Therefore, both the glycine molecules at positions 452 and 456 were collectively substituted with alanine (Table 6.3) in order to see the effect of conversion of these amino acids on CPR5 functioning. In summary, mutations in glycine residues of the predicted glycine motifs are expected to result in the loss of glycine motifs.

6.2.7 CPR5 transgenes were complemented into *cpr5-2* plants

Since a number of phenotypes and pathways are shown to be affected in *cpr5-2* plants (Jing and Dijkwel, 2008), it was hypothesised that the *SynCPR5* transgene (being identical to the wildtype protein; Figure 6.3) will complement compromised phenotypes (aberrant trichomes, spontaneous lesions and early leaf yellowing) of *cpr5-2* plants. Similarly, the disruption of the predicted motif or site will display wildtype-like phenotypes if the said motif or site has no role in the execution of that phenotype and *vice versa*. Otherwise, complemented plants will show *cpr5-2* like phenotypes. To test, *SynCPR5* and all the *CPR5* transgenes were complemented into 5-6 week old *cpr5-2* plants via *Agrobacterium*-mediated T-DNA transformations as instructed in (Zhang et al., 2007) studies. The BAR gene (a bacterial gene derived from *Streptomyces hygroscopicus*) present in the pGreen0229 plasmid will allow the confirmation of successful transformation and integration of the selected transgene in *cpr5-2* plants when sprayed with BASTA. Thus, only positive transformants are expected to grow and survive when T1 plants are sprayed with BASTA. On average, 112 positive transformants per construct were obtained, which resulted in 2800 transgenic lines (Table 6.4). The positive transgenic lines were further screened for homozygosity and stable transformation up to the fifth generation. Finally, homozygous lines of each transgene were genotyped by PCR in order to confirm the presence of the said transgene in the respective transgenic line using *cpr5-2*, *CPR5*, *SynCPR5*, and each transgene specific set of primers (Section 2.13.4).

Construct	Original sequence	Modified sequence	Position disrupted
***ckIICPR5* lines**			
ckII1-1	DTNPSNL	DANPSNL	individually
ckII1-2	DTNPSNL	DTNPDNL	individually
ckII1-3	DTNPSNL	DANPDNL	together
***ckICPR5* lines**			
DelckI-1	SSSYCSTSSTS	---YC-------	Deletion of residues
DelckI-2	SSSYCSTSSTS	S--YC---STS	Deletion of residues
ckI-1	SSSYCSTSSTS	SSSYCSAASAA	Substitution of residues
ckI-2	SSSYCSTSSTS	SSSFCSTSSTS	Individual residue

Table 6.2 Designing of CK phosphorylation site mutants.

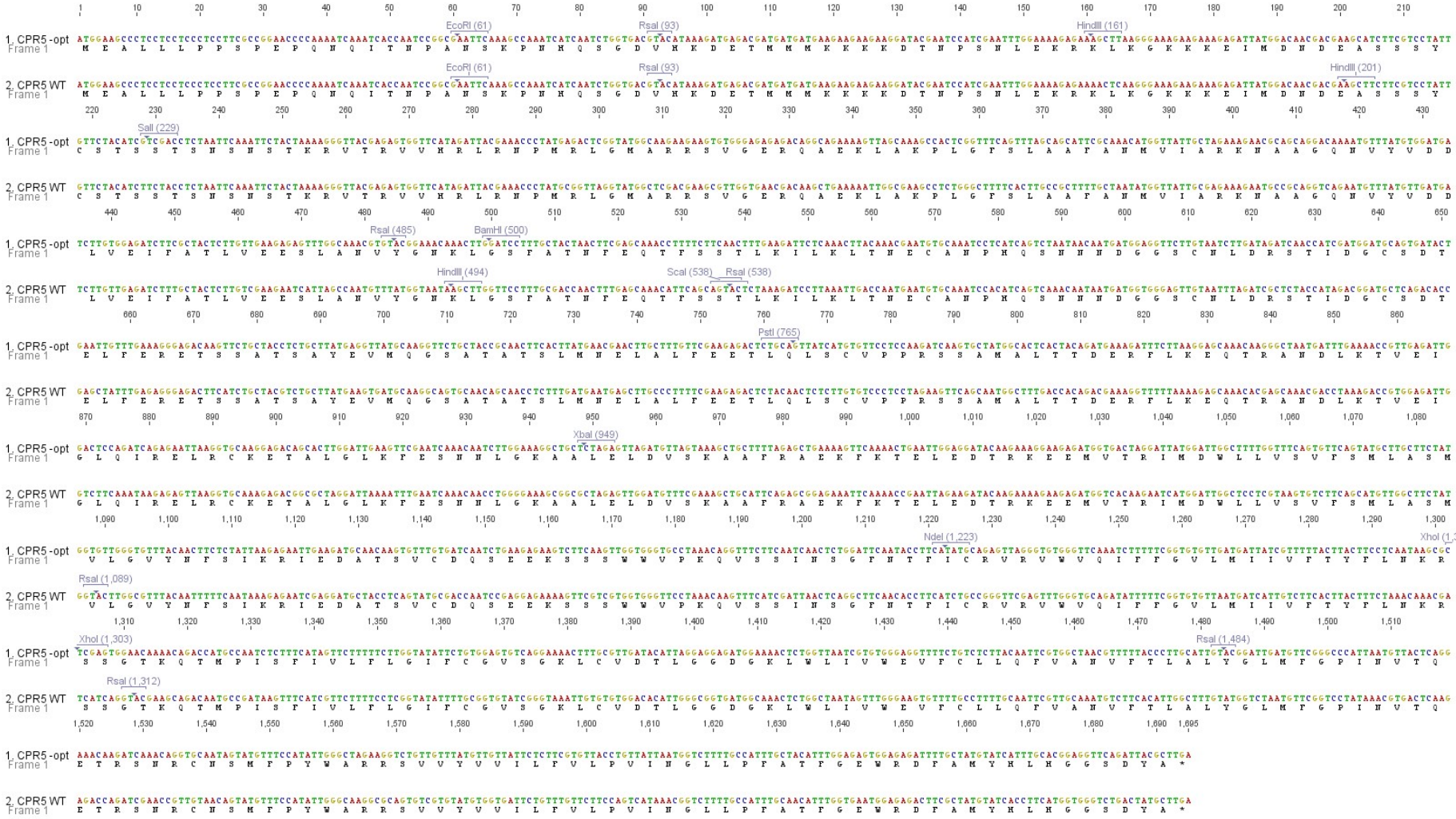
Table 6.2 summarises the residues predicted as CKI and CKII phosphorylation sites (underlined) and residues with which the predicted residues of CK sites were mutated (as underlined in the third column) through site-directed mutagenesis. Deletions of residues are denoted by dotted lines. The column named "Position disrupted" depicts if residues were mutated, and individually or collectively.

Construct	Original sequence	Modified sequence	Position disrupted
Leucine zipper mutants			
ccd1CPR5a	DDLVEIFATLVE	DDNVEIFATNVE	Individually
ccd1CPR5b	DDLVEIFATLVE	DDNVEIFATNVE	Individually
ccd1CPR5c	DDLVEIFATLVE	DDNVEIFATNVE	Collectively
ccd2CPR5	EIGLQIR	EIGNQIR	individually
ccd3CPR5	ALGLKFESNNLGKA	ANGNKFESNNNGKA	Collectively
Glycine zipper mutants			
Gln420CPR5	QIFFGVLM	QIFFAVLM	individually
Gln459CPR5	FLGIFCGVSGKLC	FLGIFCGVSAKLC	individually
Gln452CPR5	FLGIFCGVSGKLC	FLAIFCGVSGKLC	individually
Gln456CPR5	FLGIFCGVSGKLC	FLGIFCAVSGKLC	Individually

Table 6.3 Designing of leucine and glycine zipper mutants

Table 6.3 summarises the putative leucine and glycine residues (underlined) predicted to be the part of leucine and glycine zippers, respectively, and residues with which glycine or leucine residues were replaced (column 3). The column named "Position disrupted" depicts if residues were mutated, individually or collectively. Since all the coiled-coil (*ccd1CPR5a*, *ccd1CPR5b* and *ccd1CPR5c*) transgenic plants displayed similar phenotypes, therefore, *ccd1CPR5c* was renamed as *ccd1CPR5* and was used for further analyses.

A.


EcoRI (61)
RsaI (93)
HindIII (161)
EcoRI (61)
RsaI (93)
HindIII (201)
SalI (229)
RsaI (485)
BamHI (500)
HindIII (494)
ScaI (538)
RsaI (538)
PstI (765)
XbaI (949)
NdeI (1,223)
XhoI (1,3
RsaI (1,089)
XhoI (1,303)
RsaI (1,484)
RsaI (1,312)
1. CPR5-opt Frame 1
2. CPR5 WT Frame 1

B.

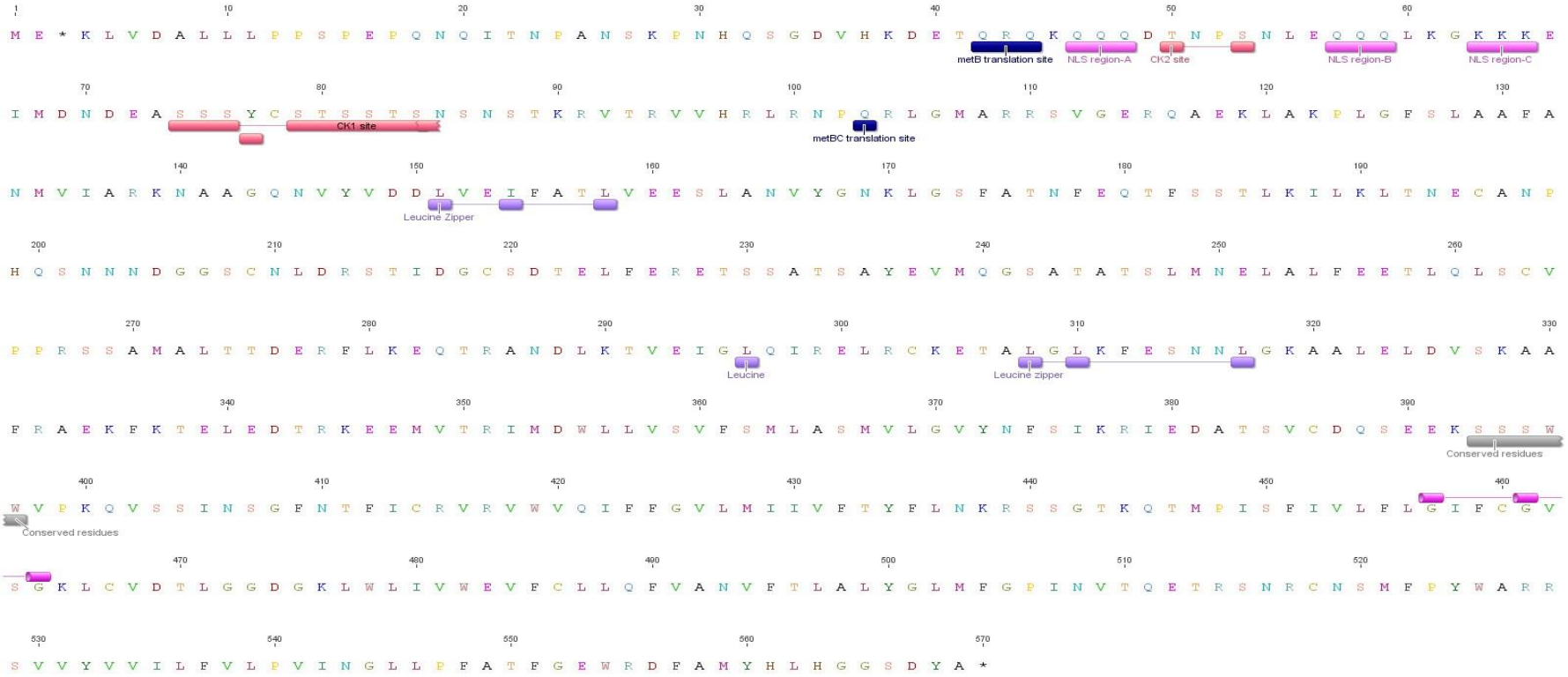


Figure 6.3 Similarities and differences between *CPR5* wildtype and synthetic gene and positions of structural sites

Figure 6.3A displays the similarities and differences of nucleotides between the *CPR5* wildtype and synthetic gene (*SynCPR5*) whereas the positions of various important predicted structural elements are also summarised in Figure 6.3B. This figure was prepared using Geneious R6 (Biomatters, Ltd).

Sr. No	Construct	Total # of transformants	Transformation efficiency %
1	*SynCPR5*	40	0.32
2	*Del37CPR5*	59	0.47
3	*Del63CPR5*	171	1.37
4	*metCPR5b*	55	0.44
5	*metCPR5bc*	71	0.57
6	*nlsCPR5A1-1*	48	0.38
7	*nlsCPR5A1-2*	49	0.39
8	*nlsCPR5A1-3*	53	0.42
9	*nlsCPR5B*	45	0.36
10	*nlsCPR5AB*	244	1.95
11	*nlsCPR5BC*	289	2.13
12	*nlsCPR5ABC*	302	2.42
13	*ckIICPR5*	39	0.31
14	*DelckCPR5I-1*	73	0.58
15	*DelckCPR5I-2*	93	0.74
16	*ckI-1CPR5*	83	0.66
17	*ckI-2CPR5*	67	0.54
18	*ckI-3CPR5*	37	0.30
19	*ccdCPR51-1*	74	0.59
20	*ccdCPR51-2*	63	0.50
21	*ccdCPR51-3*	77	0.62
22	*Gln420CPR5*	141	1.13
23	*Gln459CPR5*	203	1.62
24	*Gln452CPR5*	233	1.86
25	*Gln456CPR5*	192	1.54
	Average	**2800/25= 116**	**= 0.89**

Table 6.4 List of *SynCPR5* transgenes and their transformation efficiency

This table shows the total number of transformants obtained for each transgene, and the number of independent lines of each transgene which showed wildtype-like trichomes, no HR-like lesions and early leaf senescence. This table also documents the total number of transformants obtained as a result of the transformation of all the transgenes and the overall transformation efficiency.

6.3 DISCUSSION

6.3.1 Mutagenesis studies are expected to identify roles of various putative motifs of *CPR5*

The *in silico* studies predicted a number of structural motifs and sites, which could potentially be important for CPR5 function(s). The current chapter essentially describes the basis of how the predicted motifs of CPR5 were mutated in order to study their role in CPR5 functioning. For example, either of the basic positively charged residues of the NLS cluster, such as, lysine and arginine could act as the NLS. Therefore, seven different transgenes were developed in order to cover the majoirty of the putative residues for their role in nuclear import, individually as well as collectively. The substitutions of lysine (K) with residues, such as, glycine (Takei et al., 1999, Xiao et al., 2012), glutamine (Xiao et al., 2012), asparagine (Xaio *et al.*, 2000), and alanine (Zhang et al., 2000, Sloss et al., 2005, Shin and Reich, 2013), abolished the nuclear accumulation of the mutant proteins. Likewise, the conversions of the NLS in encoding the amino acid arginine into cysteine, serine (Takei et al., 1999), and alanine (Sloss et al., 2005, Shin and Reich, 2013), also resulted in the disruption of the NLS, and protein failed to localise in the nucleus, in contrast to the wildtype protein (Xiao et al., 2000). On the other hand, the substitution of lysine and arginine with glutamic acid (E) failed to prevent nuclear localisation (Xaio et al., 2000). As shown in results, putative NLS residues, lysine and arginine into glutamine (Q), were converted into alanine and expected to abolish CPR5 localisation if the NLS cluster(s) has a role in CPR5 localisation.

In addition to the NLS, all of the nine residues, such as, serine (S), threonine (T) and tyrosine (Y) predicted as casein kinase phosphorylation sites, were removed in order to completely disrupt the putative CKI phosphorylation sites. Alternatively, the CKI and CKII putative residues, serine and threonine, were substituted with alanine (A) on the basis of reported studies. For example, the conversion of CKI residues into alanine (A) successfully abolished the CKI site in numerous studies (Zhang et al., 2000, Blattner et al., 2002, Okamura et al., 2004, Bustos et al., 2005, Zhao et al., 2010a, Loughery et al., 2014). Additionally, the tyrosine

(Y) residue of CPR5 predicted CKI site, was substituted with phenylalanine. Therefore, keeping in view these studies, alanine substitutions will disrupt phosphorylation in *ckICPR5* and *cKIICPR5* lines.

In order to prevent translation from the putative alternative translation initiation sites, the predicted amino acid residues (methionine), were altered to either stop codons, or alanine, during the characterisation of DNA ligase 1 isoforms (Sunderland et al., 2004, Sunderland et al., 2006). Whereas, putative methionine residues were converted into isoleucine and leucine in (Chabregas et al., 2003) and (Christensen et al., 2005) studies respectively. In contrast to the aforementioned previous reports, the putative alternative translation initiation encoding residues of the CPR5 protein were mutated to glutamine (Q). In summary, glutamine would abolish translation initiation from the said sites.

As mentioned in Chapter 4, CPR5 is predicted to contain a number of coiled-coil domains, and the position of leucine residues in coiled-coil domains is crucial for *heptad* formation. In order to disrupt *heptad* formation, leucine was converted into proline in several studies (Harbury et al., 1993, Cheng et al., 2001, Lupas and Gruber, 2005, Tanaka et al., 2007, Ciani et al., 2010). The conversion of leucine (L) into alanine (A) has not always abolished *heptad* formation, since no difference was observed in the helix stability when leucine at both positions, *a* and *d*, of the leucine *heptad* was replaced with alanine (Zhou et al., 1992). In the current study, leucine (L) residues of the predicted coiled-coil domains were changed into asparagine (N) and were expected to abolish *heptad* formation.

Glycine zippers or motifs are an integral part of the transmembrane regions of proteins, and mutations in glycine residues of glycine zipper are deleterious (Kim et al., 2005). Kim and co-workers have shown that valine (V), leucine (L), Isoleucine (I), serine (S), and threonine (T) are not good substitutes of glycine for glycine zipper disruptions. The same study proposed alanine as one of the good substitutes to abolish helix formation. However, Harmeier (Harmeier et al., 2009) found that the conversion of glycine into alanine actually promoted glycine helices dimerization, whereas, the replacement of glycine with leucine disrupted the packing of the glycine-zipper. Recently, Venco and co-workers (Venco et al., 2015) have found that glycine

substitution with serine abolishes the dimerization of glycine zippers. The putative glycine residues present in the second transmembrane region of the CPR5 protein were changed into alanine and are expected to abolish the glycine zipper.

In summary, all the introduced mutations for the disruption of the selected regions or motifs are expected to identify their role in CPR5 functioning, if any. The transgenic lines of each transgene were expected to result in the execution of *cpr5*-like phenotypes or wildtype-like phenotypes depending on the implication or role of the said motif or site.

6.3.2 *CPR5* transgenes were introduced into *cpr5-2* plants using floral dipping

In addition to the biolistic method, *Agrobacterium tumefaciens*-mediated transformation is one of the successful techniques for the gene transfer between *Agrobacterium* and plant(s) to develop transgenic plants (Birch, 1997, Hansen and Wright, 1999, Bent, 2000, Weigel and Glazebrook, 2002, Gelvin, 2003, Zhang et al., 2006). Transgenes can also be transformed by virus and microinjection (Bent, 2000, Van Loock et al., 2010). The *Agrobacterium*-mediated floral dip method of transformation is efficient, fast and relatively easy but involves a few tedious steps, such as, large volumes of cultures and the dipping process (Sheludko, 2008, Li et al., 2009, Tsuda et al., 2012). Additionally, the floral dipping method avoids the use of tissue culture, which requires aseptic conditions and different types of media preparations (Clough and Bent, 1998, Harrison et al., 2006, Logemann et al., 2006, Van Loock et al., 2010). It is possible that a successful transformation method for one species could fail for other species or even among the ecotypes of a same species (Wroblewski et al., 2005). For instance, the *Agrobacterium*-mediated transformation of Ws-0 has been shown easier than for Col-0 (Wroblewski et al., 2005, Tsuda et al., 2012). On the other hand, Col-0 showed 10 fold higher efficiency than L*er*-0 with respect to *Agrobacterium*-mediated transformation (Desfeux et al., 2000).

Transformation efficiency depends on different factors, such as, plant growth stages, growth conditions, *Agrobacterium* strains and plant immunity level (Tsuda et al., 2012). Plant immunity, i.e. the ability to recognise and defend against pathogens, plays an important role in the determination of transformation efficiency, for instance, plants having upregulated levels

of plant defence hormones, such as, salicylic acid, ethylene and other defence-related genes, which restrict *Agrobacterium*-mediated transformation (Yuan et al., 2007, Anand et al., 2008, Nonaka et al., 2008, Lee et al., 2009, Tsuda et al., 2012). Similarly, *Agrobacterium* infiltrations during *Agrobacterium*-mediated transformations could also trigger plant immune response and result in the restriction of *Agrobacterium* entry into the plant cell (Citovsky et al., 2007, Gelvin, 2010, Pitzschke and Hirt, 2010). For example, mutants defective in pathogen effectors recognition were found to be more amenable to *Agrobacterium*-mediated transformation than the wildtype (Zipfel et al., 2006), and the co-transformation of a pathogen effector, such as, *AvrPto* to dampen host resistance, has dramatically enhanced *Agrobacterium*-mediated transformation efficiency in wildtype plants (Tsuda et al., 2012).

cpr5 alleles are available in both Columbia (Col-0) and *Landsberg erecta* (Ler-0) backgrounds. Additionally, *cpr5* plants show upregulated levels of *PR* genes and resistance to *PstDC3000* (Bowling et al., 1997). Based on the abovementioned transformation challenges, *cpr5-2* (Bowling et al., 1997) in Col-0 background was chosen for *Agrobacterium*-mediated floral-dip transformation, instead of *old1* Ler-0 background (Jing et al., 2002), due to the fact that Col-0 has higher transformation efficiency than Ler-0 (Desfeux et al., 2000). Moreover, lower transformation efficiency was expected due to induction of resistance in *cpr5* plants. However, despite having upregulated levels of resistance in *cpr5-2* plants, hundreds of transgenic lines with an average number of 116 transformants per transgene were obtained in this study.

Taken together, the development of set of 25 transgenes and 2800 independent lines of each transgenes are expected to prove a great resource for AtCPR5 functional characterisation.

Chapter 7 AtCPR5 regulates balance between plant growth and disease resistance

ABSTRACT:

A mutation in *CONSTITUTIVE EXPRESSER of PATHOGENESIS-RELATED GENES 5* (*CPR5*) is shown to exert multiple effects on plant phenotypes, such as, aberrant trichomes, HR-like lesions, reduced ploidy levels, and increased oxidative stress. Additionally, *cpr5-2* plants were shown to be resistant to bacterial as well as fungal pathogens. Despite CPR5 implication in several pathways, little is known about the mechanism of how CPR5 participates in different pathways. In the present study, a *CPR5* cDNA was synthesised (*SynCPR5*) and fused to the *CPR5* native promoter. Subsequently, the whole cassette (*CPR5*::*SynCPR5*) was transformed into *cpr5-2* plants in order to complement c*pr5*-compromised phenotypes. The results showed that the native promoter resulted in variable *SynCPR5* expression levels, which helped to isolate some CPR5 functions or phenotypes separate from other functions or phenotypes. Of 40 independent transgenic lines, the majority of *SynCPR5* lines displayed wildtype-like phenotypes (called complemented lines), whereas, a few lines displayed *cpr5*-like phenotypes (non-complemented). The physiological characterisation of the *SynCPR5* complemented lines showed that, compared to the wildtype, some of these lines have higher *SynCPR5* transcript abundance than others. Further quantification of leaf size, trichome number and disease resistance showed that *SynCPR5* lines, which have transcript abundance lower or equal to wildtype levels, had smaller leaves, a less number of trichomes and enhanced resistance compared to *SynCPR5* lines, which had higher transcript abundance than the wildtype. *SynCPR5* lines, which had higher transcript abundance than the wildtype, displayed wildtype-like phenotypes. Intriguingly, the majority of *SynCPR5* lines showed increased levels of *pathogenesis-related genes1* and *5* (*PR1* and *PR5*) as compared to the wildtype. On the other hand, *metCPR5B* and *metCPR5BC* lines in which putative start codons were mutated, exhibited vigorous growth and *metCPR5* transgenic plants showed larger leaves, bigger epidermal cells,

and an increased number of trichomes. Contrary to *SynCPR5* transgenic plants, *metCPR5B* and *metCPR5BC* plants conferred extreme susceptibility to *P. syringae* when compared to the wildtype. Despite having higher transcript abundance, none of the *SynCPR5* lines displayed bigger leaves and enhanced susceptibility, and ectopic leaf size and disease resistance shown by the *metCPR5* lines suggests the role of the putative start codon encoding region in maintaining the balance between growth and resistance. Thus, this finding allowed me to propose that CPR5 regulates the balance between plant growth and disease resistance in *Arabidopsis*, which is a novel role of AtCPR5.

7.1 INTRODUCTION

Being sessile, plants have developed a number of sophisticated mechanisms in order to escape or combat pathogens. When attacked by pathogens, plants activate local and systemic resistance. Activation of resistance generally involves growth-related costs, which affect plant development. AtCPR5 is proposed to be a master regulator of different pathways, as mutation(s) in *CPR5* (*cpr5*) exert pleiotropic effects on a number of pathways (Jing and Dijkwel, 2008). Of *cpr5* functions and phenotypes, stunted growth and aberrant trichomes are two of the conspicuous *cpr5* phenotypes. Additionally, At*CPR5* is shown to negatively regulate resistance (Boch et al., 1998; Wang et al., 2014) and leaf senescence (Jing et al., 2007, Jing et al., 2002).

As shown by several reports (Bowling et al., 1997; Boch et al., 1998; Jing et al., 2007), resistance is acquired by *cpr5* at the cost of reduced growth, which indicates a disruption in the balance between growth and immunity in *cpr5* plants. Thus, AtCPR5 somehow maintains the balance or trade-off between growth and resistance. The trade-off between growth and plant immunity is generally shown to be mediated through manipulations in the synthesis and signalling networks of plant hormones (Lozano-Durán and Zipfel, 2015). As shown by Jing et al., (2007), *cpr5* is defective in the perception and signalling of plant hormones, such as, SA, ABA, JA and ET, thus, it could be possible that an imbalance in growth and resistance in *cpr5* is a consequence of defects in the aforementioned plant hormone signalling. AtCPR5 may also influence the biosynthesis or signalling of either SA, JA, or ET hormones. This prompted me to investigate the role of AtCPR5 in the regulation of the trade-off between growth and plant resistance, if any.

The current chapter covers the characterisation of two types of At*CPR5* transgenic lines, *SynCPR5* and *metCPR5*. *SynCPR5* is a synthetic version of At*CPR5*, which contains several genetic changes introduced during codon optimization. However, SynCPR5 is identical to wildtype AtCPR5 with respect to amino acid composition. On the other hand, *metCPR5* involves substitutions of amino acids in annotated start codons in *SynCPR5* background. The complementation of *SynCPR5* has resulted in variable *SynCPR5* transcript abundance, leaf size,

and *PR1* and *PDF1.2* levels. Despite having higher transcript abundance than the wildtype, none of the *SynCPR5* lines show better growth than the wildtype. In contrast to *SynCPR5*, *metCPR5* show higher expression levels and bigger leaves but enhanced susceptibility as compared to the wildtype. Based on these evidences, AtCPR5 is proposed to be involved in the regulation of balance between growth and immunity.

7.2 RESULTS

7.2.1 *CPR5* native promoter is able to drive *SynCPR5* expression

In the past, a number of studies (Jing et al., 2007; Gao et al., 2011; Perazza et al., 2011) have used the overexpression *Cauli Mosaic Virus 35S* promoter, which resulted in the ectopic expression of CPR5 in the cell. Therefore, a *CPR5* native promoter was chosen to drive *SynCPR5* (synthetic *CPR5* gene) in order to obtain a range of expression of the *SynCPR5* transgene due to the position effect (Peach and Velten, 1991, Gelvin, 2003). The levels of *SynCPR5* were expected to be lower than the wildtype, wildtype-like or higher than the wildtype in the obtained *SynCPR5* transgenic lines. It was hypothesised that 1.5 kb long *CPR5* native promoter is sufficient to drive variable expression of the *SynCPR5* transgene.

To test this hypothesis, a fragment of 1.5 kb, downstream of the *CPR5* gene (At5G64930) was amplified and fused with the *SynCPR5* gene. Subsequently, the whole cassette was transformed into 4 week old *cpr5-2* plants through *Agrobacterium*-mediated transformation. Relative transcript abundance of *synCPR5*, *CPR5* (wildtype) and *cpr5-2* genes was quantified using *synCPR5*, *CPR5* and *cpr5-2* gene-specific sets of primers through real-time qRT-PCR, as described in materials and methods (Section 2.2.3). It is important to note that all three sets of primers were designed on the same region of *CPR5* mRNA in order to minimize difference in expression due to change in position of mRNA and primer efficiency. The C_T or C_q values were normalized with two housekeeping genes in order to measure relative levels of expression (Section 2.2.3).

Following successful *SynCPR5* transformation in *cpr5-2* plants, 40 independent *SynCPR5* complemented lines were obtained. However, initially a few representative lines were analysed for qRT-PCR transcriptomic analyses. As shown in Figure 7.1A, the relative transcript abundance of the *SynCPR5* transgene in *SynCPR5* independent transgenic lines, is shown to be variable when compared to the wildtype. For example, *SynCPR5L2* and *SynCPR5L3* display *SynCPR5* transcript abundance higher than the wildtype, whereas, the transcript abundance of *SynCPR5L5* was shown to be less than the wildtype. Despite having higher transcripts than the

wildtype, the transcript abundances of *SynCPR5L4* were statistically insignificantly different from the wildtype. The expression levels of *SynCPR5L4* and *SynCPR5L6* were less than the wildtype. However, the difference was found to be insignificant from the wildtype (Figure 7.1A). Based on transcript abundance, *SynCPR5L1*, *SynCPR5L2* and *SynCPR5L3* were placed in one group, named *SynCPR5G1*, whereas, *SynCPR5L4*, *SynCPR5L5* and *SynCPR5L6* were placed in a second group called *SynCPR5G2*. Overall, the majority of the *SynCPR5G1* lines show higher expression levels than the *SynCPR5G2* lines. To summarise, these results confirm that 1.5 kb *CPR5* native promoter was able to drive *SynCPR5* expression in a wide range.

7.2.2 *SynCPR*5 complements aberrant trichomes, lesions and early leaf senescence

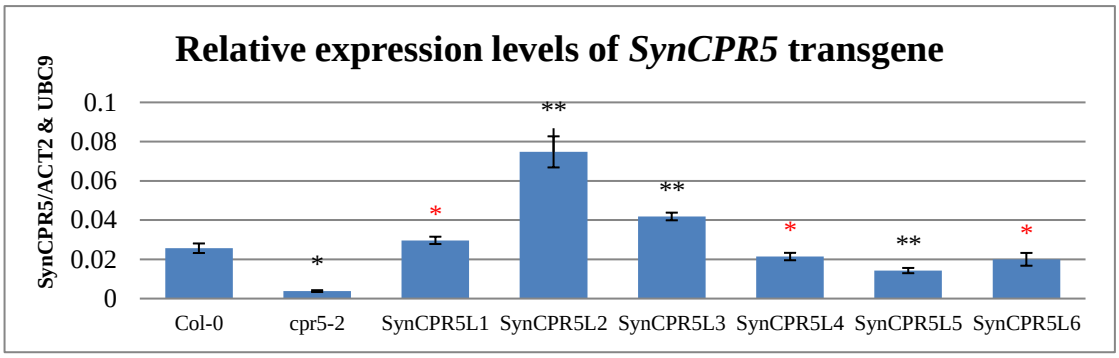
In addition to the native promoter, the synthetic *CPR5* gene (*SynCPR5*) was used in the current study. SynCPR5 protein is identical to the wildtype CPR5 protein with respect to amino acid composition. However, the *SynCPR5* gene differs from the wildtype *CPR5* gene in nucleotide composition, due to the introduction of nucleotide changes during codon optimization of the *SynCPR5*. Having a similar protein sequence (amino acid composition), the *SynCPR5* transgene is expected to complement the compromised phenotypes of *cpr5-2* plants in a wildtype fashion. To test this hypothesis, *SynCPR5* transgenic lines were morphologically characterised in order to see the restoration of aberrant phenotypes, such as, wildtype-like trichomes, no early leaf yellowing and HR-like lesions.

The morphological investigations of *SynCPR5* plants depict that the majority (80%; 32 lines out of 40 lines) of the *SynCPR5* complemented lines had leaves without HR-like lesions or early yellowing compared to *cpr5-2* (Figure 7.1B). Regardless of variations in total number, trichomes with all four types of appendages were present on the majority of the *SynCPR5* lines (Figure 7.1C). Compared to the majority (80%) of *SynCPR5* lines, the leaves on the remaining 20% (8 lines) of the *SynCPR5* complemented lines had aberrant trichomes, HR-like lesions or early leaf yellowing, similar to *cpr5-2* plants. In summary, these results confirm that the *SynCPR5* transgene managed to restore aberrant trichomes, HR-like lesions and early leaf yellowing in a wildtype fashion, on the majority of the *synCPR5* transgenic plants.

7.2.3 Leaf size and number of trichomes on *SynCPR5* plants is consistent with their *SynCPR5* expression levels

As shown in Section 7.1, *SynCPR5* transgenic lines displayed a range of *SynCPR5* expression levels, which prompted me to investigate the effect of *SynCPR5* differential expression on leaf size and trichome number. Based on *SynCPR5* differential expression, it was hypothesised that the leaf size and trichome number of *SynCPR5* lines will vary according to their transcript abundance. To test this hypothesis, the leaf area and number of trichomes from *SynCPR5* lines were quantified. Based on their leaf size and number of trichomes, *SynCPR5* lines could be classified into two major groups. As shown in Figure 7.2, the third and fourth rosette leaf pairs on the majority of lines from *SynCPR5G1* were shown to be indifferent in size from the wildtype, except leaf 3 from *SynCPR5L1*. Similarly, wildtype-like trichomes (carrying all four types of appendages) were shown to be present on the majority of *SynCPR5G1* lines (Figure 7.2). On the other hand, the third and fourth rosette leaf pairs from *SynCPR5G2* lines were intermediate in size compared to the wildtype and *cpr5-2*. In addition to leaf area, *SynCPR5G1* lines displayed more numbers of trichomes with three appendages, whereas, the *SynCPR5G2* lines showed an intermediate number of 3-branched trichomes when compared to the wildtype and *cpr5-2*. The number of trichomes carrying four appendages was shown to be less on *SynCPR5G2*, whereas, the number of the same type of trichomes from *SynCPR5G1* lines was also less but the difference was significantly indifferent when compared to the wildtype. In summary, leaf area, trichome numbers and transcript abundance trends in *synCPR5* transgenic lines establish a positive correlation between leaf size and trichome number with *SynCPR5* mRNA levels.

A.



B.

C.

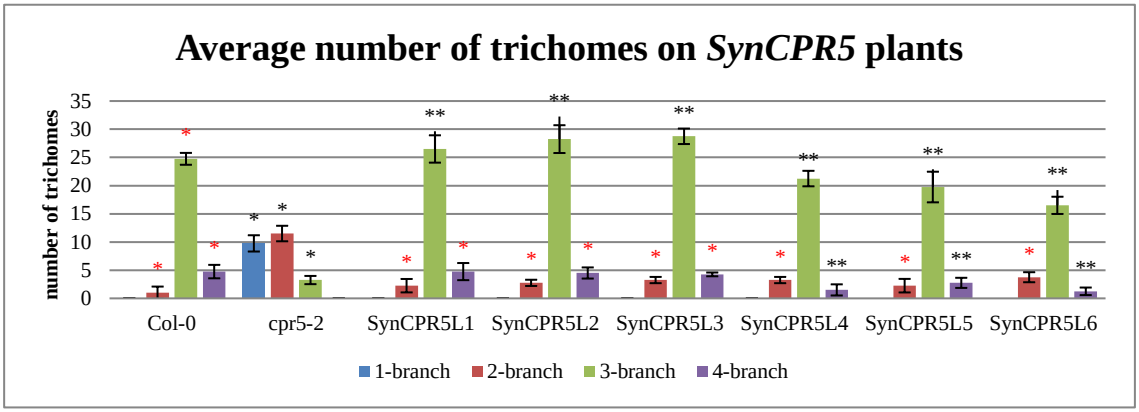


Figure 7.1 *SynCPR5* complementation, transcript and trichomes quantification

Figure 7.1A shows the quantification of transcript abundance of *SynCPR5*, Col-0 and *cpr5-2* using *SynCPR5*, Col-0 and *cpr5-2* specific sets of primers. Three biological and three technical replicates were included and values were normalized to the average expression values of *ACT2* and *UBC9* housekeeping genes. Figure 7.1B displays the complementation and rescue of *cpr5-2* plants with the *SynCPR5* transgene. Figure 7.1C shows the average number of trichomes on leaf discs of 0.68cm diameter, from the leaves of five different plants of each individual line. The bars represent the standard error and asterisks indicate the level of statistical significance at $p < 0.05$ (Student's t-test. *= significant from Col-0; * = significant from *cpr5-2*; ** = significant from Col-0 and *cpr5-2*. See Appendixes 11.3.1 and 11.3.2 for further t-test values and comparisons.

7.2.4 *SynCPR5* fails to complement upregulated levels of *PR1* in *SynCPR5* lines

Several semi-quantitative studies (Boch et al., 1998; Bowling et al., 1997; Clarke et al., 2000; Jing et al., 2007) report that *cpr5-2* mutants have upregulated levels of defence-related genes, such as, *pathogenesis related gene1* (*PRI*). Being identical to the wildtype CPR5 protein, SynCPR5 protein was expected to restore *PR1* levels in a wildtype manner. To check *PR1* levels, total RNA was extracted from 31-day old whole *SynCPR5* plants, and cDNAs were prepared as mentioned in materials and methods (Section 2.2.2). Analyses were carried out to record transcript abundance of *PR1* mRNA amplified using real-time qRT-PCR. Like leaf size and transcript abundance, *PR1* levels were also shown to vary among *SynCPR5* transgenic lines. In contrast to the hypothesis, all the *SynCPR5* transgenic lines displayed increased levels of the *PR1* gene compared to the wildtype (Figure 7.3). Additionally, the *PR1* levels or trends of the *SynCPR5* lines are not consistent with their respective transcript abundances. In conclusion, *SynCPR5* plants carry upregulated *PR1* levels and there is no correlation between *PR1* and *SynCPR5* transcript abundances.

7.2.5 *SynCPR5* and *cpr5-2* plants both show reduced *PDF1.2* levels

As shown in Section 7.2.4, none of the *SynCPR5* transgenic plants managed to be rescued for *PR1* levels, so it was hypothesised that *SynCPR5* transgenic plants will have upregulated levels of *PDF1.2* as reported previously (Bowling et al., 1997, Clarke et al., 2000). To test this, relative levels of *PDF1.2* transcripts were quantified using real-time qPCR as described in materials and methods (Section 2.2.3). In contrast to my prediction and previous reports (Bowling et al., 1997, Clarke et al., 2000), surprisingly, the *cpr5-2* plants exhibited lower *PDF1.2* transcript abundance compared to the wildtype (Figure 7.4). *PDF1.2* transcript levels were found to be lower in the majority of the *SynCPR5* transgenic plants (except *SynCPR5L4*) compared to *cpr5-2* (Figure 7.4). Furthermore, *PDF1.2* transcript abundance was shown to vary among *SynCPR5G1* and *SynCPR5G2* lines. In summary, these results are inconsistent with my hypothesis, since *PDF1.2* transcript abundance was found to be lower in *cpr5-2* and *SynCPR5* lines compared to the wildtype.

7.2.6 *SynCPR5* regulate resistance independent of *SynCPR5* and *PR1* levels

As reported previously, CPR5 (wild-type Col-0) plants are susceptible to the bacterial pathogen *Pseudomonas syringae* pv *DC3000* (*PstDC3000*), compared to *cpr5-2* mutant plants, which restrict *PstDC3000* growth (Bowling et al., 1997; Wang et al., 2014). Since all of the *SynCPR5* transgenic lines displayed upregulated levels of *PR1* (Figure 7.3), it is thus hypothesised that *SynCPR5* transgenic plants will confer resistance to *PstDC3000*, according to their *PR1* levels. To test this hypothesis, four-week-old *SynCPR5* transgenic plants were infected with *PstDC3000*.

The resistance level was scored by counting the number of colony forming units (cfu) as mentioned in materials and methods (Section 2.14). Despite having higher *SynCPR5* and *PR1* transcript abundance (Figures 7.1A and 7.3), *synCPR5G1* transgenic lines displayed susceptibility insignificantly different from the wildtype (Figure 7.5). Contrary to *SynCPR5G1* lines, *SynCPR5G2* lines displayed resistance to *PstDC3000* bacterial infection (Figure 7.5), regardless of having *SynCPR5* transcript abundance equal to or lower than the wildtype (Figure 7.1A). Furthermore, the exhibition of resistance by *SynCPR5G2* transgenic lines is in line with *PR1* levels. These results confirm that *SynCPR5G1* lines display susceptibility, whereas, *SynCPR5G2* show resistance. Additionally, the level of resistance partly depends on *SynCPR5* expression levels, but the execution of resistance seems to be independent of *PR1* levels.

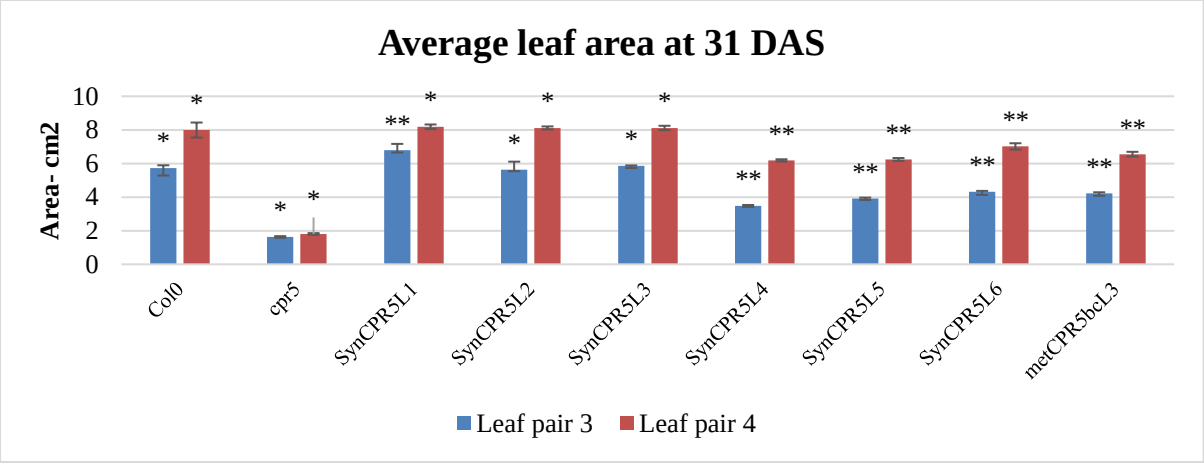


Figure 7.2 Mean leaf area of third and fourth rosette leaf pairs

This figure shows the mean leaf areas of third and fourth rosette leaf pairs at 31 DAS. The leaves were plucked from 3-5 different plants of the same line, and areas were recorded using a leaf area machine. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant from Col-0; * = significant from *cpr5-2*; ** = significant from Col-0 and *cpr5-2*. See Appendixes 11.3.3 and 11.3.4 for further t-test values and comparisons.

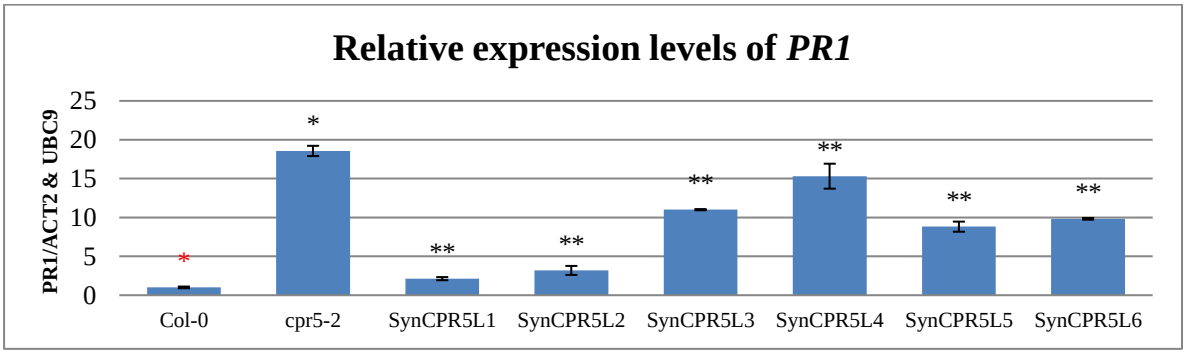


Figure 7.3 Relative expression levels of *PR1* in *SynCPR5* plants

Total RNA was extracted from the whole plants, and relative levels of *PR1* mRNA were measured. Three biological and three technical replicates were included, and values were normalized to the average expression values of *ACT2* and *UBC9* housekeeping genes. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant from Col-0; * = significant from *cpr5-2*; ** = significant from Col-0 and *cpr5-2*. See Appendix 11.3.5 for further t-test values and comparisons.

7.2.7 *SynCPR5* plants show wildtype-like ploidy levels

In the current study, the *SynCPR5* transgenic plants show a range of leaf size, which could be the consequence of difference in ploidy levels in *SynCPR5* lines. Thus, investigation of ploidy levels may reveal the role of ploidy level in the execution of wildtype-like or intermediate-sized leaves. Keeping in view leaf size, *SynCPR5G1* were hypothesised to have wildtype-like ploidy levels, whereas, *SynCPR52* may have reduced ploidy levels compared to the wildtype.

To test this, third and fourth leaves were harvested from 24-day (24 DAS) old *SynCPR5* plants. Nuclei were extracted from leaves by chopping leaves in a chopping solution, and were subsequently stained with propidium iodide, as mentioned in materials and methods (Section 2.16). Nuclei were analysed using a flow-cytometer (Partec, USA) that counts nuclei based on nuclei sizes, and displays the number of total nuclei of similar sizes in the form of peaks (Figure 7.6A). 15000 events were recorded and analysed, and the percentage of nuclei from each peak was calculated using a peak analyses plugin of Flow-max software. Average percent values from the three biological replicates were plotted in the form of graphs (Figure 7.6B).

At 24 DAS, all of the *SynCPR5G1* lines represent wildtype-like levels of 16C nuclei (Figure 7.6B). In contrast to *SynCPR5G1* lines, the number of 16C nuclei from the majority of the *SynCPR5G2* lines was found to be intermediate, when compared to the wildtype and *cpr5-2* (Figure 7.6B). Among *SynCPR5G2* lines, *SynCPR5L4* had a 16C nuclei population insignificant from the wildtype, which is consistent with *SynCPR5L4* expression levels (Figure 7.1). In addition to 16C, the majority of the *SynCPR5* lines displayed similar levels of 2C and 8C nuclei, except the *SynCPR5L2* line, which had more 2C nuclei than the wildtype (Figure 7.6B). Compared to the wildtype, the average number of 4C nuclei in the majority of *SynCPR5* lines was found to be indifferent from the wildtype, except *SynCPR5L2* and *SynCPR5L5*, which were shown to have more 4C nuclei than the wildtype (Figure 7.6B). In summary, the overall ploidy level trends of *SynCPR5* lines, are in line with the hypothesis.

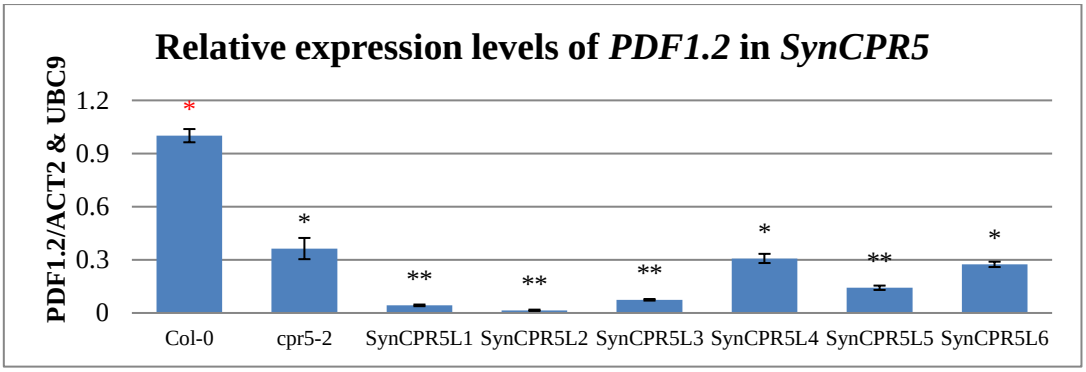


Figure 7.4 Relative expression levels of *PDF1.2* in *SynCPR5* plants

Total RNA was extracted from whole plants, and relative levels of *PDF1.2* mRNA were measured. Three biological and three technical replicates were included, and values were normalized to the average expression values of *ACT2* and *UBC9* housekeeping genes. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant from Col-0; * = significant from *cpr5-2*; ** = significant from Col-0 and *cpr5-2*.

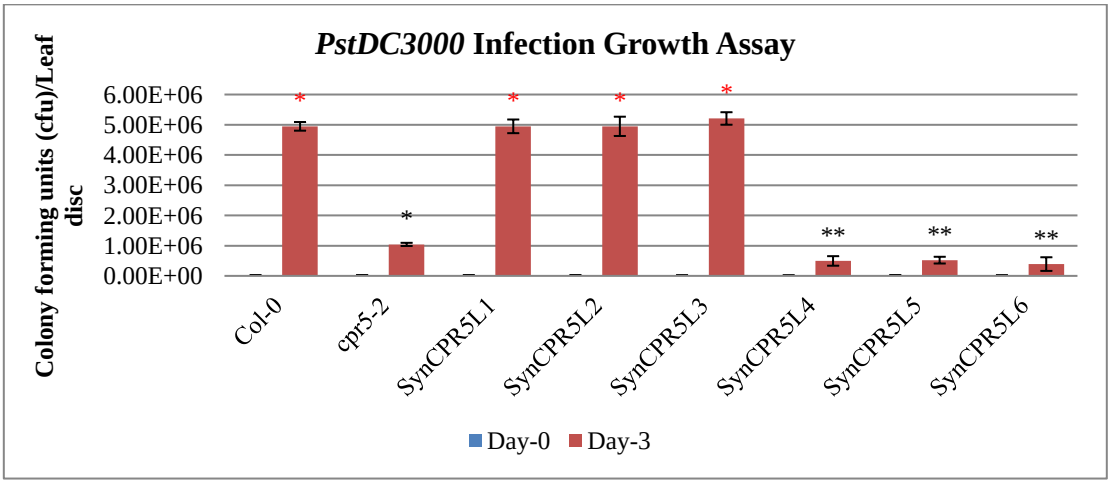
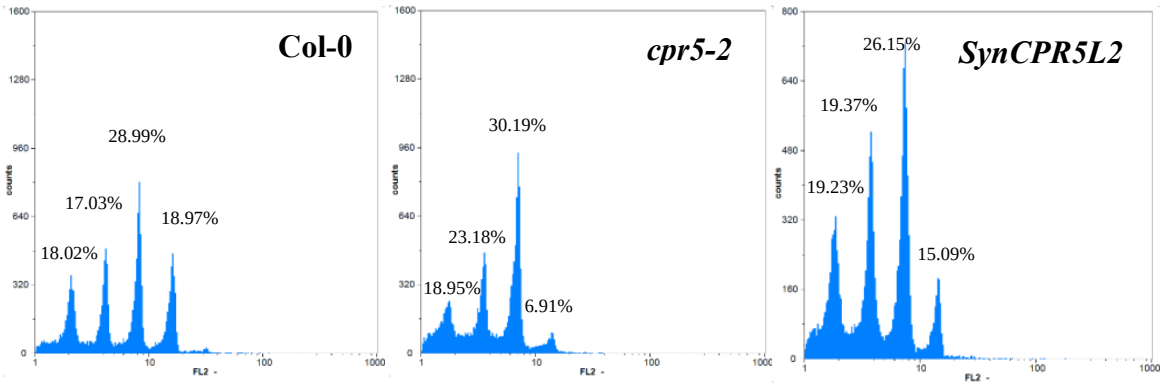


Figure 7.5 *PstDC3000* infection growth assay

This figure shows the growth of *PstDC3000* bacteria in the form of colony forming units (cfu) formed on the agar plate. Three discs of leaves were cut from random leaves of three plants that were inoculated with PstDC3000. Leaf discs were pooled and ground in order to extract bacteria in 500 µL of $MgCl_2$. Dilutions were made and plated as described in materials and methods. The bars represent the standard error and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant from *cpr5-2*; * = significant from Col-0; ** = significant from Col-0 and *cpr5-2*.

A.



B.

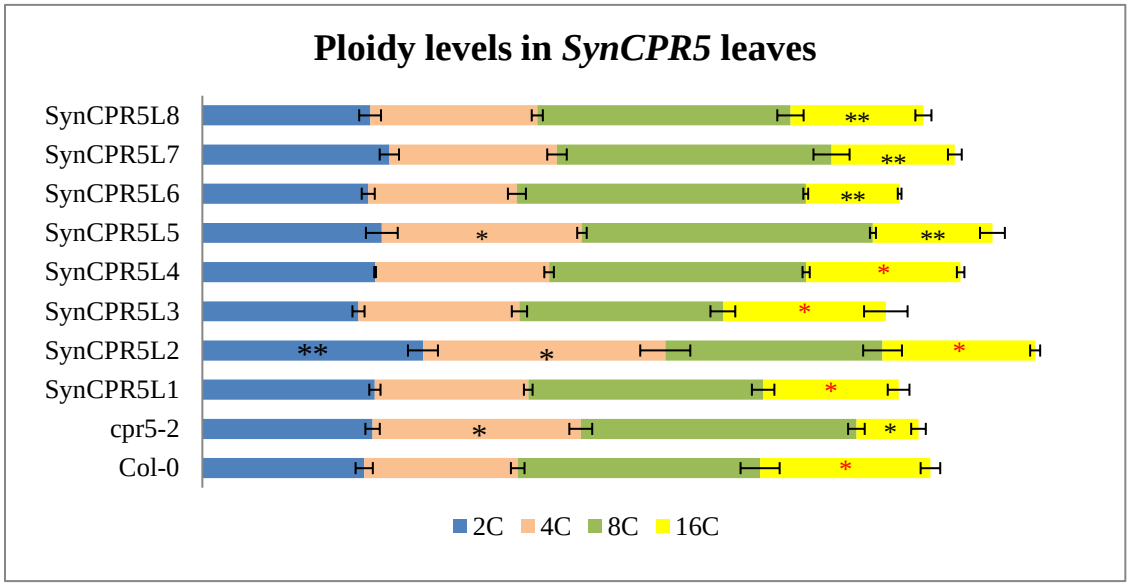


Figure 7.6 Ploidy levels in leaves of *SynCPR5* transgenic plants

Part **A** of the figure 7.6 shows the images obtained from the Flow-max peak analyses (only representative samples are shown). Part **B** shows the percentages of nuclei populations in the leaf cells of Col-0, *cpr5-2*, and *SynCPR5* plants. The nuclei were extracted from the leaf cells of 24 old plants and then stained with propidium iodide (PI) as described in materials and methods. Following overnight incubation at 4°C (in the fridge), the extracted nuclei were analysed by flow-cytometry according to the instructions provided by the manufacturer. The bars represent the standard error and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant from *cpr5-2*; * = significant from Col-0; ** = significant from Col-0 and *cpr5-2*.

7.2.8 *metCPR5* complement aberrant trichomes, HR-like lesions and leaf yellowing shown by *cpr5-2*

As shown in Chapter 3 (Section 3.2.2), the *in silico* analyses shows that the N-termini of AtCPR5 homologues are highly polymorphic in composition, and the length (amino acid sequence) of these highly polymorphic N-termini varies among AtCPR5 homologues (Figure 3.1). Moreover, the N-terminal region (~100 amino acids) of AtCPR5 is predicted to possess three putative in-frame start codons or translation start sites (section 3.2.5) in addition to the presence of predicted RNA secondary structures between these start codons (Figure 3.3). It was assumed that mutations in putative start codons may influence stability of At*CPR5* mRNA and consequently result in *cpr5*-like phenotypes in *metCPR5* lines. In order to see the effect of mutations in the putative start codons on the stability of *metCPR5* mRNAs and the subsequent effect on phenotypes, the putative start codons present at positions B and C (B = 38th, 39th and 40th, C = 63rd amino acid of CPR5) were mutated in the *SynCPR5* gene, individually (*metCPR5b*) as well as collectively (*metCPR5bc*), leaving the first start codon present at position A unchanged. The *metCPR5* transgenes were expected to encode a full length CPR5 protein. However, *metCPR5b* was unable to be translated from position B, whereas, in *metCPR5bc*, both of the putative translation start sites were mutated collectively. Therefore, the potentially functional but shorter CPR5 proteins would not be translated in the *metCPR5* lines. However, the *metCPR5* transgenic plants were shown to have normal trichomes, no HR-like lesions or early leaf yellowing (Figure 7.7 and 7.8B). Overall, *metCPR5* lines were shown to have all typical types of trichomes as found on the wildtype (Figure 7.8B). The number of trichomes having three-appendages was found to be higher on the leaves of the majority of *metCPR5* lines, except *metCPR5bcL2*, where trichomes with three-appendages were shown to be indifferent from the wildtype (Figure 7.8B). On the other hand, the number of trichomes having four-appendages was shown to be lower on the majority of *metCPR5* lines than the wildtype, except *metCPR5bL3* (Figure 7.8B). In conclusion, *metCPR5* transgenes restored wildtype-like trichomes and suppressed HR-like lesions and early leaf yellowing.

7.2.9 *metCPR5b* lines show higher transcript abundance than wildtype

Keeping in view, the variable transcript abundances shown the *SynCPR5* lines (Section 7.2.1), *metCPR5* lines were expected to show variable expression levels of their transgenes. To check *metCPR5* expression, relative transcript abundance of the *metCPR5* transgene, At*CPR5* and *cpr5-2* was measured using a transgene specific set of primers. As shown in Figure 7.8A, *metCPR5b* lines are shown to have higher transcript levels of *metCPR5b* transgenes compared to the wildtype (Figure 7.8A). Contrary to *metCPR5b* lines, *metCPR5bc* transgenic lines displayed wildtype-like levels of *metCPR5bc* transcripts, compared to *metCPR5b* lines (Figure 7.8A). A closer investigation of expression levels depict that expression levels vary among *metCPR5b* lines. In summary, these results support my hypothesis that *metCPR5* lines would have variable or wildtype-like expression levels of the *metCPR5* transgene.

7.2.10 *metCPR5b* and *metCPR5bc* have bigger leaves than wildtype

While characterising *metCPR5* phenotypes, the leaves of the *metCPR5* plants were observed to be bigger than the wildtype. As shown in Section 7.1.3, despite having higher transcript abundance than the wildtype, none of the *SynCPR5* lines displayed larger leaves. The At*CPR5* overexpression (OE) lines developed by Jing et al., (2007) display a 1000-fold increase in their transcript abundance compared to the wildtype, yet nothing is mentioned about any remarkable change in leaf size, by Jing et al., (2007). In the current study, the *metCPR5b* plants show about 5-fold higher transcript levels compared to the wildtype. Thus, measurements of the average leaf area of *metCPR5b*, *metCPR5bc*, OE lines and control (Col-0 and *cpr5-2*) plants will show leaf size difference among these lines. If *metCPR5* leaves are indeed bigger, the increase in size could be the consequence of either transcript abundance, or be solely the effect of mutations in the putative start codon. It was hypothesised that the structural motif containing the putative start codons is responsible for bigger leaves on *metCPR5* plants. To test this hypothesis, leaf areas at different growth stages were recorded. To reiterate, the area of leaves was measured in pairs, for example, leaf 1 and 2 were taken as one pair. Similarly, leaf 3 and 4 were considered the second rosette leaf pair and so on. In addition to Col-0 and *cpr5-2*, two independent *CPR5* overexpression lines (OE1 and OE2) were also included in the study.

As shown in Figure 7.9A, some of the leaf pairs (LP) from the overexpression lines (OE1 and OE2) were shown to be indifferent from the wildtype, except leaf pair 5 from OE1, which was bigger than the wildtype. Similarly, LP4 and LP5 from the OE2 lines were also bigger than the wildtype. As expected, the majority of leaf pairs from the *metCPR5b* plants were significantly bigger than the wildtype, as well as from *metCPR5bc* lines at 31 DAS (Figure 7.9A). On the other hand, areas of majority of third and fourth leaf pairs from *metCPR5bc* plants were shown to be indifferent from the wildtype (Figure 7.9A). Overall, LP5 is shown to be bigger on all the *metCPR5* lines compared to the wildtype. In conclusion, these results indicate that *metCPR5* especially *metCPR5b* plants have bigger leaves compared to the wildtype. Additionally, leaf size regulation appears to be independent of At*CPR5* transcript abundance. However, bigger LP5 on OE lines indicate a degree of dependence of leaf size regulation on *CPR5* transcript abundance.

7.2.11 *metCPR5b* and *metCPR5bc* have wildtype-like epidermal pavement cells

Generally, the final cell size is an outcome of increases in cell division, cell expansion or the combination of both processes (Sugimoto-Shirasu et al., 2003). The finding of bigger leaves on *metCPR5* transgenic lines compared to the wildtype and *cpr5-2* plants prompted me to check the sizes of leaf epidermal pavement cells of *metCPR5* transgenic plants. Keeping in view, the bigger leaves, it was hypothesised that the *metCPR5* transgenic lines will have bigger cells compared to the wildtype. To test this hypothesis, the images of four (3^{rd}, 4^{th}, 5^{th} and 6^{th}) rosette leaves were taken using a scanning electron microscope (SEM), and areas of the selected cells from the SEM images were recorded using ImageJ as described in materials and methods (Section 2.15). Since the average area of epidermal cells from leaf 3 was found to be similar to the cells from leaf 4 of the same plant, and similarly the cells from leaf 5 were of similar sizes to leaf 6 (data not shown), images and leaf area data is shown only for leaf 3 and leaf 6. Leaves from OE lines were missed from being included for epidermal pavement cell measurement.

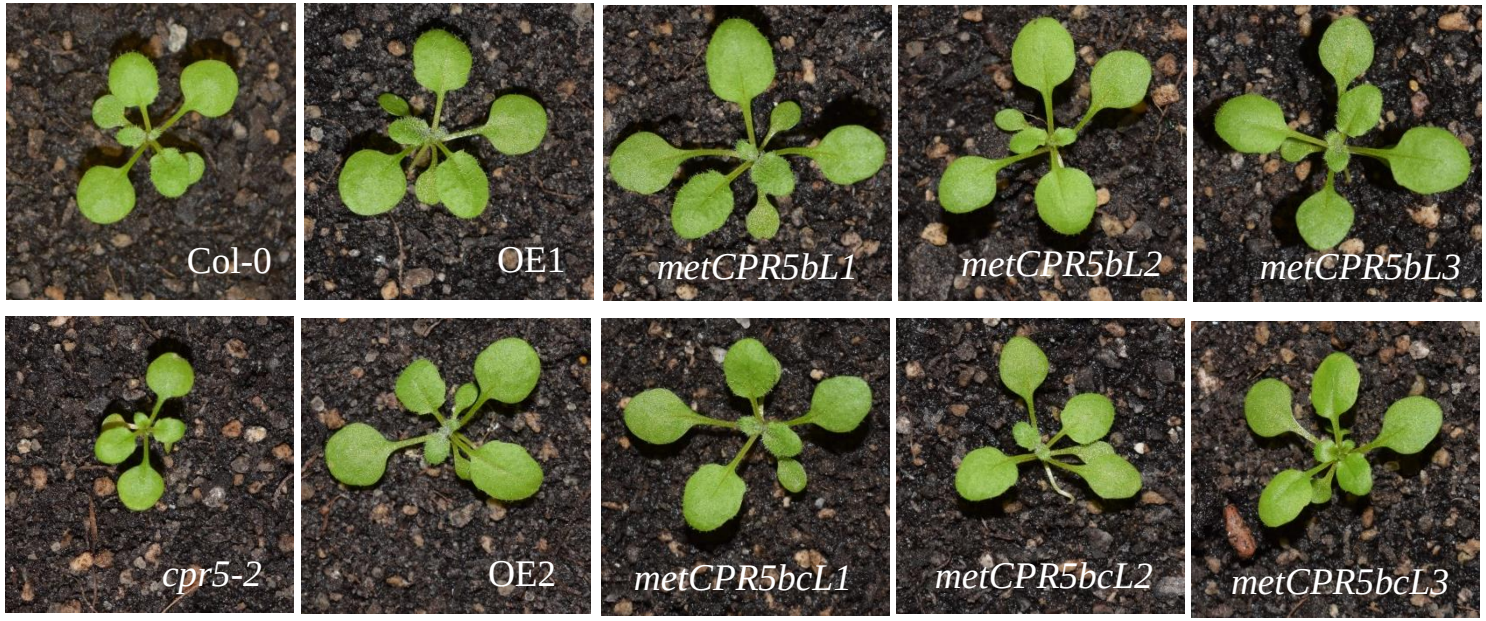
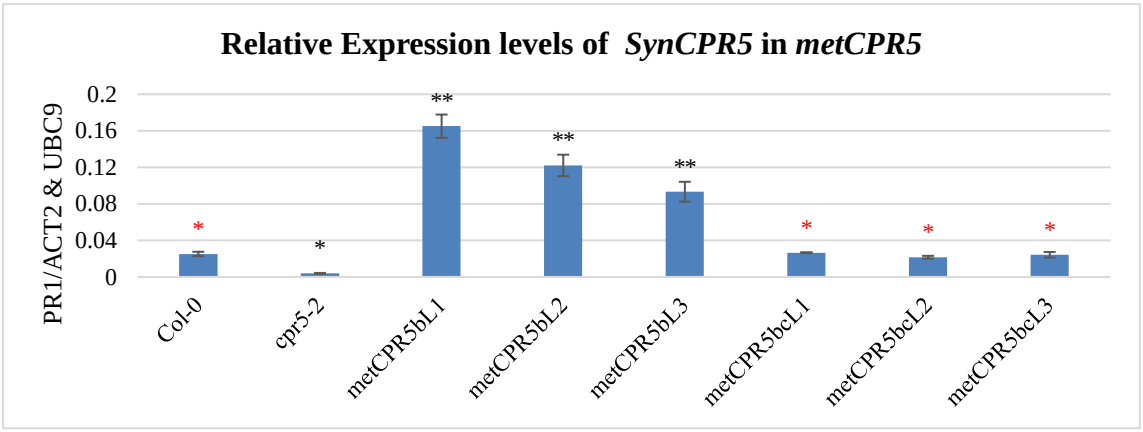


Figure 7.7 Overexpression, *metCPR5*, Col-0 and *cpr5-2* plants at 17 DAS

Plants were sown in soil under long day conditions (14 hours light:10 hours dark). Plants were photographed by digital camera (Nikon D5300) fixed on a stand. OE1 and OE2 represent two independent overexpression lines in CPR5 gene was expressed under the 35S promoter and were developed by Jing et al., (2007).

A.



B.

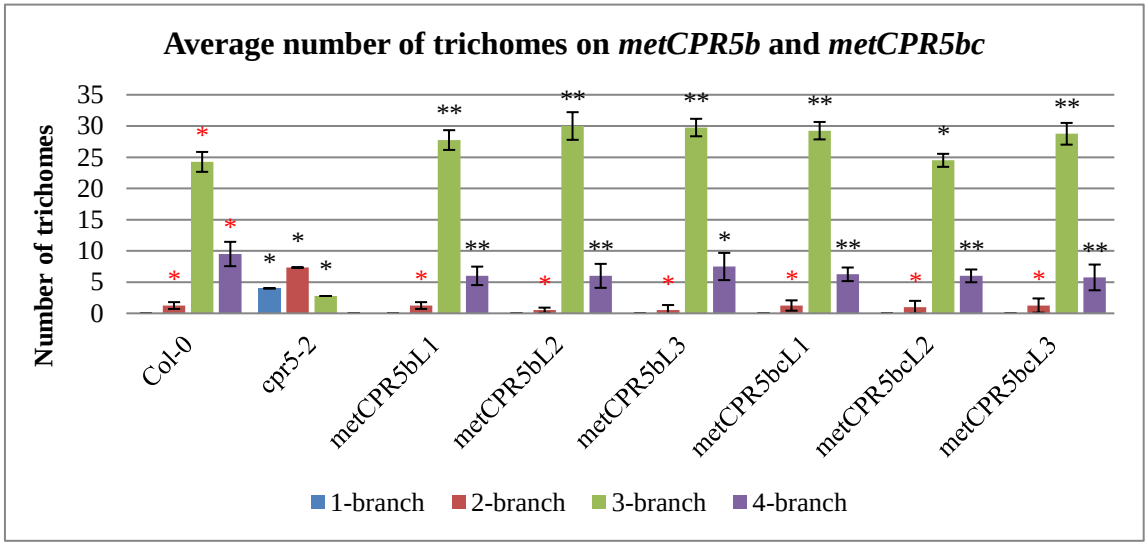


Figure 7.8 *metCPR5* transgene and trichome quantification in *metCPR5* plants

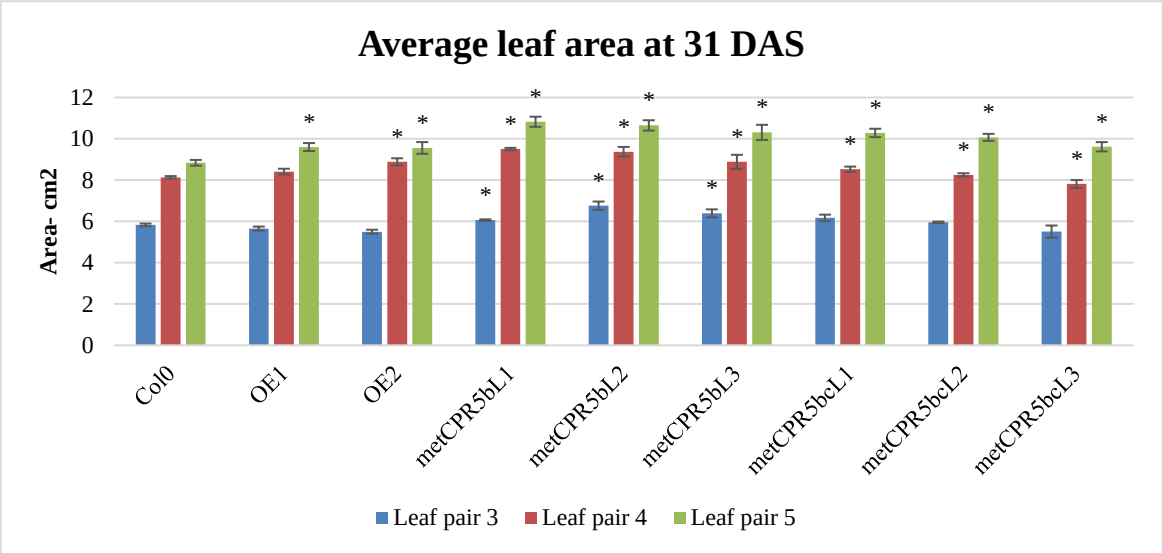
Figure **7.8A** shows the quantification of transcript abundance of *SynCPR5*, Col-0 and *cpr5-2* using *SynCPR5*, Col-0 and *cpr5-2* specific sets of primers. Three biological and three technical replicates were included, and values were normalized to the average expression values of *ACT2* and *UBC9* housekeeping genes. Figure **7.8B** shows the average number of trichomes on leaf discs of 0.68cm diameter from the leaves of 5 different plants of each individual line. *SynCPR5* = synthetic *CPR5* gene. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Student t-test). * = significant compared to wildtype, * = significant compared to *cpr5-2*; ** = significant compared to the wildtype Col-0 and *cpr5-2*.

As shown in Figure 7.9B, the average size of epidermal pavement cells from leaf 3 and leaf 6 of wildtype CPR5 plants was shown to be significantly bigger than *cpr5-2* mutant leaf cell size. The average size of pavement cells of leaf 6 from *metCPR5b* and *metCPR5bc* plants was found to be bigger but statistically insignificant from the wildtype. On the other hand, the area of leaf 3 cells from *metCPR5b* was found to be indifferent, but cells of the same leaves from *metCPR5bc* plants were significantly bigger compared to the cells from the wildtype (Figure 7.9B).

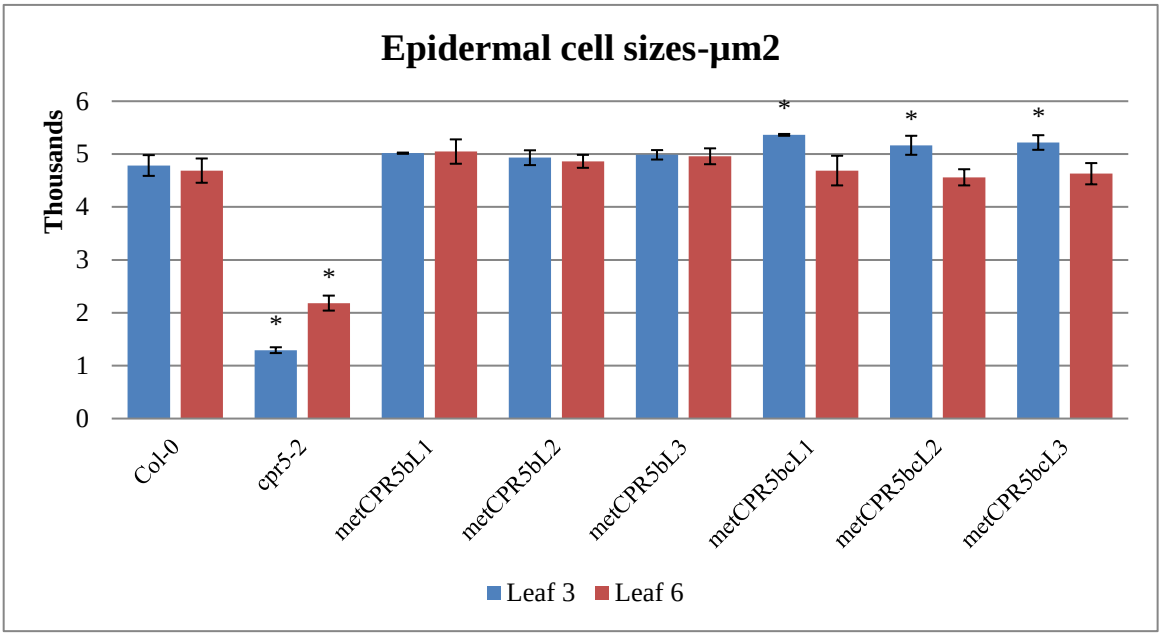
In addition to the cell area, the average number of cells from the same images (used for area measurements) of each line was also calculated. It is important to note that some of the cells closer to the boundary of the images did not have complete boundaries when imaged (Figure 7.9C). Thus, only cells with complete boundaries were counted. As shown in Table 7.1, the average number of cells from leaf 3 and 6 of the CPR5 wildtype was counted to be 28 and 31 respectively, compared to *cpr5-2*, which had 145 and 85 cells per image, respectively. In contrast to *cpr5-2* mutants, the average number of cells per image from *metCPR5* leaf 3 and 6 was shown to be less than the wildtype (Table 7.1). For instance, the average number of cells belonging to leaf 3 and leaf 6 on *metCPR5b* transgenic plants was 25. Similarly, the average number of cells on *metCPR5bc* leaves was counted to be 26 and 27, respectively.

In conclusion, these results indicate that the overall cell sizes of *metCPR5* plants are indifferent from the wildtype, though the cells from the third leaf of *metCPR5bc* are bigger than the wildtype. Additionally, the lower number of cells per image in *metCPR5* plants indicates that genetic changes in *metCPR5b* and *metCPR5bc* have somehow contributed in producing bigger cells, though size differences are insignificant when compared to the wildtype.

A.


Average leaf area at 31 DAS
Area- cm2
12
10
8
6
4
2
0
Col0
OE1
OE2
metCPR5bL1
metCPR5bL2
metCPR5bL3
metCPR5bcL1
metCPR5bcL2
metCPR5bcL3
Leaf pair 3
Leaf pair 4
Leaf pair 5

B.


Epidermal cell sizes-µm2
Thousands
6
5
4
3
2
1
0
Col-0
cpr5-2
metCPR5bL1
metCPR5bL2
metCPR5bL3
metCPR5bcL1
metCPR5bcL2
metCPR5bcL3
Leaf 3
Leaf 6

C.

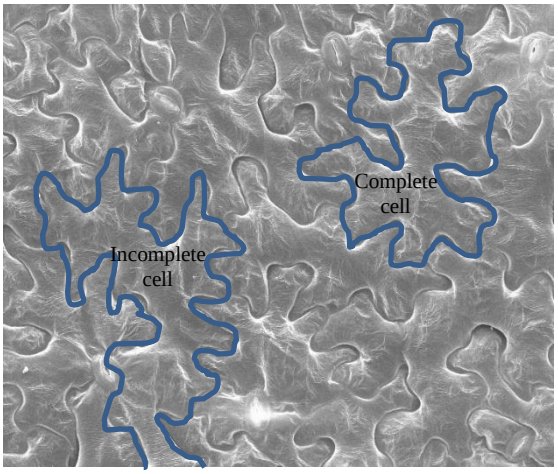


Figure 7.9 Mean area of leaves and epidermal pavement cells of *metCPR5* plants

Part (**A**) of figure **7.9** shows the mean leaf areas of cotyledons and first five rosette leaf pairs at 31 DAS. The leaves were plucked from 3-5 different plants of the same line, and areas were recorded using a leaf area machine. Figure **7.8B** shows the average area of epidermal pavement cells measured using a scanning electron microscope, as described in materials and methods. Areas of the cells were measured which had complete boundaries. Overexpression lines were not included for cell size measurement since these lines were not available when the other lines were grown. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Student t-test) compared to the wildtype.

Average number of cells per image								
	Col-0	***cpr5-2***	***metCPR5b***			***metCPR5bc***		
			L1	**L2**	**L3**	**L1**	**L2**	**L3**
Leaf 3	28	145*	26*	25*	25*	26*	26*	27
Leaf 6	31	85*	25*	25*	26*	28*	28*	27*

Table 7.1 Average number of cells per image

This table shows the average number of epidermal pavement cells, which were counted from 3-5 SEM images using ImageJ as described in materials and methods (Section 2.15). * indicates the level of statistical significance at $p < 0.05$ (Student t-test) compared to wildtype Col-0.

7.2.12 *metCPR5b* and *metCPR5bc* leaves have higher ploidy levels

Leaf size is positively correlated with ploidy levels (Sugimoto-Shirasu and Roberts, 2003). In *Arabidopsis thaliana*, leaves generally undergo three to four rounds of endoreduplication and contain nuclear populations of DNA content with 2, 4, 8, 16 and 32 C. Since *cpr5-2* plants show reduced ploidy levels, thus, the ploidy levels of nuclei from *metCPR5* leaves were measured in order to see the effect of genetic changes in *metCPR5* on ploidy levels. It was hypothesised that *metCPR5* leaves have higher ploidy levels than the wildtype. To test this hypothesis, third and fourth leaves were harvested from 20-day (20 DAS) old *SynCPR5* plants. Nuclei were extracted from the leaves by chopping the leaves in a chopping solution, and were subsequently stained with propidium iodide as mentioned in materials and methods (Section 2.16).

Compared to the wildtype, the majority of *metCPR5* plants, including OE and *cpr5-2* lines, displayed a significantly less number of 2C nuclei. A closer view of the 2C nuclei populations reveals that the majority of the *metCPR5* lines (except *metCPR5L1* and *metCPR5L4*) have intermediate levels of 2C nuclei as compared to *cpr5-2*, wildtype as well as OE lines (Figure 7.10). The OE lines show a population of 2C nuclei indifferent from *cpr5-2*. In contrast to 2C nuclei, there is a lot of variation among *metCPR5* lines with respect to 4C populations, making it difficult to draw a line between lines. With respect to 8C nuclei, the majority of the *metCPR5* lines displayed insignificant levels when compared to the wildtype (Figure 7.10). However, OE lines showed a higher number of 8C nuclei compared to the wildtype and *cpr5-2*. Compared to the wildtype, the presence of a 16C population only in *metCPR5* lines concludes the most striking result of this section. In conclusion, these results partly support the formulated hypothesis, since *metCPR5* lines do not show increases for all types of nuclei populations. Moreover, the presence of 16C nuclei in *metCPR5* lines in contrast to the wildtype, *cpr5-2* and OE lines, indicates that in contrast to *metCPR5* nuclei, nuclei in the wildtype, *cpr5-2* and OE lines did not enter the third round of endoreduplication.

7.2.13 *metCPR5b* and *metCPR5bc* have lower or wildtype-like *PR1* levels

In addition to leaf area, cell size and ploidy levels, *metCPR5* transgenic lines were also characterised in order to the see the complementation of *PR1* gene levels. All of the *SynCPR5* lines have shown upregulated levels of *PR1*. Thus, it was hypothesised that like *SynCPR5* lines, *metCPR5* lines will also show higher *PR1* transcript levels. To test this hypothesis, the transcript abundance of *PATHOGENESIS RELATED GENE1* (*PR1*) was quantified from *metCPR5b* and *metCPR5bc* transgenic plants using real-time qRT-PCR.

As shown in Figure 7.11, the majority of *metCPR5b* transgenic lines displayed *PR1* transcript abundance insignificant from the wildtype, except *metCPR5bL3*, which has significantly lower *PR1* levels. Contrary to *metCPR5b*, the majority of *metCPR5bc* lines exhibited reduced levels of *PR1* compared to the wildtype, except *metCPR5bcL4* (Figure 7.11). In summary, these results do not support the formulated hypothesis, since the upregulated levels of *PR1* in *cpr5-2* as well as in *SynCPR5* plants, are restored to wildtype-like levels or lower than wildtype levels in the *metCPR5* transgenic lines.

7.2.14 *metCPR5* plants conferred enhanced susceptibility to *P syringae*

As shown above in Section 7.1.4, *SynCPR5* plants displayed a range of *PR1* transcript abundance, and half of all of these lines conferred resistance to *PstDC3000*, despite having upregulated *PR1* levels. On the other hand, *metCPR5* plants have shown wildtype-like or lower than wildtype levels of *PR1* transcripts. Therefore, *metCPR5* lines were hypothesised to show susceptibility to the *PstDC3000* bacterial pathogen. To test this hypothesis, *metCPR5* plants were grown in short days under 12:12 (Light:dark) conditions. Leaves from 4-5 week old plants were infiltrated with *PstDC3000*, and resistance was scored by measuring the number of colony forming units (cfu) as described in materials and methods (Section 2.15). As shown in Figure 7.12, the *metCPR5b* and *metCPR5bc* lines displayed susceptibility to *PstDC3000* significantly higher than the wildtype (Figure 7.12). In conclusion, these results are in line with our notion that these plants, having wildtype-like or lower than wildtype levels of *PR1*, should be susceptible to *PstDC3000*.

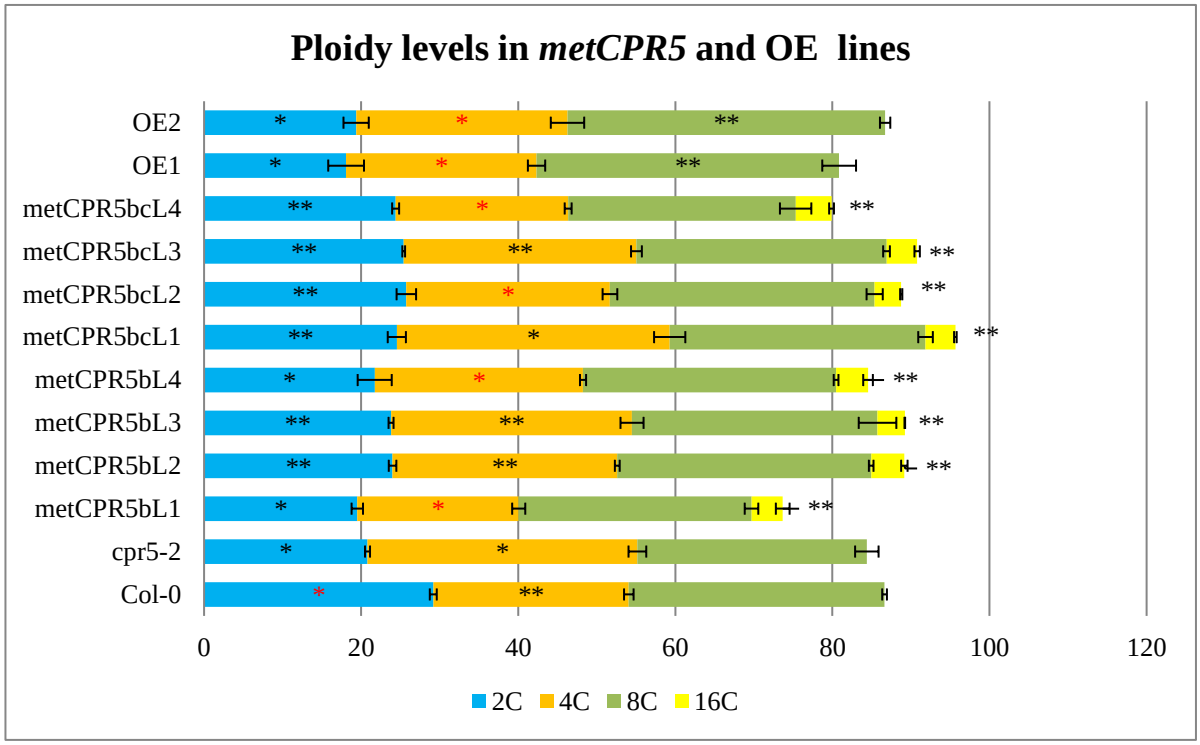


Figure 7.10 Ploidy levels in leaves of *SynCPR5* transgenic plants

This figure shows the proportion or percentages of nuclei populations in the leaf cells of Col-0, *cpr5-2*, OE and *metCPR5* plants. The nuclei were extracted from leaf cells of 24 old plants by mincing the leaves with a chopping blade in a glass petri plate and then staining with propidium iodide (PI), as described in materials and methods. Following overnight incubation at 4°C (in the fridge), the extracted nuclei were analysed by flow-cytometer (Partec, USA) according to the instructions provided by the manufacturer. Percentages of each population from each sample were calculated and average (percentage) values were shown in graph form. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Student t-test). * = significant compared to wildtype, * = significant compared to *cpr5-2*; ** = significant compared to wildtype Col-0 and *cpr5-2*.

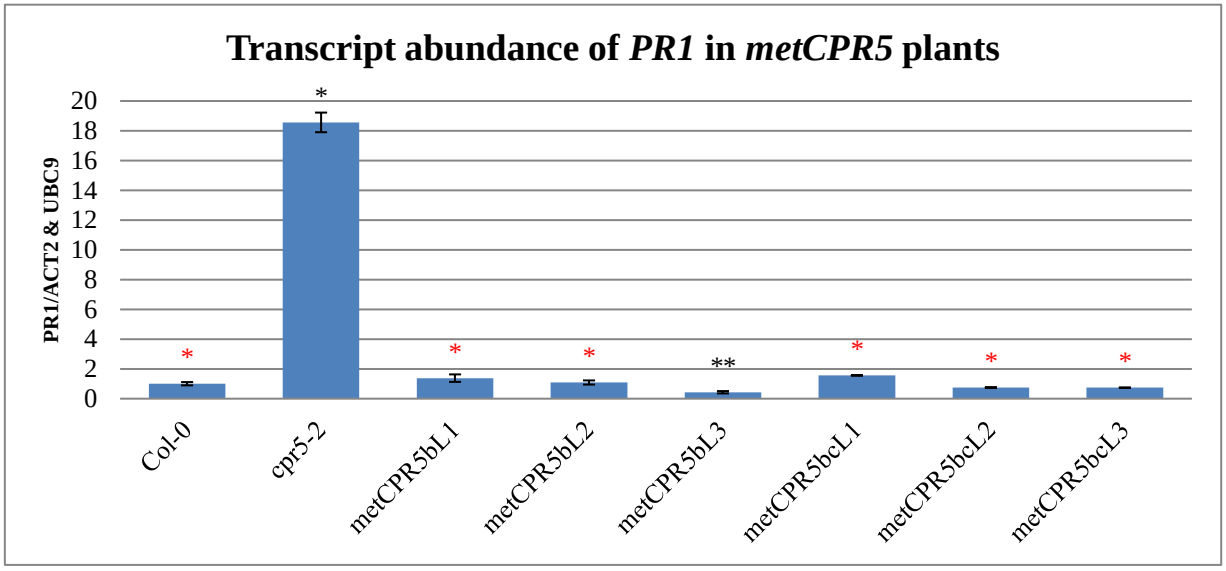


Figure 7.11 Relative expression levels of *PR1* in *SynCPR5* plants

Total RNA was extracted from whole plants and relative levels of *PR1* mRNA were measured. Three biological and three technical replicates were included, and values were normalized to the average expression values of *ACT2* and *UBC9* housekeeping genes. The bars represent the standard error and the asterisks indicate the level of statistical significance at $p < 0.05$ (Student t-test). * = Significant from Col-0; * = significant from *cpr5-2*; ** = significant from Col-0 and *cpr5-2*.

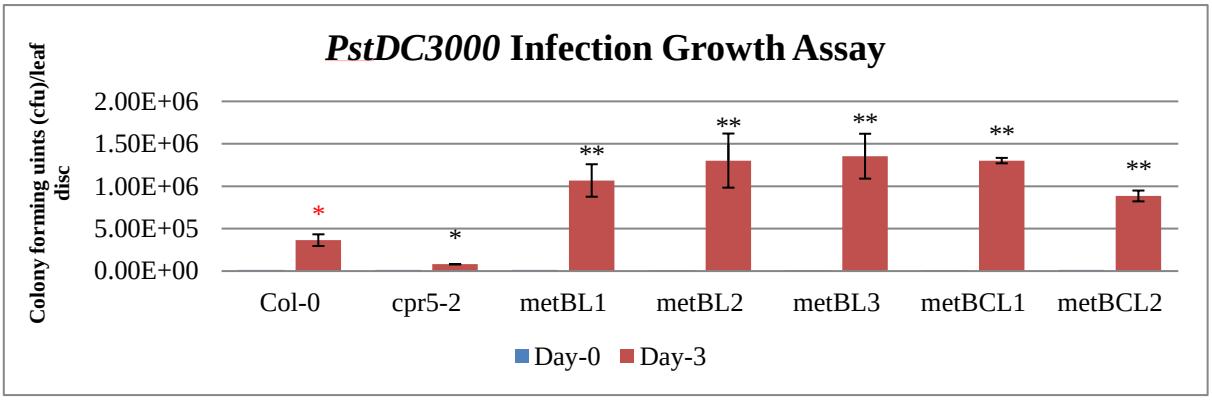


Figure 7.12 *PstDC3000* infection growth assay

This figure shows the growth of *PstDC3000* bacteria in the form of colony forming units (cfu) formed on the agar plate. Three discs of leaves were cut from random leaves of three plants that were inoculated with *PstDC3000*. Discs were cut immediately form the leaves for samples at day 0. Leaf discs were pooled and grounded in order to extract bacteria in 500 µL of $MgCl_2$. Dilutions were made and plated as described in materials and methods. The bars represent the standard error and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test).* = significant from Col-0; * = significant from *cpr5-2*; ** = significant from Col-0 and *cpr5-2*.

7.3 DISCUSSION

7.3.1 *SynCPR5* transgenic plants have variable *SynCPR5* expression levels

CPR5-fused with GFP and driven by the 35S promoter have resulted in the ectopic distribution of the CPR5 protein in the cell such as cytoplasm, plasma membrane, nucleus and in the proximity of nuclear lemma (Dijkwel's lab, results unpublished). The distribution of CPR5 in plasma as well as nuclear lemma could be the result of overexpression driven under 35S as observed in many studies (Pino et al., 2007a, Pino et al., 2007b, Gookin and Assmann, 2014, Dutt et al., 2014, Liu and Stewart Jr, 2016). Since the random integration of transgenes through *Agrobacterium*-mediated transformation in the host results in variable expression (position effect) (Peach and Velten, 1991, Gelvin, 2003), expression of *SynCPR5* transgenes under a native promotor was expected to result in variable expression levels in the current study. The *CPR5* native promoter has already been used in some previous studies (Jing et al., 2007; Gao et al., 2011; Perazza et al., 2011; Wang et al., 2014). However, only one of these studies report the length of the promoter (741 bp upstream of the *CPR5* start codon), which successfully complemented *cpr5-1* (Gao et al., 2011). Additionally, the relative transcript abundance of *CPR5* under the native promoter was not quantified and shown by Gao et al., (2011). Contrary to Gao et al., (2011), a longer segment (1.5 kb) of DNA sequence upstream of the *CPR5* start codon was used as the *CPR5* native promoter.

The results show that the *CPR5* native promoter successfully managed to result in *SynCPR5* variable expression, such as, **1**) wildtype-like; **2**) higher than the wildtype and **3**) lower than the wildtype (Figure 7.1A). This difference in transgene expression levels could be due to the position of integration in the host genome. The transformation of transgenes using the *Agrobacterium*-mediated T-DNA method, results in variable expression of the transgenes, since T-DNA cassette(s) carrying transgene(s) integrate in the genome randomly and consequently, their levels of expression of transgene depend on the position and region in the host genome, known as position effect (Peach and Velten, 1991, Gelvin, 2003). A transgene integrated closer to, or far from, transcriptional activating elements or enhancers could result in high or low expression. A transgene could also fail to be expressed when integrated into a

transcriptionally silenced region (Gelvin, 2003, Zhu et al., 2003). In the current study, a large number of independent transgenic lines of each transgene were isolated (Table 6.4) and studied in order to choose lines with similar levels of expressions. To conclude, these results confirm that the *CPR5* native promoter was able to drive *SynCPR5* expression to varying levels, likely due to position effect.

7.3.2 *SynCPR5* complements majority of *cpr5-2* compromised phenotypes

In previous studies, *cpr5-2* mutant plants have been complemented with the *CPR5* gene under a native (Gao et al., 2011), as well as 35S promoter (Boch et al., 1998; Yoshida et al., 2002; Jing et al., 2007; Gao et al., 2011; Perazza et al., 2011; Wang et al., 2014). In the majority of these studies, complementation was confirmed or scored on the basis of phenotypes shown to be compromised in *cpr5-2* or *cpr5-1* plants, such as, aberrant trichomes, HR-like lesions on leaves and precocious leaf senescence (Boch et al., 1998; Yoshida et al., 2002). In addition to the progression of normal leaf senescence and trichomes, Yoshida et al., (2002) observed a wildtype-like response to sugars in complemented lines of the *cpr5* allele *hys1-1*. Lesions on leaves and ABA hypersensitivity in *cpr5-1* were found to be restored in *cpr5-1* complemented lines, in which the *CPR5* gene was driven by the native and 35S promoters (Gao et al., 2011). The chlorophyll content of overexpression lines was shown to be lower than wildtype levels, despite having higher *CPR5* transcript abundance in the overexpression lines (Jing et al., 2007). Likewise, accelerated senescence was observed in these overexpression lines (Jing et al., 2007). Additionally, the overexpression lines also displayed partial complementation with respect to JA/ABA responses (Jing et al., 2007), which suggests that 35S promoter results in ectopic phenotypes. These studies show that majority of the *cpr5*-compromised phenotypes were managed to be rescued in the complemented lines. In the current study, complementation of *cpr5-2* plants with *SynCPR5* transgene(s) has successfully rescued the majority of compromised phenotypes and resulted in execution of wildtype-like phenotypes. For example, in contrast to aberrant trichomes on *cpr5-2* (Kirik et al., 2001, Brininstool et al., 2008), wildtype-like trichomes were found on *SynCPR5* plants. The leaves of *SynCPR5* transgenic plants were found to be healthy, displaying no early leaf senescence or HR-like lesions as compared to *cpr5-2* (Boch et al., 1998; Bowling et al., 1997; Jing et al., 2002). The cotyledons

and first rosette leaf pairs on *SynCPR5* transgenic plants were found green and senesced in a wildtype manner, compared to the cotyledons and first rosette leaf pair on *cpr5-2* plants (data not shown). These results confirm that *SynCPR5* transgene has successfully complemented *cpr5-2* compromised phenotypes.

However, in the majority of the previous *cpr5* complementation analyses, transcript abundance of *PR1* gene was ignored except by Jing et al., (2007). All of the overexpression lines (driven under 35S) showed upregulated levels of *PR1* (Jing et al., 2007). Similarly, a slightly higher but insignificant increase in *PR1* levels can be observed in Perazza et al., (2011) studies when *CPR5* was expressed under the *VP16* constitutive transcriptional activation domain. These results suggest that *PR1* upregulation was not rescued in complemented lines or *CPR5* overexpression resulted in *PR1* upregulation. These results are in line with our finding, where *SynCPR5* transgenic lines also showed enhanced levels of *PR1*, the exact reason or mechanism of which is yet unknown. However, it could be possible that truncated cpr5 protein in the background could be interfering and/or competing for interaction with the wildtype or SynCPR5 proteins by acting as dominant negative allele (Lorrain et al., 2003, Belkhadir et al., 2004, Palma et al., 2010) and hence inducing *PR1* upregulation.

In summary, *SynCPR5* transgenic plants managed to show healthy leaves with normal trichomes and without HR-like lesions and early leaf senescence. At the transcriptional level, these plants showed upregulated *PR1* levels. Moreover, the execution of wildtype-like growth with enhanced *PR1* expression is an interesting finding.

7.3.3 Phenotypes suggest quantitative as well as qualitative roles for CPR5

SynCPR5 variable expression levels have allowed me to uncouple the phenotypes regulated quantitatively from those regulated qualitatively. As shown in the results, the *SynCPR5* transgenic lines have shown a range of leaf size, trichome numbers and ploidy levels. For example, the production of three types of leaves (wildtype-like, intermediate and smaller i.e. *cpr5*-like) by *SynCPR5* plants suggests a positive correlation between *SynCPR5* transcript abundance and leaf size regulation. Despite having fewer trichomes with four appendages in *SynCPR5L3*, all the *SynCPR5* plants which have higher transcript abundance have exhibited

trichomes in the wildtype manner. Whereas, an intermediate number of trichomes was observed on *SynCPR5* plants carrying wildtype-like or lower than wildtype *SynCPR5* expression levels, except *SynCPR5L1* (Figure 6.1B). Similarly, bigger leaves and more number of trichomes shown by *metCPR5* plants also suggest leaf size and trichomes to be quantitatively regulated traits by AtCPR5. Lastly, wildtype-like and intermediate levels of ploidy levels shown by *SynCPR5* and the presence of 16C nuclei in *metCPR5* also reinforce the quantitative role of CPR5. In addition to leaf size, trichome numbers and ploidy levels, variable trends (differential expression) were observed with respect to defence-related genes *PR1* and *PDF1.2*. However, expression trends of *PR1* and *PDF1.2* are shown to be independent of *SynCPR5* transcript abundance. For instance, *SynCPR5L2* and *SynCPR5L3*, despite having higher *SynCPR5* transcript abundance, these lines had higher levels of *PR1* than wildtype (Figure 6.3).

On the other hand, there are a number of traits, which are shown to be qualitatively regulated by At*CPR5* i.e. some phenotypes were either shown to be fully complemented or failed to be complemented by *SynCPR5*, such as, HR-like lesions, early leaf senescence and resistance to the bacterial pathogen, *Pseudomonas syringae*. For example, despite having variable *SynCPR5* expression levels, there were no intermediate levels of HR-like lesion and early leaf senescence. Similarly, *SynCPR5G1* lines displayed susceptibility phenotypes, whereas, *SynCPR5G2* plants exhibited resistance regardless of *SynCPR5* and *PR1* transcript abundances. In summary, these results show that leaf size, trichome number and ploidy levels are regulated quantitatively, whereas, the execution of HR-like lesions, early leaf senescence and resistance is controlled qualitatively by *CPR5*. Additionally, the regulation of *PR1* and *PDF1.2* is independent of *CPR5* expression levels.

7.3.4 *PDF1.2* is expressed at lower levels in *SynCPR5* and *cpr5-2* than *CPR5*

In previous studies (Bowling et al., 1997, Clarke et al., 2000), *cpr5-2* is shown to produce higher levels of *PDF1.2* as compared to *PDF1.2* levels in the wildtype. However, the quantification of *PDF1.2* using quantitative real-time qRT-PCR, displays contrasting results. In the current study, the relative expression levels of *PDF1.2* in *cpr5-2* plants were found to be lower than in wildtype plants in contrast to previous studies. In addition to *cpr5-2*, *SynCPR5*

complemented lines have also exhibited lower levels of *PDF1.2* compared to the wildtype. A closer view of Figure 7.4 shows that the relative levels of *PDF1.2* vary with the change in *SynCPR5* expression levels (Figure 7.1). However, no correlation can be established between *SynCPR5* and *PDF1.2* transcript abundance.

Regardless of variable *PDF1.2* expression levels, *PDF1.2* downregulation in *cpr5-2*, *SynCPR5* and *metCPR5* transgenic complemented lines has not been reported previously. It could be the absence of functional *CPR5* that positively regulates *PDF1.2* or the suppression of *PDF1.2* by SA/*PR1* antagonism. *PDF1.2* is generally induced by either necrotrophic fungal pathogens or wounding through ethylene (ET) or jasmonic acid (JA) signalling pathways (Jing et al., 2007). Jing et al., (2007) report that *cpr5-2* (*Arabidopsis* Col-0 *cpr5* allele) as well as *old1-1* (*Arabidopsis* L*er*-0 *cpr5* allele) both show wildtype levels of ethylene in contrast to upregulated JA levels in these *cpr5* alleles. These results indicate that *cpr5-2* should have upregulated *PDF1.2* levels due to enhanced JA levels. The downregulation of *PDF1.2* in *cpr5-2*, *SynCPR5* and especially in *metCPR5* lines points to another (unknown) mechanism(s), for instance SA and JA/ET antagonism, being responsible for *PDF1.2* downregulation.

PATHOGENESIS-RELATED GENE1 (*PR1)* is induced in response to SA and is a marker gene for the confirmation of SA-mediated defence system activation (Palma et al., 2010). *cpr5-2* plants show elevated SA and *PR1* levels (Bowling et al., 1997; Boch et al., 1998; Clarke et al., 2001; Jing et al., 2007; Wang et al., 2014). Additionally, SA plays a major role in suppressing JA/ET signalling and subsequently *PDF1.2* downregulation (Navarro et al., 2008). Thus, keeping in view SA and *PDF1.2* antagonism, *PDF1.2* suppression in *cpr5* and *SynCPR5* plants is likely to be the consequence of increased SA and/or *PR1* levels, since *cpr5-2* and all of the *SynCPR5* transgenic lines show elevated *PR1* levels (SA levels not measured). Post-thesis submission analyses show that *cpr5-2* plants contain upregulated levels of *WRKY70*, a transcription factor that is known to interact with *WKRY11* and subsequently repress *PDF1.2* levels. Thus, these results conspicuously support *PDF1.2* suppression by *PR1* and/or SA antagonism.

A number of studies (Bowling et al., 1997; Boch et al., 1998) show that resistance is conferred by *cpr5-2* against the necrotrophic pathogen *H. parasitica* (*Noco2*) which is believed to be activated through JA/ET signalling pathway. However, SA-mediated resistance is also sufficient to restrict *Noco2* growth, it could be possible that cpr5-mediated resistance to *Noco2* could be through the activation of resistance mediated by SA. Therefore, quantification of *cpr5-2* resistance against different necrotrophic pathogens, such as *A. brassicicola*, will help to confirm the activation of JA-ET-mediated signalling pathway and understand the misregulation of *PDF1.2* in *cpr5-2* plants. Keeping in view *PDF1.2* downregulated levels, *cpr5-2* and *SynCPR5* plants would be expected to exhibit susceptibility to the necrotrophic pathogen, *Alternaria brassicola*. In summary, based on the current results, it could be proposed that CPR5 may be a positive regulator of *PDF1.2*, which is a novel finding. Despite evidence of the role of *CPR5* in the regulation of *PDF1.2*, how *CPR5* regulates *PDF1.2* expression at a genetic level needs further investigation.

7.3.5 Putative roles for nucleotide residues associated with alternative start codons

All the *metCPR5b* plants have shown higher transcript abundance and bigger leaves than the wildtype. So, how could nucleotide changes in the putative start codons (*metCPR5b*) affect transcript abundance and consequently result in bigger leaves on *metCPR5* plants, is an interesting question? Below are some plausible explanations to this question.

The first factor that could affect the transcription of a gene, includes the natural redundancy of the genetic code i.e. the ability of an mRNA to be transcribed or translated from more than one start codons (Gingold and Pilpel, 2011). Natural redundancy provides evolution an opportunity to tune the efficiency and accuracy of protein production at various levels without any change in amino acid composition (Gingold and Pilpel, 2011). However, a change in the genetic code could lead to a change in the amount of corresponding tRNA in the cell and the speed by which a ribosome recognises a specific codon (Varenne et al., 1984, Sørensen et al., 1989). Additionally, these alternative nucleotide sequences (synonymous change) of certain codons could give rise to transcripts with a different secondary structure and stability that could subsequently affect translation (Kudla et al., 2009) and even protein folding (Komar et al.,

1999, Kimchi-Sarfaty et al., 2007). As mentioned in Chapter 5, the *SynCPR5* gene was codon optimized except for the first three hundred base pairs of the At*CPR5* coding region. Thus, it could be possible that nucleotide substitutions in At*CPR5*, due to codon optimization have improved or degraded At*CPR5* mRNA stability. However, as mentioned in Section 7.3.1, the presence of variable *SynCPR5* expression levels around wildtype levels is likely not because of changes in *SynCPR5* expression levels due to codon optimization and/or genetic code redundancy, but due to position effect.

Secondly, there is an increasing number of evidence confirming the role of RNA secondary structures in the regulation and production of protein isoforms from single mRNA, by the use of alternative translation start site(s) (Chabregas et al., 2003, Sunderland et al., 2004, Christensen et al., 2005, Wamboldt et al., 2009, Asano, 2014). In Arabidopsis, THI1, POLγ1, POLγ2, and DNA ligase 1 are some of the examples of protein isoforms produced by the RNA secondary structure or alternative translation sites (Chabregas et al., 2003, Sunderland et al., 2004, Christensen et al., 2005). As shown in Chapter 3 (section 3.2.5), At*CPR5* mRNA is predicted to contain a number of strong stem-loop structures and four start codons in the first 300 bp region. *metCPR5b* includes mutations of 9 nucleotides in order to disrupt 3 consecutive start codons present at the second position (position B), which were annotated to be part of the predicted stem-loop structure of At*CPR5*. However, when the first 500 bp of both versions (wildtype and *metCPR5b*) of the gene were analysed, no difference in stem-loop structure formation and energy between At*CPR5* and *metCPR5* was predicted. These results suggest that the difference in transcript abundance of *metCPR5* transgenic lines is independent of stem-loop structure existence and disturbance. The variable expression levels of *SynCPR5* and *metCPR5* lines indicate position effect as one of the processes responsible for such variations in their transcripts.

Assuming that the increase in transcript abundance of *SynCPR5* and *metCPR5b* plants is independent of nucleotide(s) changes introduced during codon optimization as well as specific selected substitutions, the question arises as to what causes leaves to be bigger on *metCPR5* plants compared to wildtype, OE lines and *SynCPR5* plants?

Like stem-loop structure, the presence of favourable nucleotides required for the initiation of translation in the context of a start codon and the preference of one start codon over the second codon, also plays an important role in controlling the balance of isoforms of a protein (Chabregas et al., 2003; Sunderland et al., 2004; Christensen et al., 2005; Wamboldt et al., 2009; Asano et al., 2014). For example, RNA stem-loops and the presence of favourable residues in the context of initiation sites favoured distinct THI1 isoforms (Chabregas et al., 2003; Sunderland et al., 2004). As shown, AT*CPR5* is predicted to contain multiple start codons. The annotation of favourable nucleotides for transcription/ translation initiation in the vicinity of codons at position B (*metCPR5b*) indicates a higher probability of this site to be preferred over the first site (section 3.2.5). Thus, this preference could result in competition between these sites. For instance, the presence of three consecutive start codons at position B could have higher preference over the first start codon (position A), or equal probability of being transcribed/translated like the first one. Thus, it could be possible that disruption of the start codons in *metCPR5b* and *metCPR5bc* might have removed or reduced the preference competition between start codons, ultimately improving the efficiency of the first start codon, allowing the production of more full length protein.

In addition to isoform generation, the presence of multiple translation initiation sites also allows proteins to be targeted to different or multi-cellular locations, for instance, distinct POLγ1 isoforms were found to be targeted to mitochondria or plastids (Christensen et al., 2005). Similarly, translation initiating from the first AUG directs THI1 to the chloroplast, whereas, translation initiation from the second AUG, targets the protein in the mitochondria (Chabregas et al., 2003). Arabidopsis DNA ligase 1 (AtLIG1) carries three alternative translation initiation sites, and wildtype AtLIG1 is reported to be targeted to nucleus and mitochondria. The translation initiation from the first methionine was shown to be mitochondrial, whereas, from the second and third translation start sites, the protein was targeted to the nucleus only (Sunderland et al., 2004). For isoform-mediated multiple targeting, isoforms were shown to have signature sequences (signal peptides).

The plant mitochondrial pre-sequence (signal peptide) is generally enriched in arginine and serine residues in addition to leucine and alanine, in order to form an amphiphilic helix.

Similarly, the chloroplast signal peptide is enriched in serine and alanine, and is low in acidic amino acids (Christensen et al., 2005). As mentioned in Chapter 3, the first 300bp of At*CPR5* cDNA is rich in residues, such as, arginine, serine, leucine and alanine. Additionally, AtCPR5 is also predicted to be targeted to the nucleus, mitochondria, chloroplasts and/or plasma membrane when truncated versions of AtCPR5 protein were analysed using a number of subcellular localisation tools (data not shown). Thus, it is conceivable that AtCPR5 translation from multiple sites could target to multi-cellular locations and perform different functions. It could be possible that nucleotide changes (disruption of methionine) in *metCPR5b* could have resulted in the stability and abundance of a version of AtCPR5 that is desired in the regulation of leaf size. To date, there are no reports on AtCPR5 isoform production and the current thesis also lacks in studies showing AtCPR5 isoform formation and their localisation to different cellular organelles. Follow-up AtCPR5 localisation studies, including the identification of the CPR5 protein in various cellular fractions using CPR5 antibodies in *metCPR5* (*metCPR5b* and *metCPR5bc*) and *cpr5-2,* will reveal the implication of the putative start codons in AtCPR5 isoform production and/or dual or multi-targeting.

Similarly, the tuning of translation efficiency by genetic code substitutions during evolution helps different genes to tune their expression according to their desired levels. For example, some genes are required in higher amounts than others, whereas, the expression of others, such as, regulatory proteins tend to be low. However, more challenging are the genes which are needed to be translated at various (differential) levels under different conditions (Takagi et al., 2005, Lü et al., 2006, Ingolia et al., 2009). AtCPR5 has been shown to be implicated in a number of pathways. Thus, it could be possible that AtCPR5 is produced at different levels and from different start codons under different conditions. Nucleotide changes in *metCPR5b* could have improved translation efficiency and enriched the level of the desired AtCPR5 version. It could be possible that full length AtCPR5 protein is essential for leaf size regulation and its enrichment by nucleotide modifications have resulted in improved AtCPR5 transcript/ protein and eventually better growth.

The exact mechanism of how nucleotides associated with putative start codons improve leaf size is yet unknown and is recalcitrantly difficult, as the quantification of AtCPR5 and Atcpr5

proteins under a native promotor has been shown to be refractory, both by CPR5 antibody and mass-spectrometry (post thesis submission results). To conclude, the presence of multiple alternative start codons, translation initiating favourable residues and the annotation of multiple targeting of AtCPR5, reinforces that these three consecutive start codons somehow are important for AtCPR5 functions.

7.3.6 Putative roles of CPR5 in regulation of balance between growth and resistance

In *cpr5*, plant growth is severally compromised, for instance, *cpr5* plants are stunted in growth and have smaller leaves with aberrant trichomes (Bowling et al., 1997). On the other hand, *cpr5* plants display upregulated levels of *PR1* and *PR5*, and hence confer resistance to *P. syringae*, and *H. parasitica* pathogens. These results propose *cpr5* to be defective in growth regulation, due to the gain-of-resistance function. That CPR5 negatively regulates resistance in Arabidopsis (Yoshida et al., 2002) further reinforces the role of AtCPR5 in the regulation of resistance. The bigger leaves and enhanced susceptibility shown by *metCPR5b* plants lay the foundation of the hypothesis that AtCPR5 plays a central role in controlling the balance between growth and resistance.

The balance between plant growth and immunity is generally controlled through plant hormone crosstalk and manipulations in their signalling (Lozano-Duran and Zipfel, 2015). Phytohormones, such as, SA, JA, ET, and ABA including growth promoting hormones, brassinosteriods (BRs), gibberellins (GAs), auxins and cytokinins, play a major role in regulating growth and resistance trade-off (Lozano-Duran and Zipfel, 2015). SA is induced in response to a biotrophic pathogen attack, whereas, JA/ET responds to necrotrophic pathogen and wounding (Jing et al., 2007). SA-or JA/ET-mediated resistances were shown to be accompanied by a loss in plant growth (Song et al., 2014, Schmiesing et al., 2016). At the molecular level, the WRKY70 transcription factor is a common target of SA and JA during SA- and JA/ET-mediated resistances (Li et al., 2004). *WRKY70* overexpression results in enhanced and constitutive resistance to *PstDC3000*, combined with reduced expression of JA-responsive genes, such as, *PDF1.2*. *WRKY70* antisense suppression renders susceptibility to *PstDC3000* (Li et al., 2006, Li et al., 2004). SA activates WRKY70 either directly or through

NPR1, whereas, JA represses WRKY70 via COI1 (Li et al., 2006, Li et al., 2004). Recently, *CPR5* overexpression has resulted in the retention of NPR1 in cytoplasm and enhanced susceptibility to *P. syringae* (Gu et al., 2016). *cpr5* plants have higher SA and JA but wildtype levels of ethylene (Jing et al., 2007). The perception and integration of SA, ethylene and JA is proposed to be perturbed in *cpr5* plants (Jing *et al.*, 2007). Additionally, *cpr5* plants are shown to have elevated expression levels of *WRKY70* (post-thesis submission results). Thus, these results suggest that the upregulation of SA, JA and *WRKY70* could have led to the activation of resistance and repression of growth in *cpr5*.

In the current thesis, *cpr5-2*, as well as *SynCPR5* and *metCPR5* plants, were shown to have lower *PDF1.2* levels than the wildtype (Figure 7.4), indicating some unknown signalling intrusion, which appears to exert suppressive effects on *PDF1.2* transcript levels in *cpr5*, *SynCPR5* and *metCPR5*. *WRKY70* overexpression results in the activation of *PR1* but the downregulation of *PDF1.2* (Li et al., 2004), which suggests that SA and/or *WRKY70* could be upregulated in *SynCPR5* and *metCPR5*. This hypothesis appears to be in line with *cpr5* and *SynCPR5* plants, which show higher *PR1* as well as increased resistance. Contrarily, the execution of enhanced susceptibility by *metCPR5* plants is inconsistent with the proposed SA upregulated levels. On the other hand, JA can also induce *WRKY70* and SA-responsive genes via the activation of *AtMYB44* and *COI1*, i.e. JA induces *AtMYB44* TF that activates *WRKY70* through *COI1*, which subsequently activates SA-responsive genes and resistance (Shim et al., 2013). Since *cpr5* has upregulated levels of JA but wildtype-like levels of ethylene, JA could activate *WRKY70* via *AtMYB44* which, in turn, induces SA-mediated responsive genes (*PRs*) and supresses JA/ET signalling and JA-responsive genes, like *PDF1.2* (Li et al., 2006). This notion is again consistent with the *SynCPR5* scenario but is inconsistent with *metCPR5*, as *PR1* levels are repressed, or with wildtype-like levels in *metCPR5* lines. In addition to SA and JA, *AtMYB44* can also be induced by ABA and wounding, and AtMYB44 physically interacts with PLY8 (ABA receptor). As shown by Jing et al., (2007) *cpr5* plants show hypersensitivity to SA, JA and ABA exogenous applications, and exhibit arrested seed germination and reduced hypocotyl lengths. These results suggest that SA, JA and ABA misregulation could be the consequence of AtMYB44 and subsequent WRKY70 induction, and vice versa.

In addition to SA, WRKY70 and AtMYB44, WRKY11, WRKY15 and WRKY18 also play an important role in SA and JA/ET signalling crosstalk and were shown to be negative regulators of WRKY70 and basal resistance (Park et al., 2005, Journot-Catalino et al., 2006, Kim et al., 2006b). PTI (PAMP-triggered immunity) induction by *flg22*, induces *WRKY11*, *WRKY15*, and *WRKY18* which, in turn, are believed to suppress defence signalling in order to reduce loss of fitness (Lozano-Duran and Zipfel, 2015). BRs detection by BR-receptors and *flg22* recognition by the FLS-receptor, both target the same signalling cascade and share components, such as, BRI1. Downstream in the same signalling cascade, *BZR1* (one of the crucial TFs) activates *HBI1* TF, which promotes growth and/or *WRKY* (WRKY11, WRKY15 and WRKY18) TFs that negatively regulate plant immunity (Lozano-Duran et al., 2013; Lozano-Duran and Zipfel, 2015). Additionally, BZR1 is also proposed to interact with WRKY18, WRKY60, WRKY40, and WRKY70 in order to regulate the trade-off between growth and immunity (Lozano-Duran *et al.*, 2013; Figure 9.2). *WRKY40* or *WRKY70* both get induced by SA and activate SA-mediated resistance. Thus, it could be possible that *WRKY40/70* is somehow suppressed by *BZR1*, whereas *HBI1* gets upregulated in order to promote growth. Since CPR5 is shown to restrict NPR1, JAZ1 and ABI5 (Gu et al., 2016), thus, *metCPR5b* plants showing enhanced susceptibility, could have managed to upregulate *BZR1* and /or *HBI1* through the suppression of SA and/or *WRKY40/70* TFs (Figure 9.2) and eventually resulted in bigger leaves by the upregulations of growth promoting genes (*PRE*s). On the other hand, a mutation in *CPR5* (*cpr5*), results in SA activation. Thus it is tempting for me to propose that the upregulation of resistance and suppression of growth in *cpr5* could be the result of *WRKY40/70* activation (Figure 9.2). These results further shed light on the complexity of the mechanism of how SA, JA, ABA, *WRKY* (*WRKY 11, 15, 18, 40* and *70*), and/or *AtMYB44* TFs may coordinate in order to execute SA-mediated resistance and *PDF1.2* suppression in *cpr5*, the exact mechanism of which is yet unknown. However, current discussion allows me to propose AtCPR5 to be a SA suppressor or regulator, through which CPR5 regulates balance between growth and resistance.

DELLAs (transcription factors) are plant growth repressors and play an important role in the regulation of plant growth (Sun and Gubler, 2004). Gibberellins (GAs) degrade DELLAs, release repression of growth and promote plant growth (Sun and Gubler, 2004). DELLAs and

GAs both regulate a number of cell cycle-related genes in rice, such as, *CYCA1;1* and *CDKB2;1* (Fabian et al., 2000). Similarly, DELLAs and GAs activate several growth and cell cycle related genes in Arabidopsis (Ogawa et al., 2003). During vegetative growth in *Arabidopsis*, DELLAs were shown to control (mainly repress) growth through a number of cell cycle inhibitors e.g. KRP2, SIM1, SMR1, SMR2 etc. (Achard et al., 2009). AtCPR5 physically interacts with KRP2 and regulates immunity (Wang et al., 2014). *cpr5* plants have misregulated levels of a number of cell cycle genes, including KRP2 (Bao and Hua et al., 2014 and post-thesis submission results). Since *cpr5* and *SynCPR5* transgenic lines (having intermediate leaf and cell sizes) display upregulated levels of *KRP2* transcripts, these results indicate that *cpr5* and *SynCPR5* plants could have relatively higher levels of DELLAs, reduced levels of GAs or defects in the cell cycle. These results are in line with Gu et al., (2016) who found suppression in a large number of auxin and GAs genes in *cpr5*, which indicates a higher accumulation of DELLAs. Keeping in view these results, it is tempting to speculate that AtCPR5 could regulate growth, possibly by interacting and subsequently repressing DELLAs. In the absence of functional CPR5 (*cpr5*), DELLAs could be stabilized to promote the activation of cell cycle inhibitors (KRP2) and repress cell proliferation and expansion.

The overexpression of the gibberellin synthetic gene (*AtGA20OXI*) resulted in an increase of cell division, as well as cell expansion in Arabidopsis leaves (Gonzalez et al., 2010). The presence of 16C cells in *metCPR5* lines indicates that *metCPR5* were advanced in endocycles, as compared to the wildtype and OE1 lines. Like ploidy levels, epidermal cells were also bigger on *metCPR5*. Thus, the bigger cells and leaves on *metCPR5* plants are an indication of higher GA levels, which could have resulted in increased ploidy levels and expansion.

In conclusion, this current study establishes a new role of AtCPR5 in the regulation of balance between growth and resistance. Taken altogether, it can be proposed that AtCPR5 regulates growth and resistance through SA suppression. Additionally, SA antagonism is proposed to be a reason of reduced *PDF1.2* levels. The present study identifies a novel role of CPR5 which is responsible for enhanced growth but reduced immunity. AtCPR5 appears to be a key gene that could be a target for pathogens, which upon recognition, activate SA and SA-mediated resistance and suppress growth. Lastly, how the genetic changes or disruption of methionine

residues of putative transcription or translation initiation sites regulates leaf size is still obscure and needs further investigation.

Chapter 8 AtCPR5 putative NLS clusters appear to have limited function in CPR5 nuclear localisation and may be part of a structural component of the Intrinsically Disordered Region

ABSTRACT

The *CONSTITUTIVE EXPRESSER* of *PATHOGENESIS-RELATED GENES 5* (*CPR5*) is proposed to be a master regulator of many integral processes in *Arabidopsis*. CPR5 was suggested to be a nuclear protein, since CPR5 was predicted to contain a bipartite nuclear localisation signal (NLS) in its N-terminus (41-61 aa). However, studies have not shown the role of the predicted NLS clusters in CPR5 localisation or function. Therefore, the CPR5 protein was analysed for the confirmation of the predicted NLS. In contrast to previous results, the current results showed that CPR5 protein contains three putative NLS clusters, which could act as a mono-, bi- or tripartite NLS. Putative NLS clusters were mutated individually (*nlsCPR5A* and *nlsCPR5B*) and collectively (*nlsCPR5AB*, *nlsCPR5BC* and *nlsCPR5ABC*) to test the role of each cluster in CPR5 nuclear localisation. Results show that collective mutations in NLS clusters, such as, *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* have severe effects on *CPR5* functions, compared to *nlsCPR5A* and *nlsCPR5B* lines, in which NLS-encoding clusters were mutated individually. *nlsCPR5A* and *nlsCPR5B* lines displayed wildtype-like trichomes compared to aberrant trichomes on the *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* lines. *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* showed early bolting compared to *nlsCPR5A*, *nlsCPR5B,* wildtype and *cpr5-2* plants. However, upregulated levels of *PR* genes and disease resistance was not complemented in any of the *nlsCPR5* lines. In contrast to the expected *cpr5*-like phenotypes, the complementation of some of the phenotypes in the *nlsCPR5* lines suggests that NLS clusters may not be implicated in CPR5 localisation and may have limited functions. In addition to NLS clusters, the first 98 amino acids of CPR5 (including NLS-encoding clusters) were predicted as an intrinsically disordered region (IDR). Therefore, CPR5 IDR was truncated in three parts; 1-37aa (*Del37CPR5*), 1-63 aa (*Del63CPR5*) and 1-98 aa (*Del98CPR5*).

Del37CPR5 lines displayed phenotypes similar to the *nlsCPR5A* and *nlsCPR5B* lines i.e. *Del37CPR5* leaves had wildtype-like trichomes without any HR or early leaf yellowing. Similarly, *Del63CPR5* lines displayed phenotypes similar to the phenotypes of *nlsCPR5AB*, *nlsCPR5BC* and *nlsCPR5ABC* lines. Based on the similarities in phenotypes, it is tempting to propose that, 1) NLS clusters may function as part of an intrinsically disordered region, rather than as an NLS or 2) the NLS may have a dual function. Additionally, these results show that putative CPR5 IDRs appear to be crucial for CPR5 proper functions.

8.1 INTRODUCTION

The central dogma of gene expression illustrates that a gene is transcribed in the nucleus, then translated into a protein in the cytoplasm and then is translocated to its destined cellular compartment(s), such as, the nucleus, chloroplast, endoplasmic reticulum and vesicles for proper function or interaction (Crick, 1970, Pemberton et al., 1998, Kosugi et al., 2009, Guo, 2014). The proteins being translocated to either of the aforementioned cellular compartments diffuse passively or are carried over actively. For active translocation to various cellular compartments, proteins contain specific signals, known as signal peptides, which are recognized by their import or export machinery to facilitate or mediate the localisation of the proteins in their respective organelles (Pemberton et al., 1998, Moroianu, 1999, Pemberton and Paschal, 2005). For instance, nuclear proteins have a short stretch of positively charged peptides, lysine and arginine, known as nuclear localisation signals (NLS). Likewise, chloroplast targeting proteins have signal peptides that guide the proteins to the chloroplast (kosugi *et al.*, 2009).

A large number of proteins can be translocated in to the nucleus passively, for instance, smaller sized (40-60 kDa) macromolecules, including histone proteins (Breeuwer and Goldfarb, 1990, Jakel et al., 1999), different types of RNAs (tRNA); (Arts et al., 1998, Kutay et al., 1998, Zasloff, 1983) and transcription factors are shuttled between the cytoplasm and nucleus passively (Bonner, 1978, Mattaj and Englmeier, 1998, Gorlich and Kutay, 1999, Mallet and Bachand, 2013) through the nuclear pore, which has a size of 9.0 nm in diameter (Moroianu, 1999). On the other hand, bigger proteins require an opening of 25 nm in diameter or translocate via active transport (Moroianu, 1999).

The active transport of proteins includes proper cargo systems, such as, signals, receptors, and the energy (Ran-GTPase) needed for transportation (Kalderon et al., 1984, Lanford and Butel, 1984, Dingwall and Laskey, 1991, Robbins et al., 1991, Gorlich and Kutay, 1999). For active nuclear transport, the NLS is recognised by receptors called importins or karyopherins, which mediate the nuclear import of proteins. Nuclear import is mediated when a ternary complex is formed upon the binding of karyopherins with the NLS and Ran-GTP (Weis, 2003, Pemberton

and Paschal, 2005). The asymmetrical distribution of Ran-GTP and Ran-GDP controls the entry and exit of proteins between the nucleus and the cytoplasm (Kalab et al., 2002).

Nuclear localisation signals (NLSs) have been classified into classical and non-classical (non-canonical) NLSs. The classical NLS comprises of a further two sub-groups; monopartite and bipartite (bi-NLS) (Mallet and Bachand, 2013). Monopartite NLSs contain a single cluster of basic amino acid residues, whereas bipartite NLSs include two clusters of basic amino acids separated by a 10–12-amino acid linker (Lange et al., 2007). Typical monopartite NLSs contain four consecutive basic amino acids, such as, SV40 large T antigen NLS (PKKKRKV) (Kalderon et al., 1984), or three basic amino acids, such as, c-Myc NLS (PAAKRVKLD) ((Dang and Lee, 1988, Dingwall and Laskey, 1991, Robbins et al., 1991, Kosugi et al., 2009). On the other hand, a typical bipartite NLS, such as, the nucleoplasmin NLS, is divided into two parts separated by a linker, such as, KRPAATKKAGQAKKKK (Lange et al., 2007).

Contrarily, non-classical NLSs contain protein-tyrosine residues (PY-NLS), which are comprised of 3 regions or epitopes (Mallet and Bachand, 2013). The first region is composed of 4-20 residues enriched in basic residues or hydrophobic motifs. Whereas, the second region is found to contain arginine, histidine or lysine residues and is present 2-5 residues upstream of the third region. The third region is well conserved and contains proline-tyrosine residues. Though a few examples of PY-NLSs have been demonstrated experimentally, *in silico* predictors suggest that there is a substantial fraction of proteins, which carry a signature for the PY-NLS mediated nuclear import (Lee et al., 2006, Lange et al., 2007).

CPR5 is a nuclear as well as a membrane protein (Wang et al., 2014). Therefore any change in the putative NLS cluster(s) may alter or destroy CPR5 native localisation and disrupt CPR5 normal functioning. As mentioned in Chapter 3, CPR5 is predicted to contain a monopartite or bipartite NLS in the N-terminal half of the protein. In the current study, all the putative clusters of CPR5 NLS were altered in order to study their effect(s) on CPR5 functioning. *nlsCPR5* transgenes carrying mutations in the NLS-cluster(s) were transformed in *cpr5-2*. Plant growth phenotypes of the obtained *nlsCPR5* transgenic lines were compared and evualated on the basis of restoration of wildtype-like phenotypes, which managed to be rescued in *cpr5-2*

complemented plants, for example, normal trichomes, leaf area, epidermal pavement cell size, plant height and drought resistance, HR-like lesions, leaf yellowing and susceptibility to bacterial pathogen (*Pseudomonas syringae*). Results indicate that mutations in the putative NLS-cluster(s) greatly affect CPR5 proper function. However, their phenotypes were shown to be intermediate between *cpr5* and CPR5 plants. Contrary to our expectation, none of the *nlsCPR5* transgenes compeletely abolished CPR5 function in the *cpr5* manner. The partial complementation by *nlsCPR5* prompts me to propose that putative NLS-clusters are important for CPR5 functioning and could be part of some other structural motifs, such as, intrinsically disordered regions (IDRs).

8.2 RESULTS

IN SILICO CHARACTERISATION

8.2.1 AtCPR5 is predicted to contain 3 NLS-encoding clusters

AtCPR5 has been predicted to contain a bipartite nuclear localisation signal (NLS) (Yoshida et al., 2002) and was shown to be localised in the nucleus by Perazza et al., (2011) and in the nuclear membrane by Wang et al., (2014). AtCPR5 was originally proposed to embed in the outer nuclear membrane, which is cleaved off and then translocated to the nucleus via the NLS-mediated nuclear import (Kirik et al., 2001; Yoshida et al., 2002). However, none of these studies (Kirik et al., 2001, Yoshida et al., 2002, Perazza et al., 2011, Wang et al., 2014) provides any functional evidence of the NLS being implicated in the AtCPR5 nuclear import. In order to validate the previous prediction about the prediction of a bipartite NLS, an extensive *in silico* analysis of the AtCPR5 protein using the latest computer tools was done. A close and careful examination of the AtCPR5 N-terminus shows the presence of a number of clusters of positively charged residues. Thus, it was hypothesised that these positively charged amino acids could act as an NLS. To check the possibility of positively charged stretches as a putative NLS, the amino acid sequence of AtCPR5 was analysed by NLS-prediction *in silico* tools.

As shown in Figure 8.1, the AtCPR5 N-terminus displays three clusters of positively charged residues (pink-coloured bars). These clusters are enriched with lysine (K) and arginine (R) residues. Region-A contains four lysine residues KKKK at positions 41-44, region-B contains KRK residues at 53-55 and region-C carries KKK at the 59-61 positions of the protein (Figure 8.1). In addition to manual scanning, the AtCPR5 protein sequence was also analysed by two different computer logarithms trained to predict a nuclear localisation signal (NLS). These tools proposed AtCPR5 to be comprised of two putative regions that contain the consensus sequence for the NLS; the first ranges from 41-50 (regions A and B together; Figure 8.1), whereas the second, ranges from 59-61 (region-C; Figure 8.1). According to annotation, both of these regions could act as a monopartite NLS (data not shown). Concomitantly, the same tools also propose AtCPR5 to contain a bipartite NLS; the first part of which (bipartite NLS) ranges from 41-44 (region-A), whereas the second half ranges from 53-55 (region-B) amino acids of the

AtCPR5 protein (Figure 8.1). In conclusion, the manual and *in silico* predictions indicate that the AtCPR5 protein contains 3 clusters of basic positively charged amino acids that could act as a monopartite, bipartite NLS or both.

8.2.2 Putative NLS clusters are highly conserved among AtCPR5 homologues

As mentioned in Section 8.2.1, AtCPR5 is shown to be a part of the genomes of many plant species. AtCPR5 homologues were shown to have a significant amount of homology among their protein sequences (Figure 8.2). Keeping in view homology, two questions could be raised; (**1**) how many of the AtCPR5 homologues found in the database contain NLS-like clusters? and, (**2**) to what extent are these NLS-like clusters conserved in AtCPR5 homologues? Thus, it was hypothesised that putative NLS-like clusters are present in AtCPR5 homologues. To test this hypothesis, the protein sequences of AtCPR5 homologues were aligned and investigated in order to find out the presence and position of predicted NLS or NLS-like clusters in AtCPR5 and its homologues. AtCPR5 homologues were divided into monocot and dicot groups.

8.2.2.1 Majority of CPR5-like proteins from dicots are predicted to contain NLS-like clusters

Arabidopsis thaliana is a dicot and belongs to the brassicaceae family. Thus, firstly AtCPR5 was compared with close relatives i.e. homologues belonging to the same family. As shown in Figure 8.1, AtCPR5 close relatives, such as, *Brassica rapa*, *Capsella rubella* and *Eutrema salsugineum* are annotated to contain NLS-like clusters. Moreover, the positions of the annotated NLS-like clusters in these AtCPR5 homologues were also found to be highly conserved (Figure 8.1 and 8.2A). In addition to the NLS clusters, even the spacer residues, such as, serine (S) and proline (P), present between the clusters (region-A and region-B) of the AtCPR5 annotated bipartite NLS residues, are also found to be conserved among AtCPR5 homologues (Figure 8.1A and 8.2A).

In addition to the comparison between AtCPR5 and its close relatives, putative NLS-encoding cluster(s) were also shown to be present in the majority of the protein sequences of AtCPR5 homologues from remaining dicot plant species (Figure 8.2B). However, the typical bipartite pattern of NLS clusters as found to be present in AtCPR5, was missing in several homologues

(Figure 8.2). Additionally, the clusters or stretches of NLS-encoding residues in GsCPR5 (*G. soja*), PvCPR5 (*P. vulgaris*), RcCPR5 (*R. communis*) and CaCPR5 (*C. arietinum*) homologues, were found to be separated by non-conserved amino acids. Alternatively, the predicted NLS of some of these homologues comprise of a single stretch of consecutive basic residues e.g. CmCPR5 (*C. melo*), CsCPR5 (*C. sativus*), NnCPR5 (*N. nucifera*), FvCPR5 (*F. vesca*), and CcCPR5 (*C. clementina*) etc. (Figure 8.1A and 8.2A). There are also a few AtCPR5-like proteins, which were annotated to possess none of the NLS-encoding-like clusters e.g. GmCPR5 (*G. max*), AtCPR5 (*A trichopoda*), BvCPR5 (*B. vulgaris*) and MtCPR5 (*M. truncatula*).

8.2.2.2 Majority of CPR5 homologues from monocots are predicted to contain no NLS

In contrast to AtCPR5 dicot homologues, the majority of CPR5-like proteins from the monocot plant species were predicted to contain no NLS-like region (Figure 8.2). Of the CPR5-like protein sequences from the monocot plant species, only VvCPR5 (*V. vinifera*), ThCPR5 (*T. hasslerana*), CcCPR5 (*C. canephora*), SICPR5 (*S. Indicum*) and TcCPR5 (*T. cacao*), were annotated to carry NLS-encoding residues. For example, the NLS-encoding sequence from VvCPR5 contains a stretch of 8 consecutive lysine residues, CcCPR5 (*C. canephora*) consists of 5, and TcCPR5 (*T. cacao*) has 4 (Figure 8.2B). In addition to the homomologues from the monocots and dicots, *Phycomitrella patens* (Moss; PpCPR5) was found to have no NLS-like cluster.

Taken altogether, it can be concluded that the majority of CPR5-like proteins from dicots carry annotated NLS-encoding regions, whereas, only a few CPR5-like proteins from the monocot group of plants are predicted to contain NLS-encoding regions.

A.

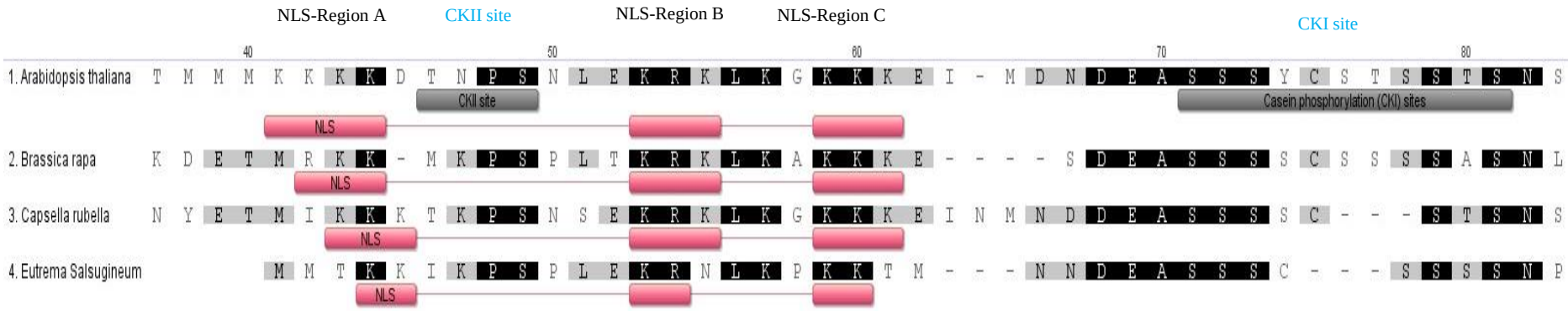


Figure 8.1 Positions of potential nuclear localisation signal (NLS) facilitating basic residues in AtCPR5 and its homologues

This figure shows the positions of positive residues, such as, lysine (K) and arginine (R), which are known to be part of the signals for nuclear localisation recepters, as idenified both by manual scanning and through computational tools (SeqNLS and PSORTII), in the protein sequences of Arabidopis CPR5 and its homologues. The pink bars shows the position of annotated NLS-encoding residues, whereas the grey bars represent the posotion of predicted casien kinase phosphorylation sites in some of the AtCPR5 homologues.

A.


A. thaliana MEALLLPPSPEPQNQITNPANSKPNHQSGDVHKDETMMMKKKKDTNPSNLEKRKLKGKKKEIMDNDEASSSYCSTSSTSNSNSTKRVTRVVHRLRNPMRLGM
C. sativa MEALLLLSSSPEPQKSITDPANSKSDHQSDEEVKDETMMKKKKKKTKQSDSGKRKHKGKKKEMNNDEASSSSCSISSTSNSNSTRRVSRVAHRLRNPTVRLG
C. rubella MEALLLLSPSPEPQNPITDPANSKSHHQSDEEVNYETMIKKKTKPSNSEKRKLKGKKKEISSSSCSTSNSNSTRRVSRVAHRLRNPTVRLGMSRRSVGERQA
E. salsugineum MMTKKIKPSPLEKRNLKPKKTMNNDEASSSCSSSSNPNYTRRVSRVAYRLRNPTVRLGMARRSVGERQAEALARPLVYIICSVYILFMRYLTFVHMVLQRNA
B. rapa MDALLLPPSPDPEESITDPANQRANHRPENSSDELKDETMRKKMKPSPLTKRKLKAKKKESDEASSSSCSSSSASNLNSTRRVSRVAHRLRSPPVRLGFTRR
B. napus MDALLLPPSPDPEKSFTDPANQGANHRPEDSSDELKDETMRKKMKPSPLTKRKLKAKKKESDEASSSSCSSSYASNLNSTRRVSRVAHRLRSPSVRLGFTRR
G. soja MAEIQNCSSEPEPTAINECESQASMKSTTQSSEASDTTSSRKENKGKGKAAAFKRRNPRVLVRRHRANNVDTIGLPLGMSFAAVMAQVMYRRDVAAESMSPS
P. vulgaris MAEIEDCSSEPTPINQCPSQISINDGMRSSEVSEGSSKRRGKSKRVLLKRRNPSVVVRRHRTNNVNTIGLLLGISFAAVTAQVLYKRDAAAESISPNHLSKM
C. melo MDEPPTCATTYDPLVDETTDDDREDIAAVYSSPGNSTGTLYDPPPCSNRNRKKKKKKKKRVAIDNVQVLPSSSSSCSSPARLLQRGVNCKRRRPKVVIAQPR
C. sativus MDEPPTCATTYSPFVDETTDDDREDIAAVDSSPGNSTRTLYDPPPCSNRNRKKKKKKKRVVIDNVQVLPSSSSSSSCSSPARLLQRGVNCKRRRPKVIIAQ
E. grandis MDAAAAAVPPPPPAGEPAPSLSPPIAAQEPIDAPNSPDRGGPSPPAEAPSPPSKPVGEKANVKKRRKLKGVASQAAAAASSSSASSSSARPLLPRPAERGAR
N. nucifera MEVATSSLATESTSPGFTLTTEDEVEEVGKGPTVQTTANHFNRTVEVPPADANAELCSSPCQNSIPINNGGTVIRKKKKKNKELTDFSSSGSGASSSLSYFS
F. vesca MFRFLPHSPSSYLNSLLNAVPFPSSHLSMDPPPPAPPPQPQISAVQSTTPRHENPIAAVENPKKKGKKKLFKDSASVSSSSSSASTSTSGFSVRGAMRVACK
R. communis MEPPPSVTPSNSTEHSNPTADNPSCLPTMIKNKNRKKKKVFKTCAQTSSSSSAATRFPYKRRNSKVVLAPVRRGGRSFGDSQMEAVALPLGMSFAAIIAKFM
C. arietinum MDVLTTLSSSSSSSDCNSNVSLRTPEHITRRKKLHEATSYHASPKGKCKRRKIVRVLIRRKRTDHVSTIGFPLGMSFAAVMAEVLYRRDAAAAHTLSPTHIS
S. lycopersicum MDHPPLSQPVDPPPVTTADAMASEILPESLNSGVMKKKTKKKKKNKACVTDPFDPHSSAASYSSSSCSATSVPPSTQRGIRVSTGRRNPRVIISSGRQKSDN
S. tubersoum MDHPPLSQPVDPPPETAGDAMASEILPESSNSGVMKKKKTKKKKKNKACVTDPFDPHSSAASYSSSSCSATSFPPSNQRGIRVSTGRRNPRVIISSGRQKSD
C. clementina MDTPRPPSSSSSPPPPPSPRCQQSTTCVSTENPLNPFAIDPNPTVDGSSAPEKRVRYEKKKKNRFFKDAASSSSSRPSCFARSGTRVASRRRSMRVDRLGPI
G. raimondii MDAMPPSSSPLSLQPNGSSACASIIPYKHDSNPTLDNHIPSSKPLIRKKRRKTVDRDAPSSSSCSAYSSSVQKGMRLSSKRRNLRVRFGPVRRADVRDVDSI
G. arboreum MDAMPSSSSPLSLQPNGSSACASIIPYNHDSNPTLDNHIPSSKPLIRKKRRKTVDRDAPSSSSCSAYSSSVQKGMRLSSKRRNLRVRFGPVRRADVRDVDSL
P. mume MDTPSSSPPSLQLLASKSCVLSDQADTVDHENAATATTVDANPNPSSRTVRNLSPCDESPAKPICRNTKKTKRKVLKDAAAAPSSSCGSTSSGPSFRGTRVA
P. euphratica MERTAPSPSLSTQDPTPTNHSNPTVENDCYPKTINKRKKKGVFKDTAMTPPPSQPTSSSSSSSSSSSSSVNKSPRVSHKRRSPKVVFAPVRRRRGIGDDGVE
P. tricocarpa MDRTAPSPSLSTQEPTPTNHPNPTVENDCNPKTINKRKKKRVFKDTAITPPPSQPTSSSSSSSSVNKSPRVSHKHRSPKVVFAPVRRRGGIGDDGVEAIALP
N. sylvestris MDHHPLSQPVDPPPEPTRDTMATEIPPESSNSGGVMKNKTKKKKKKNKACVNDPFDPHASSAASYSSSSCSTTWFSQRGIRVSASRRNPRVVIGAGRQKSDE
N. tomentosiformis MDHHHPPLSQPVDPPPEPTSDTMAKEIPPESSDSGGVMKNKPKKKKKKNKSCDNDLFDPQSSSAASYAPSSCSTTSFSQRGIRISASRRNPRVVIGAGRQKS
J. curcaas MGTPSPSASPPIQESNTTADSIPEDESPPAISKPLEDDSLNMTHRKKKKKKKKAYEGVSMISSSSSSPSTRRSTSMTYKRRNPKVLIAPVRRGGRFGNSDME
B. vulgaris MEINHSPSSPLHHSLSPNPPPPEPSDFSLKPQINGPKLTRKIKKRYPDEASSASSSSSIQRGIRIPNKRRNPRFSVRRNATDVEAIAFPLGMSIAAVLAQIL
A. trichopoda MQSLSHGFTACTNKSSRQGKESLRIDKEVGYAKDGDFLSLTNSPTPSSSSICTVADDYIEVCLQGFGNRGSKDMGECSSGILGEVRDKSGVKTDRFLSFETA
M. Truncatula MLLLFIETPMAEIETCSSQSASKNRAGLEICETNSYNSSQLQRTKVKGKGIACKRRNPRVSVRRTRGNNVAAIGFPLGMSVAAVMAQVLYRRDAAAEGMSPD
G. max MAEIQNCSSEPEPTAINEWEGQASMKSTTQSSEASDTTSSRKENKGKGKAAAFKRRNPRVLVRRHRANNVDTIGLPLGMSFAAVMAQVMYRRDVAAESMSPS

B.

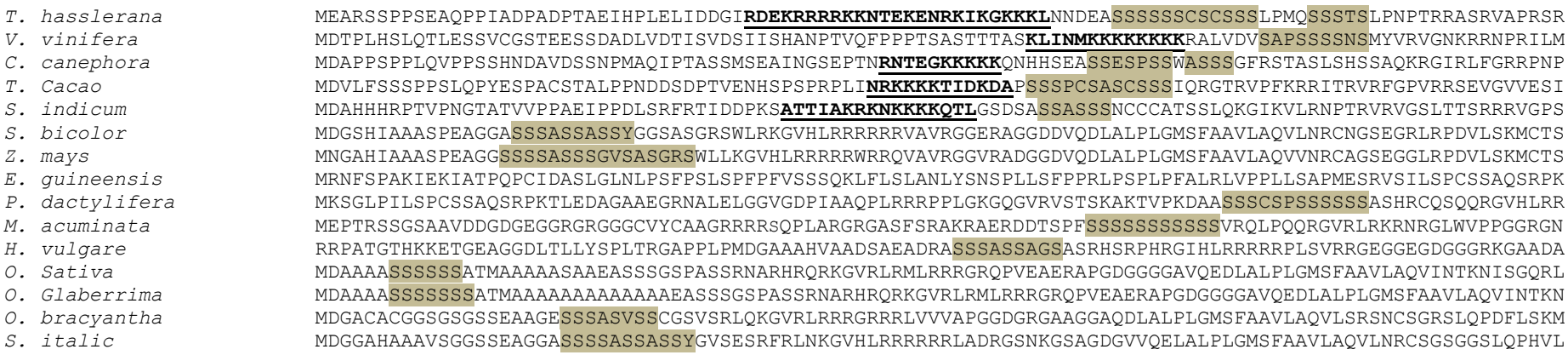


Figure 8.2 Position of annotated NLS clusters and casein kinase phosphorylations sites among CPR5-like proteins

This figure shows the position of annotated NLS clusters among CPR5-like protein sequences from dicots (**A**) and monocots (**B**) using two different predicting tools, SeqNLS and PSORTII, as mentioned in the methods. The underlined residues show the positions of annotated NLS clusters. The residues highlighted in the dark grey color show the positions of conserved regions for casein kinase phosphorylation sites.

8.2.3 CPR5 putative NLS-encoding clusters are predicted to be flanked by CKI and CKII phosphorylation sites

A number of nuclear proteins carry casein kinase phosphorylation site(s) either flanking or in the vicinity of NLS clusters in order to facilitate NLS-mediated nuclear localisation through phosphorylation (Harreman et al., 2004). Thus, AtCPR5 protein sequence was investigated in order to find out the presence of phosphorylation sites in the vicinity of the annotated NLS. As AtCPR5 is proposed to contain a bipartite NLS, it is likely that the AtCPR5 NLS will contain CKI and/or CKII phosphorlation sites. To test this hypothesis, amino acid sequences of proteins from AtCPR5 and its homologues were analysed and compared in order to confirm the presence of CKI and CKII phosphorylation site(s).

As expected, the AtCPR5 protein was predicted to contain a casein kinase (CK) I and a CKII phosphorylation site. The predicted CKII site resides at 46-49 positions between the annotated NLS cluster-A and cluster-B, whereas the CKI site locates at 71-81 positions (downstream of the annotated NLS cluster-C) in the CPR5 protein (Figure 8.1). It can be seen that putative CKI and CKII phosphorylation sites essentially flank the annotated NLS clusters, as suggested by Harreman et al., (2004). Additionally, both CKI and CKII phosphorylation sites are present in dicot CPR5 homologues, which are highly conserved (Figure 8.2). Surprisingly, monocot homologues are predicted to carry only one CK1 phosphorylation site. Based on these results, it can be documented that AtCPR5 is predicted to contain CKI and CKII phosphorylations sites, which are conserved among several of the CPR5 homologues.

PHYSIOLOGICAL CHARACTERISATION:

8.2.4 *nlsCPR5* complemented lines show variation in grwoth

In the current study, NLS-encoding clusters were mutated individually as well as collectively, and each transgene carries mutations of 2-4 residues in order to disrupt the putative NLS sites. Having a strong annotation and resemblance with the classical NLS-clusters, it was hypothesised that the *nlsCPR5* transgenic plants will show aberrant trichomes, spontaneous lesions and early leaf senescence in the c*pr5* fashion. To test this hypothesis, *nlsCPR5*

transgenes were complemented in *cpr5-2* plants, and complemented lines were analysed in order to check the restoration status of trichomes, HR-like lesions and early leaf senescence on *nlsCPR5*. In contrast to *cpr5*-like phenotypes, *nlsCPR5* plants displayed a range of phenotypes. For example, *nlcpCPR5A* and *nlsCPR5B* plants exhibited wildtype-like leaves with trichomes (Figure 8.3). The leaves of *nlsCPR5AB*, *nlsCPR5BC* and *nlsCPR5ABC* plants were shown to have no or reduced number of appendages or trichomes. *nlsCPR5A*, *nlsCPR5B*, *nlsCPR5BC* and *nlsCPR5ABC* plants showed no HR-like lesions (red arrow) and early leaf senescence (white arrow) when compared to *cpr5-2* (Figure 8.3). In conclusion, these results do not support my hypothesis, since *nlsCPR5* lines show a range of phenotypes i.e. partial complementation of phenotypes.

8.2.5 *nlsCPR5* plants show reduced *nlsCPR5* expression levels

As shown in Chapter 6, *SynCPR5* and *metCPR5* transgenic lines showed a range of transcript abundance of their transgenes. Thus, it was assumed that like *SynCPR5* and *metCPR5*, *nlsCPR5* transgenic lines will also show different levels of their transcripts. To confirm, the relative transcript abundance of the transgenes in *nlsCPR5* complemented *cpr5-2* plants, was measured using real-time qRT-PCR.

Similar to *SynCPR5* and *metCPR5* lines, *nlsCPR5* lines also displayed variable expression levels, which were found to be varying intra- (among the lines of same construct) and inter-*nlsCPR5* lines (between different *nlsCPR5* constructs) (Figure 8.4). For example, the *nlsCPR5BC* and *Del63CPR5* lines were shown to have transcript abundances indifferent from *cpr5-2* levels (Figure 8.4). *nlsCPR5AB* had higher transcript abundance than *cpr5*, *nlsCPR5BC*, and *Del63CPR5*, but less than *nlsCPR5A*, *nlsCPR5B* and the wildtype (Figure 8.4). Similarly, *nlsCPR5A* lines had higher expression levels than *cpr5*, *nlsCPR5AB*, *nlsCPR5BC* and *Del63CPR5*, but lower than *nlsCPR5B* and the wildtype. In summary, *nlsCPR5* plants displayed variable expression levels.

Figure 8.3 *nlsCPR5* and *DelCPR5* plants at 17 DAS

Plants were sown in soil under long day conditions (14:10 hours of light:dark). Plants were photographed by camera fixed on a stand. The small dots on the *cpr5* leaves represent spontaneous lesions, which are missing from the wildtype and some of the *nlsCPR5* lines. Small hair-like structures on the leaves represent trichomes. The white arrow represents early yellowing or senescence, whereas the red arrow depicts a lesion on *cpr5* leaves.

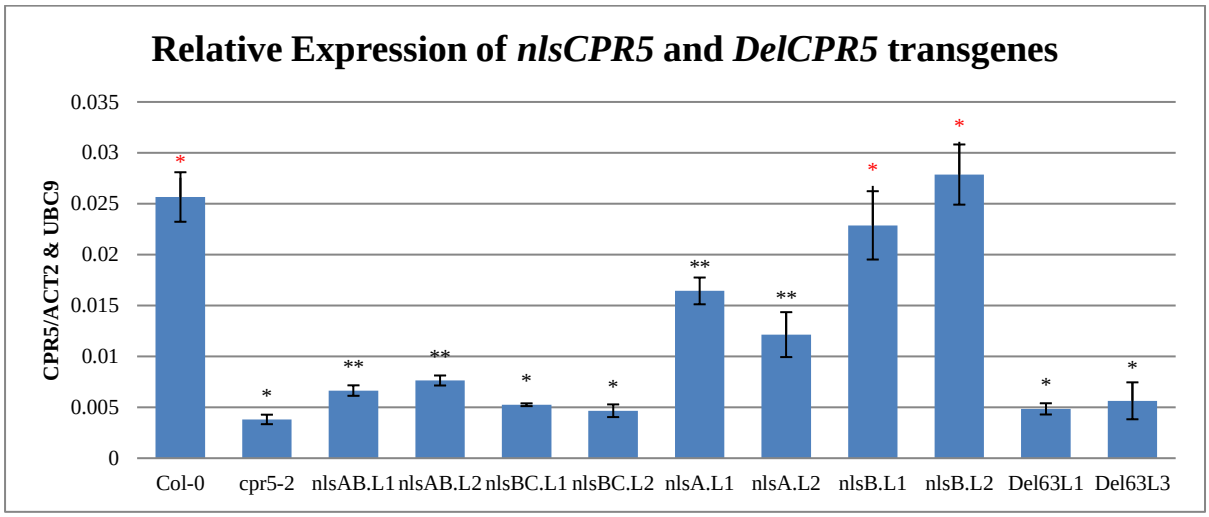


Figure 8.4 Transcript abundance of *CPR5* in *nlsCPR5* and *DelCPR5* transgenic lines

This figure shows the relative transcript levels of *nlsCPR5* and *DelCPR5* (*Del63CPR5*) transgenes using real-time quantitative reverse transcriptase, qRT-PCR. Total RNA was isolated from 31 day old plants. Values were normalized to the expression of *ACT2* and *UBC9*. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant compared to wildtype, * = significant compared to *cpr5-2*; ** = significant compared to wildtype Col-0 and *cpr5-2*. See Appendix 11.3.6 for further t-test comparisons among *nlsCPR5* lines. nlsA = *nlsCPR5A*, nlsB = *nlsCPR5B*, nlsAB = *nlsCPR5AB*, nlsBC = *nlsCPR5BC*, and nlsABC = *nlsCPR5ABC*, Del63= *Del63CPR5*.

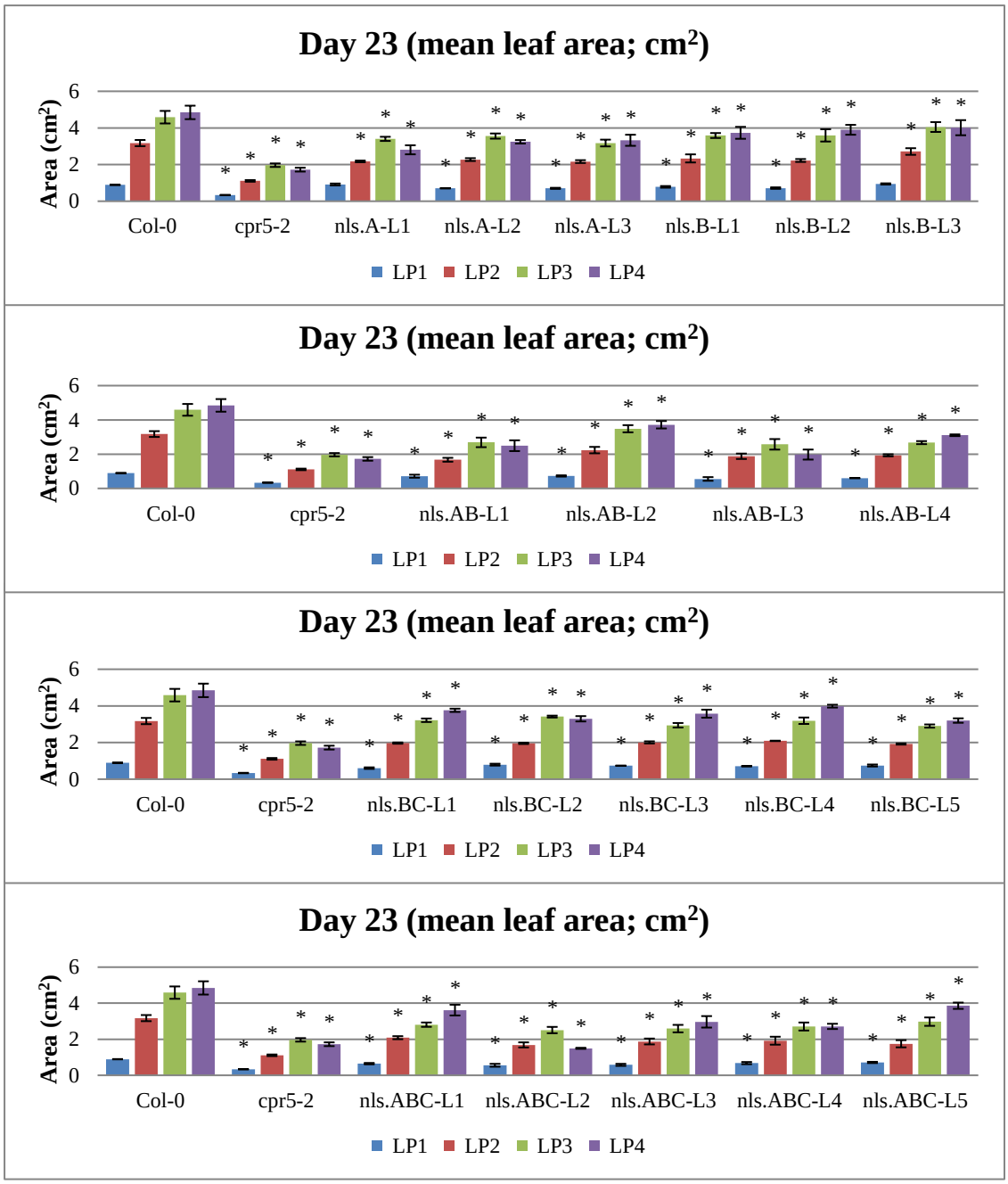


Figure 8.5 Mean leaf areas of *nlsCPR5* lines

This figure shows the leaf area of the first four rosette leaf pairs (first eight rosette leaves) of *nlsCPR5* at 23 DAS. The bars represent the standard error and the asterisk indicates the level of statistical significance at $p < 0.05$ (Students' t-test) compared to the wildtype. nlsA = *nlsCPR5A*, nlsB = *nlsCPR5B*, nlsAB = *nlsCPR5AB*, nlsBC = *nlsCPR5BC*, and nlsABC = *nlsCPR5ABC*.

8.2.6 *nlsCPR5* transgenic lines display smaller leaves

Plant growth and development is shown to be compromised in all of the *CPR5* mutants, such as, *cpr5-2*, *hys1* and *old1* (Bowling et al., 1997; Yoshida et al., 2002; Jing et al., 2002). As *nlsCPR5* plants show reduced and variable transcript abundances, the *nlsCPR5* plants were expected to show intermediate phenotypes. Therefore, it was hypothesized that mutation(s) in the proposed NLS clusters would affect leaf size. To test the hypothesis, the leaf areas (in the form of pairs of leaves as measured for *SynCPR5* and *metCPR5* plants) of *nlsCPR5* lines from each group (construct) were measured, in order to see any difference in leaf size between *nlsCPR5* and the wildtype. As expected, the majority of the *nlsCPR5* transgenic lines displayed smaller leaves compared to the wildtype at 23 DAS (Figure 8.5). Except LP1 on *nlsCPR5AL1* and *nlsCPR5BL3*, all of the first four leaf pairs on the *nlsCPR5* lines, including, *nlsCPR5A* and *nlsCPR5B*, were significantly smaller than the wildtype at 23 DAS. In conclusion, these results demonstrated that mutations in putative NLS-encoding clusters have resulted in variable expression levels and smaller leaves on *nlsCPR5* transgenic lines.

8.2.7 Epidermal pavement cells are smaller on *nlsCPR5* transgenic leaves

Cell division and endoreduplication (increase in ploidy levels) positively correlate with final leaf expansion and size (Sugimoto-Shirasu and Roberts, 2003). As shown above, the majority of the *nlsCPR5* mutants produced leaves of smaller sizes compared to the wildtype at 23 DAS. Therefore, it was hypothesised that epidermal cells of *nlsCPR5* transgenic plants will be smaller compared to the wildtype. To test this hypothesis, the sizes of epidermal pavement cells from leaves of *nlsCPR5* plants were measured by taking images of leaves using scanning electron microscopy (Figure 8.6A). Since leaf size differences become more pronounced from the second leaf pair, the epidermal pavement cells of the second leaf pair (third and fourth rosette leaves) and third leaf pair (fifth and sixth rosette leaves) were measured. Additionally, the cell areas of rosette leaf 3 were shown to be indifferent from leaf 4 cells. The cell areas of leaf 5 were similar to leaf 6 cells from Col-0 and *nlsCPR5* mutants. Therefore, only the areas of rosette leaf 3 and leaf 6 are shown here. The average cell size of the third and sixth rosette leaves of Col-0 was found to be ~4800 and ~4700 μm^2, respectively (Figure 8.6B). On the

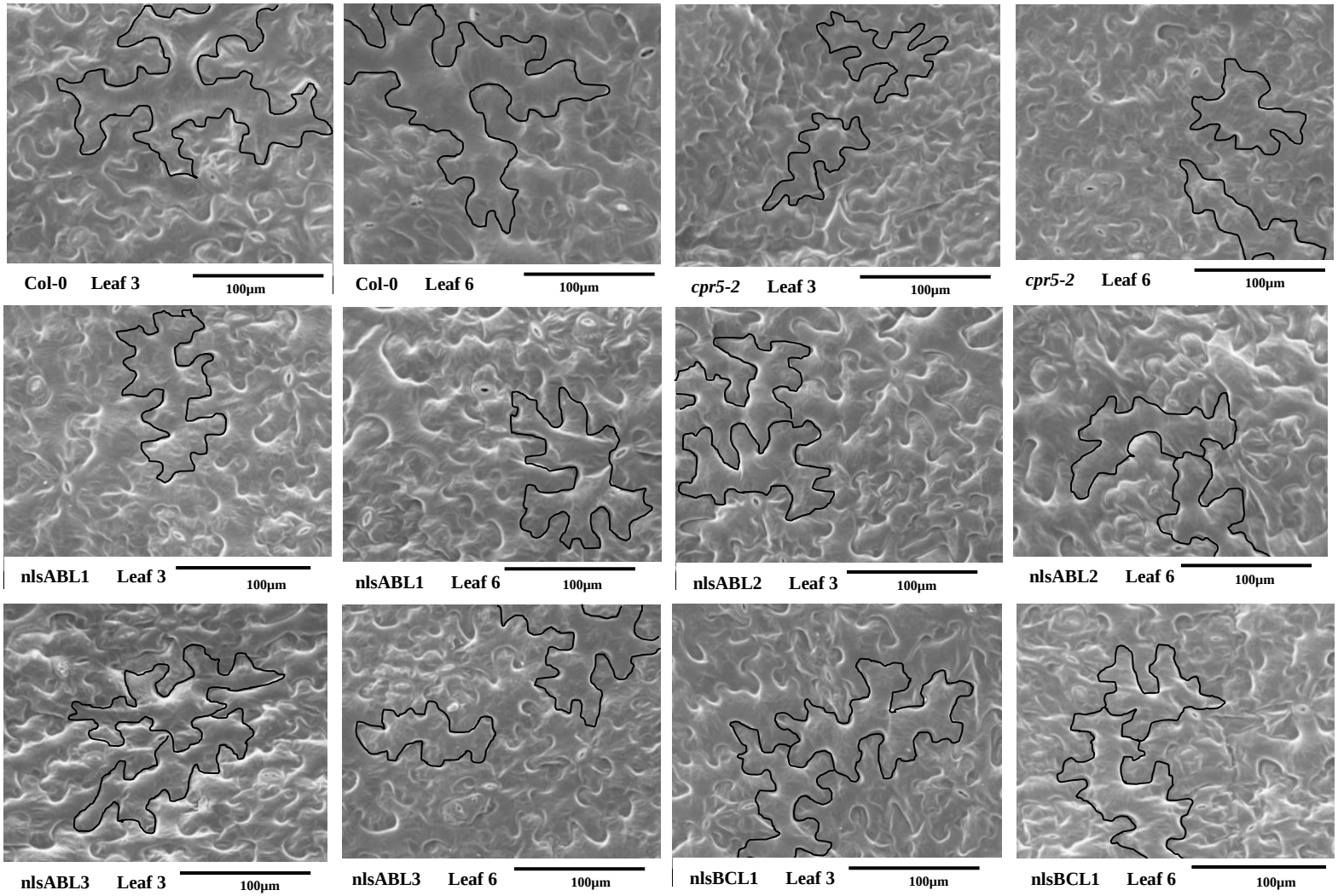
other hand, cell areas from the third and sixth leaves of *cpr5-2* mutant plants were smaller (~1300 and ~2000 µm^2) compared to the wildtype. As shown in Figure 8.6B, the cell areas of the third and sixth leaves of *nlsCPR5* lines are intermediate in size compared to the wildtype and *cpr5-2*. In summary, these results support my hypothesis that *nlsCPR5* cells are smaller than the wildtype and bigger than *cpr5-2*.

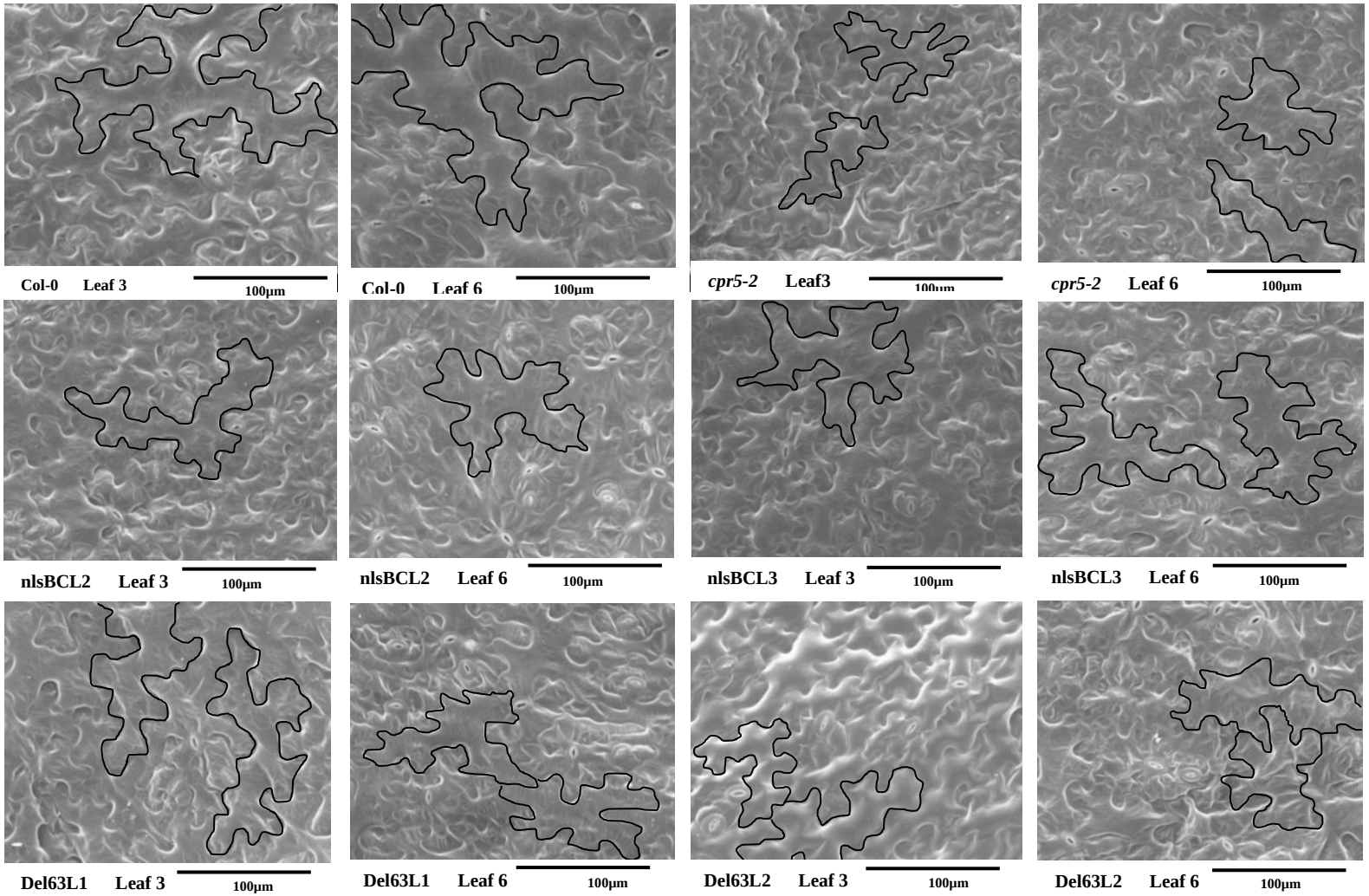
8.2.8 Cell expansion appears to be affected in leaves of *nlsCPR5* transgenic plants

Cell size positively correlates with endoreduplication, since cells expand in size due to increases in genome size through endoreduplication (Brininstool et al., 2008). Endoreduplication is a type of cell cycle in which cells enter into the next round of replication without going through cell division (mitosis), resulting in the doubling of the genome and ultimately cell expansion (Brininstool et al., 2008). Based on leaf and cell sizes, it was hypothesised that *nlsCPR5* transgenic leaves will have reduced ploidy levels. In order to investigate the ploidy levels, nuclei were extracted from leaves, stained with propidium iodide (PI), and kinematic analyses were carried out using a flow cytometry as described in materials and methods (Section 2.16).

As shown in Figure 8.7, five nuclei peaks corresponding to 2C, 4C, 8C, 16C and 32C ploidy levels, can be seen in wildtype leaves in contrast to four nuclei peaks observed in leaves from *cpr5-2* plants. The most striking difference between the wildtype and *cpr5-2* nuclei populations can be explained on the basis of 16C and 32C nuclei populations. For example, none of the *nlsCPR5* lines and *cpr5-2* mutant leaves depicted a 32C nuclei peak (Figure 8.7). Similarly, there are significant reductions in 16C nuclei populations of *nlsCPR5* and *cpr5-2* lines. Overall, the majority of *nlsCPR5* lines showed intermediate populations of 16C nuclei, whereas, no 32C nuclei were detected. In conclusion, these results support my hypothesis that cells from *nlsCPR5* plants possess reduced ploidy levels, which correlates with their cell sizes and leaf area reductions.

A.


Col-0 Leaf 3
100µm
Col-0 Leaf 6
100µm
cpr5-2 Leaf 3
100µm
cpr5-2 Leaf 6
100µm
nlsABL1 Leaf 3
100µm
nlsABL1 Leaf 6
100µm
nlsABL2 Leaf 3
100µm
nlsABL2 Leaf 6
100µm
nlsABL3 Leaf 3
100µm
nlsABL3 Leaf 6
100µm
nlsBCL1 Leaf 3
100µm
nlsBCL1 Leaf 6
100µm


Col-0 Leaf 3
100µm
Col-0 Leaf 6
100µm
cpr5-2 Leaf3
100µm
cpr5-2 Leaf 6
100µm
nlsBCL2 Leaf 3
100µm
nlsBCL2 Leaf 6
100µm
nlsBCL3 Leaf 3
100µm
nlsBCL3 Leaf 6
100µm
Del63L1 Leaf 3
100µm
Del63L1 Leaf 6
100µm
Del63L2 Leaf 3
100µm
Del63L2 Leaf 6
100µm

B.

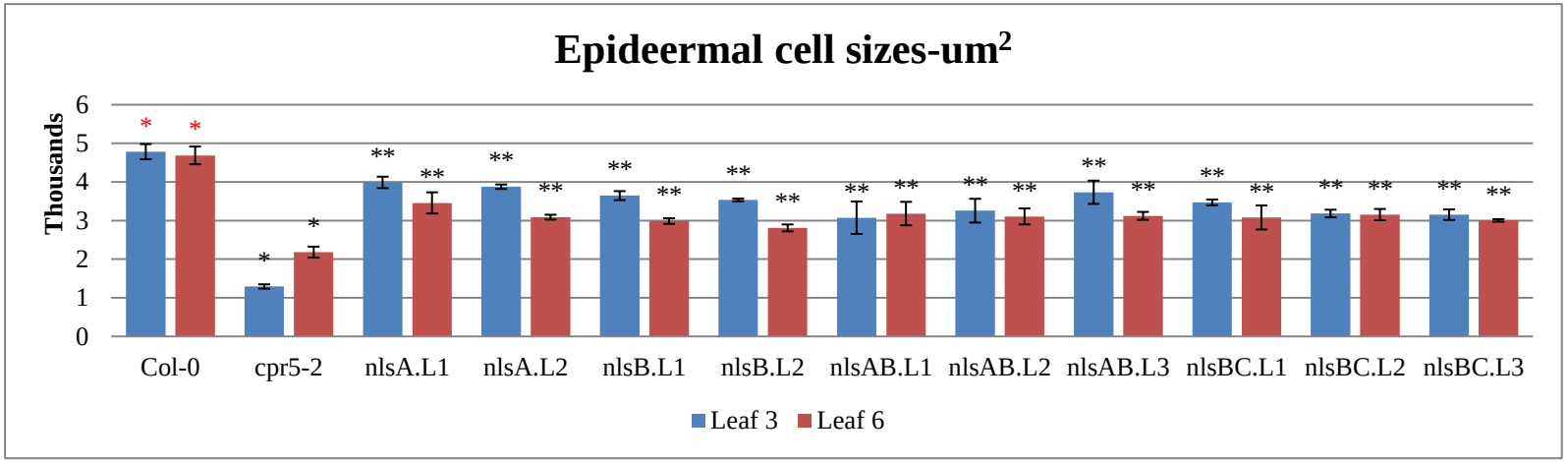


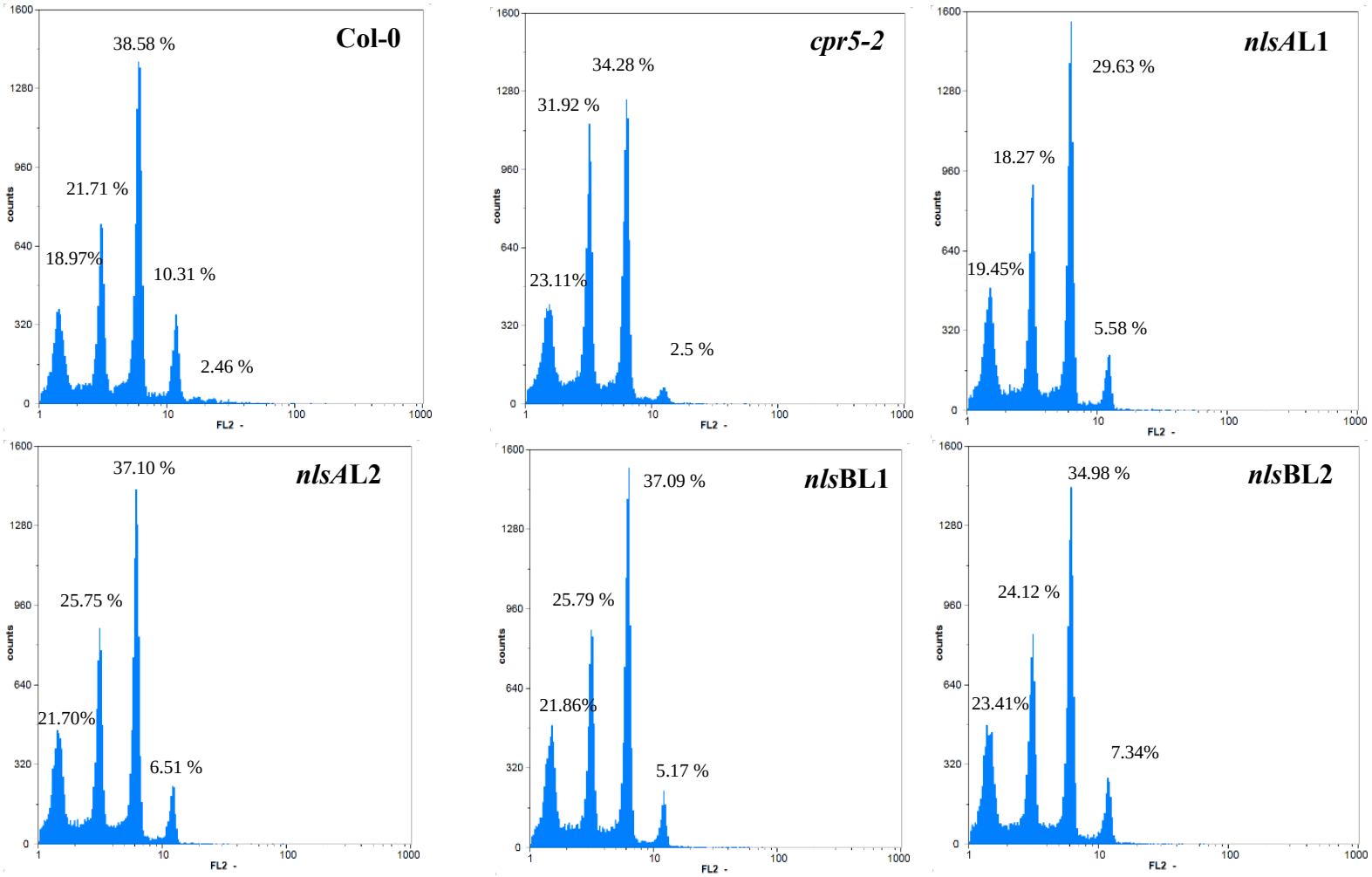
Figure 8.6 Epidermal pavement cell size

Figure 8.6A shows the images taken by the Scanning Electron Microscope (SEM). Figure 8.6B displays average cell areas measured by ImageJ as described in materials and methods (Section 2.14). The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant compared to wildtype, * = significant compared to *cpr5-2*; ** = significant compared to wildtype Col-0 and *cpr5-2*. nlsA = *nlsCPR5A*, nlsB = *nlsCPR5B*, nlsAB = *nlsCPR5AB*, nlsBC = *nlsCPR5BC*, and nlsABC = *nlsCPR5ABC*.

8.2.9 Degree of trichome aberration varies among *nlsCPR5* lines

The development of leaf trichomes and their appendages have a positive correlation with the ploidy levels of cells (Brininstool et al., 2008). It has been established that cells with a reduced level of endoreplication produce fewer trichomes and with reduced number of appendages (Brininstool et al., 2008). *Arabidopsis* trichomes have a specific branching pattern i.e. 4-5% with 1-branch, 75-80% with 2 branches and 19-20% with three appendages (Brininstool et al., 2008). Based on their ploidy levels, leaves on the *nlsCPR5* transgenic lines will have aberrant trichomes. In order to count the trichomes and their appendages, the first four rosette leaves of each independent line were analysed under a diseccting microsope at 16 DAS.

As shown in Figures 8.8A and 8.8B, wildtype-like trichomes were observed on the majority of *nlsCPR5A* and *nlsCPR5B* leaves. Like the wildtype, trichomes with three appendages were the largest population on *nlsCPR5A* and *nlsCPR5B*, in addition to the presence of trichomes with 2 and 4 appendages (Figure 8.8A and 8.8B). Contrary to *nlsCPR5A* and *nlsCPR5B*, the typcial pattern of trichomes and their density was found to be abolished on the leaves of *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* transgenic plants. The trichomes on *nlsCPR5AB, nlsCPR5BC, nlsCPR5ABC* were drastically reduced in number as well as in appendages, and displayed *cpr5-2* like characteristics (Figure 8.8C, 8.8D and 8.8E). Bifurcated trichomes were shown to be the predominent population of trichomes on the leaves of *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* lines. The trichomes on *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* lines were observed to be shorter, thinner, and present mostly on the leaf edges (Data not shown). In conclusion, these results partly support my hypothesis, as both *nlsCPR5A* and *nlsCPR5B* lines show wildtype-like trends of 3-appendged trichomes. The pattern of trichomes and their appendages on *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* leaves is also consitent with my hypothesis.


Col-0
38.58 %
21.71 %
18.97%
10.31 %
2.46 %
cpr5-2
34.28 %
31.92 %
23.11%
2.5 %
nlsAL1
29.63 %
18.27 %
19.45%
5.58 %
nlsAL2
37.10 %
25.75 %
21.70%
6.51 %
nlsBL1
37.09 %
25.79 %
21.86%
5.17 %
nlsBL2
34.98 %
24.12 %
23.41%
7.34%
counts
FL2 -

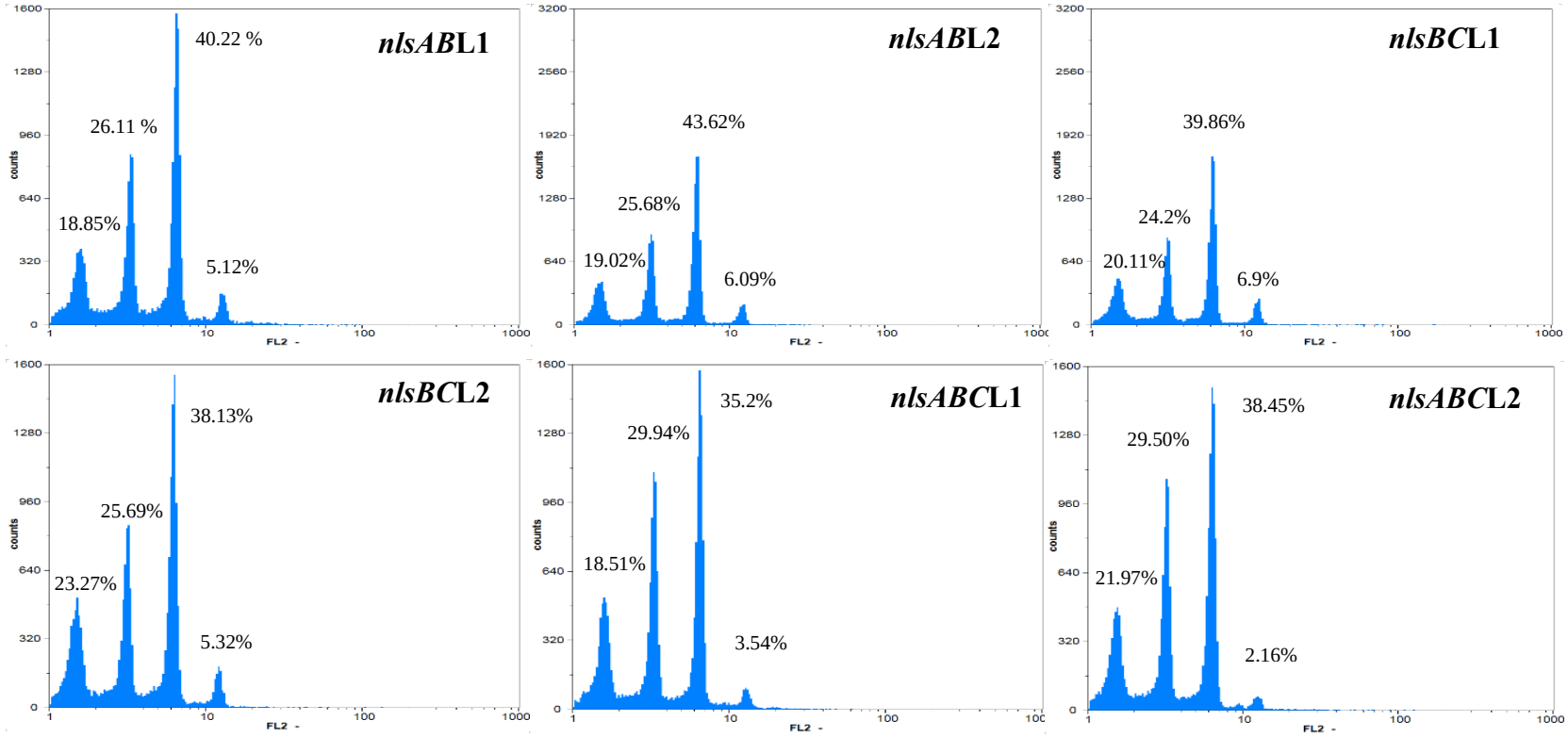
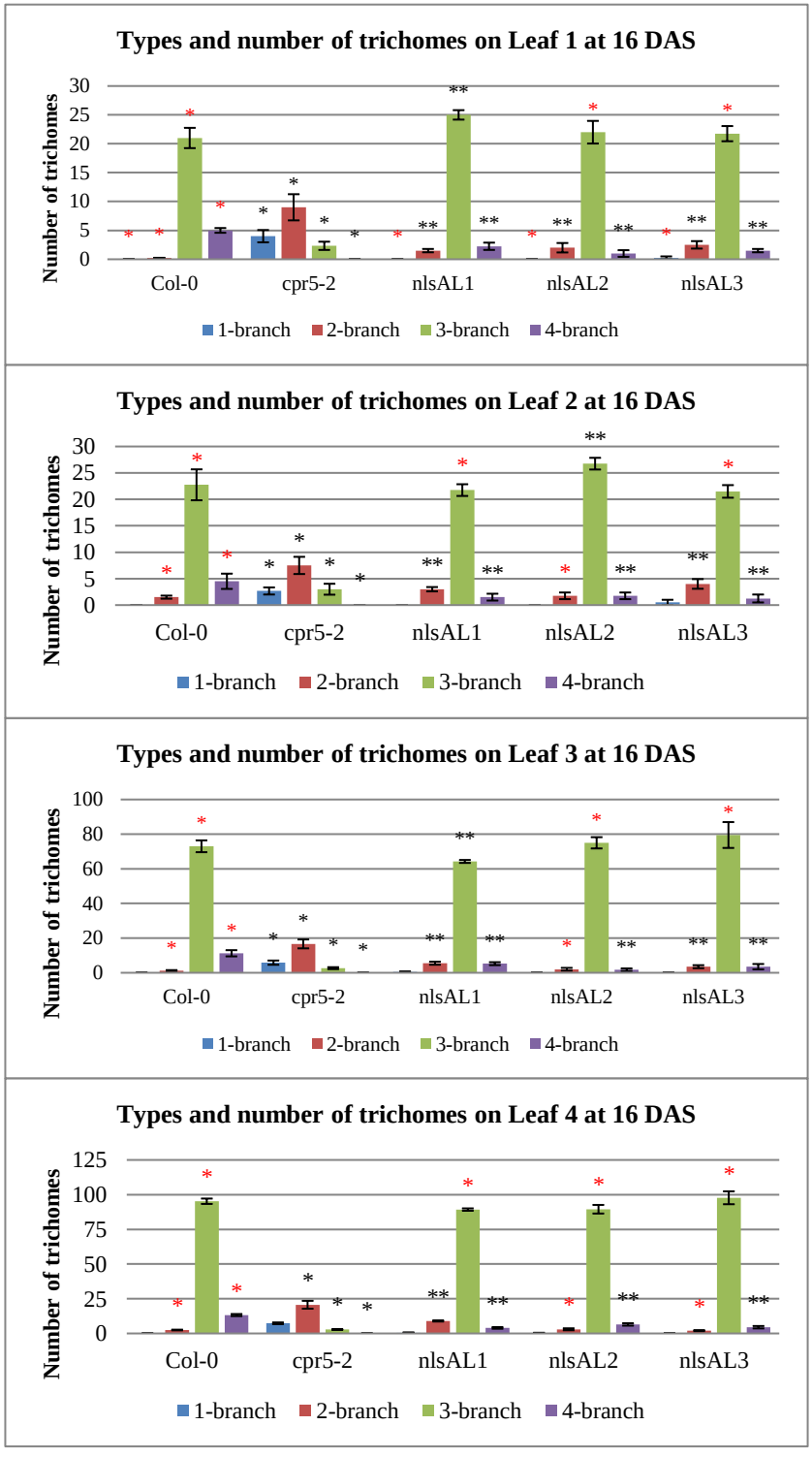


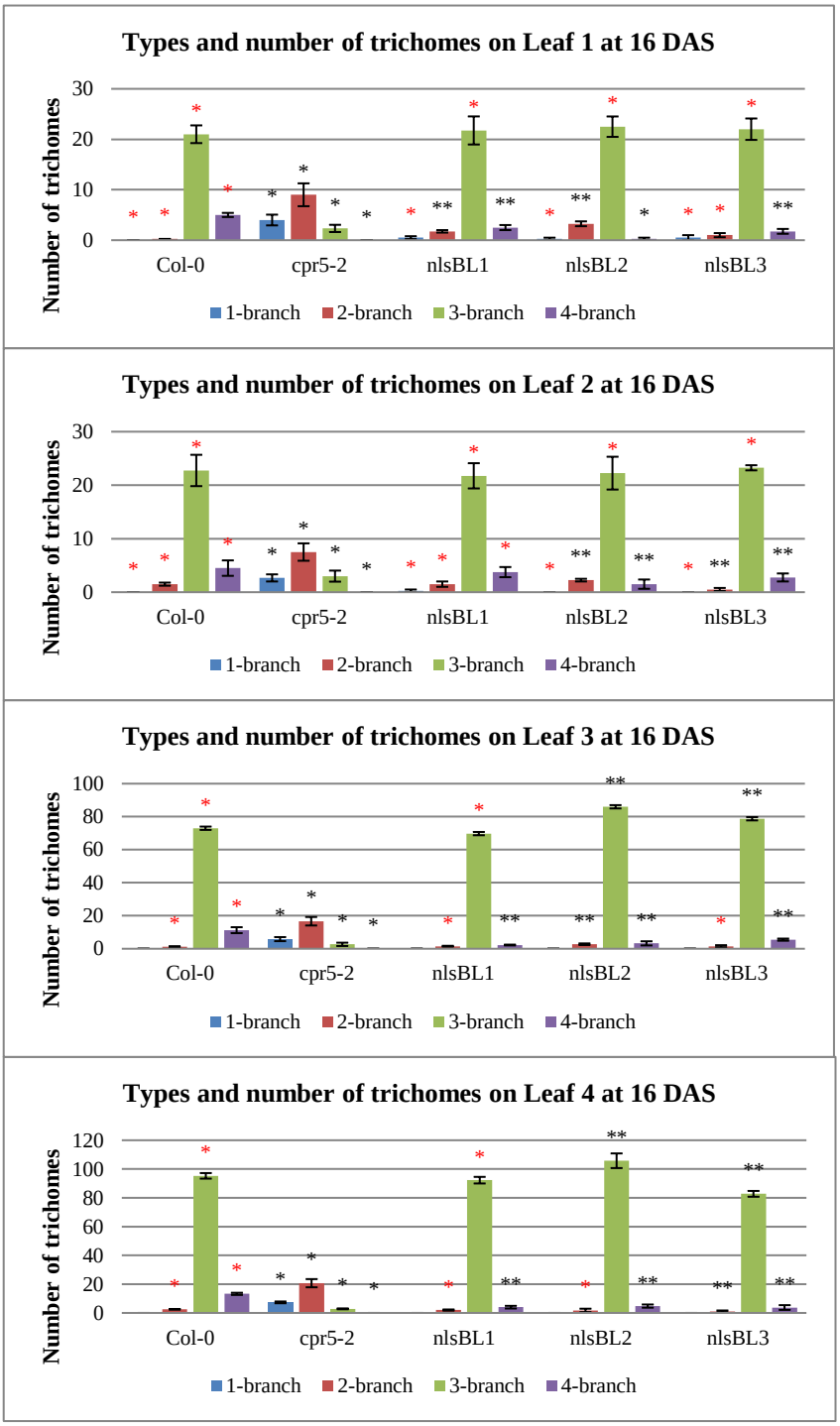
Figure 8.7 Measurements of ploidy levels in leaves of *nlsCPR5* mutants at 24 DAS

Nuclei from leaf 3 and 4 were extracted from 24 day old plants, stained in propidium iodide and analysed as described in materials and methods. The percentage of cells present in each peak was analysed and measured using FloMax®, according to the manufacturer;s instruction. Three biological replicates were analysed for each line, however, images are only shown for two replicates. nlsA = *nlsCPR5A*, nlsB = *nlsCPR5B*, nlsAB = *nlsCPR5AB*, nlsBC = *nlsCPR5BC*, and nlsABC = *nlsCPR5ABC*.

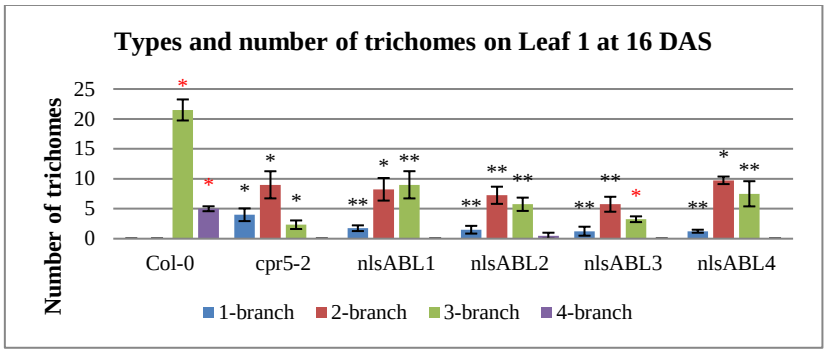
A.

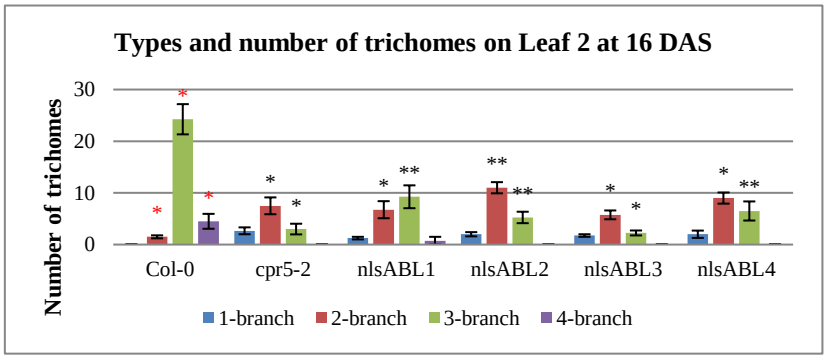

Types and number of trichomes on Leaf 1 at 16 DAS
Number of trichomes
30
25
20
15
10
5
0
Col-0
cpr5-2
nlsAL1
nlsAL2
nlsAL3
1-branch
2-branch
3-branch
4-branch
Types and number of trichomes on Leaf 2 at 16 DAS
Number of trichomes
30
25
20
15
10
5
0
Col-0
cpr5-2
nlsAL1
nlsAL2
nlsAL3
1-branch
2-branch
3-branch
4-branch
Types and number of trichomes on Leaf 3 at 16 DAS
Number of trichomes
100
80
60
40
20
0
Col-0
cpr5-2
nlsAL1
nlsAL2
nlsAL3
1-branch
2-branch
3-branch
4-branch
Types and number of trichomes on Leaf 4 at 16 DAS
Number of trichomes
125
100
75
50
25
0
Col-0
cpr5-2
nlsAL1
nlsAL2
nlsAL3
1-branch
2-branch
3-branch
4-branch

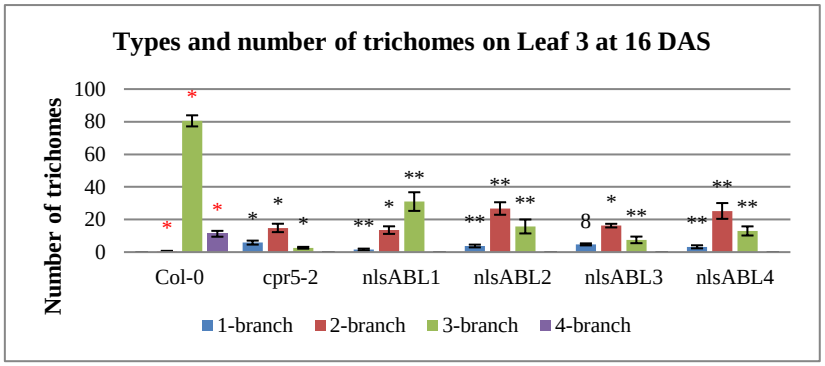
B.

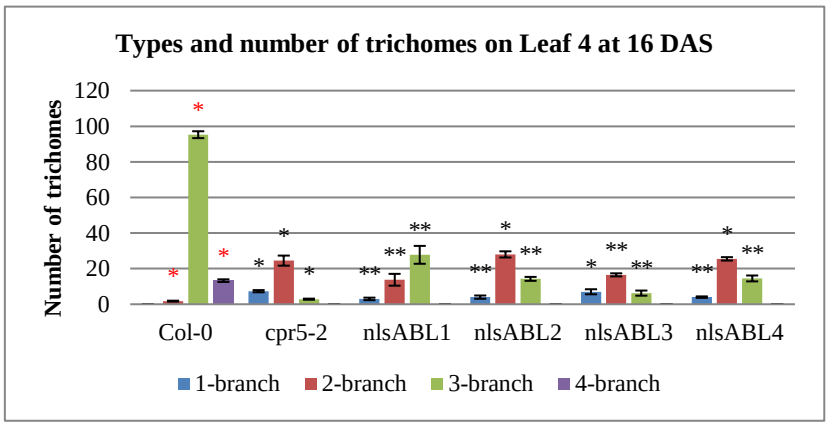

Types and number of trichomes on Leaf 1 at 16 DAS
Number of trichomes
30
20
10
0
Col-0
cpr5-2
nlsBL1
nlsBL2
nlsBL3
1-branch
2-branch
3-branch
4-branch
Types and number of trichomes on Leaf 2 at 16 DAS
Number of trichomes
30
20
10
0
Col-0
cpr5-2
nlsBL1
nlsBL2
nlsBL3
1-branch
2-branch
3-branch
4-branch
Types and number of trichomes on Leaf 3 at 16 DAS
Number of trichomes
100
80
60
40
20
0
Col-0
cpr5-2
nlsBL1
nlsBL2
nlsBL3
1-branch
2-branch
3-branch
4-branch
Types and number of trichomes on Leaf 4 at 16 DAS
Number of trichomes
120
100
80
60
40
20
0
Col-0
cpr5-2
nlsBL1
nlsBL2
nlsBL3
1-branch
2-branch
3-branch
4-branch

C.

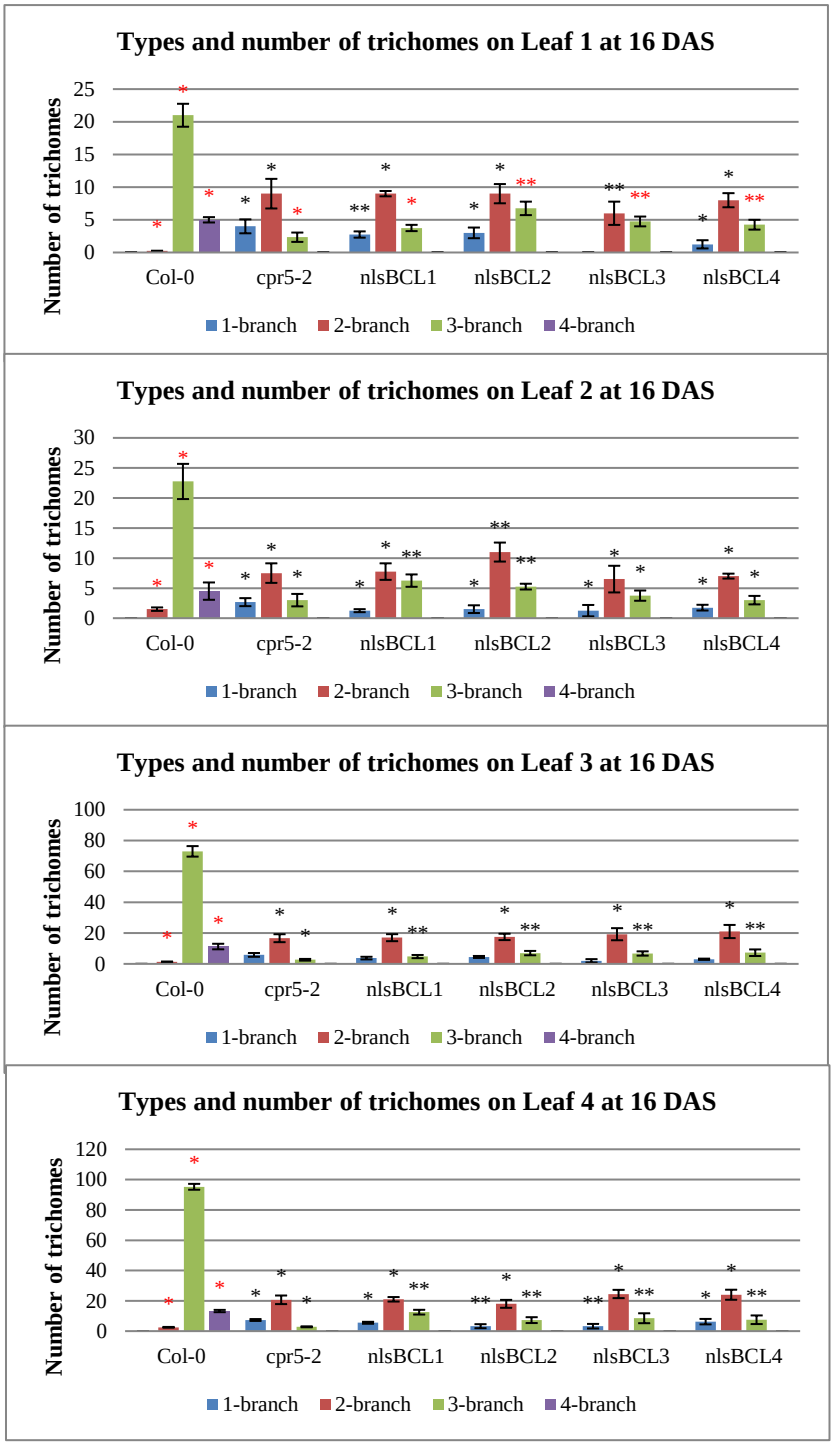

Types and number of trichomes on Leaf 1 at 16 DAS
Number of trichomes
25
20
15
10
5
0
Col-0
cpr5-2
nlsABL1
nlsABL2
nlsABL3
nlsABL4
1-branch
2-branch
3-branch
4-branch


Types and number of trichomes on Leaf 2 at 16 DAS
Number of trichomes
30
20
10
0
Col-0
cpr5-2
nlsABL1
nlsABL2
nlsABL3
nlsABL4
1-branch
2-branch
3-branch
4-branch


Types and number of trichomes on Leaf 3 at 16 DAS
Number of trichomes
100
80
60
40
20
0
Col-0
cpr5-2
nlsABL1
nlsABL2
nlsABL3
nlsABL4
1-branch
2-branch
3-branch
4-branch


Types and number of trichomes on Leaf 4 at 16 DAS
Number of trichomes
120
100
80
60
40
20
0
Col-0
cpr5-2
nlsABL1
nlsABL2
nlsABL3
nlsABL4
1-branch
2-branch
3-branch
4-branch

D.


Types and number of trichomes on Leaf 1 at 16 DAS
Number of trichomes
25
20
15
10
5
0
Col-0
cpr5-2
nlsBCL1
nlsBCL2
nlsBCL3
nlsBCL4
1-branch
2-branch
3-branch
4-branch
Types and number of trichomes on Leaf 2 at 16 DAS
Number of trichomes
30
25
20
15
10
5
0
Col-0
cpr5-2
nlsBCL1
nlsBCL2
nlsBCL3
nlsBCL4
1-branch
2-branch
3-branch
4-branch
Types and number of trichomes on Leaf 3 at 16 DAS
Number of trichomes
100
80
60
40
20
0
Col-0
cpr5-2
nlsBCL1
nlsBCL2
nlsBCL3
nlsBCL4
1-branch
2-branch
3-branch
4-branch
Types and number of trichomes on Leaf 4 at 16 DAS
Number of trichomes
120
100
80
60
40
20
0
Col-0
cpr5-2
nlsBCL1
nlsBCL2
nlsBCL3
nlsBCL4
1-branch
2-branch
3-branch
4-branch

E.

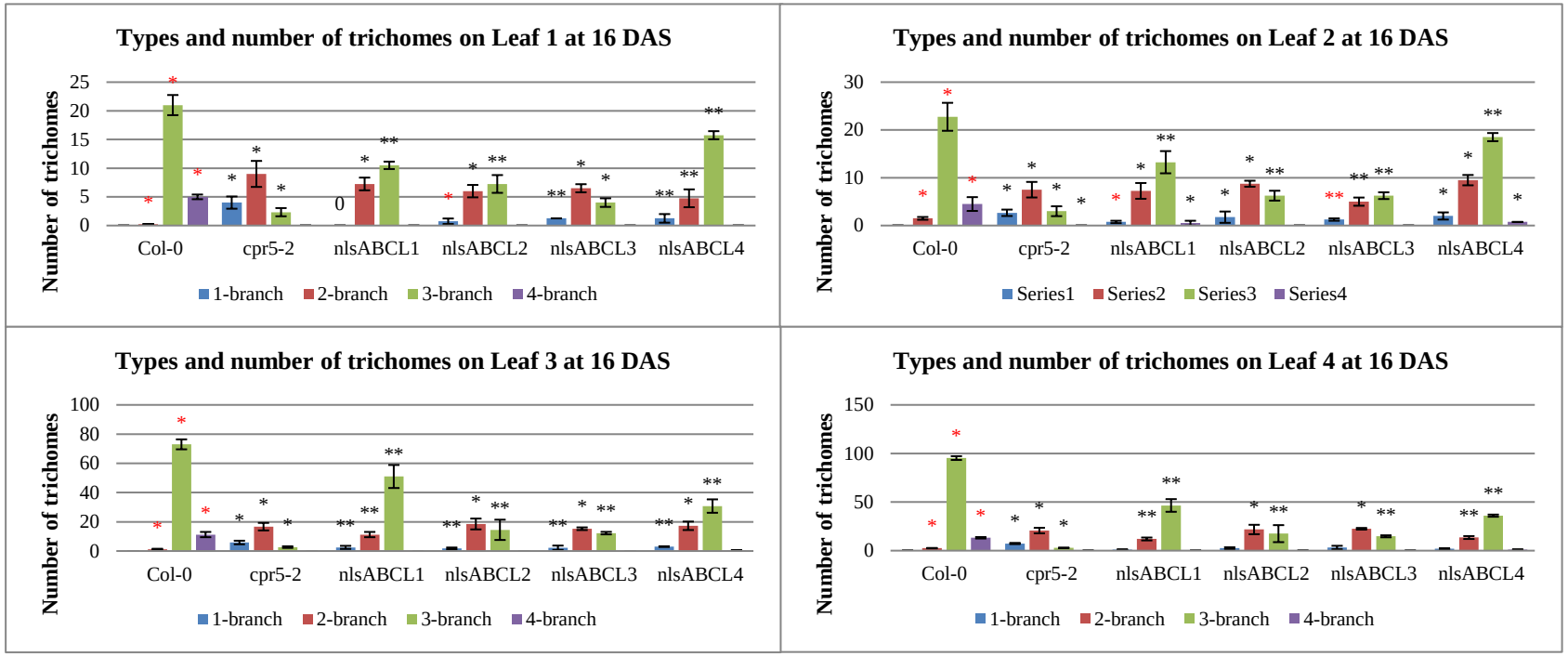


Figure 8.8 Type and number of trichome appendages at 16 DAS

These figures show the type and number of appendages on the 1st, 2nd, 3rd and 4th rosette leaves at 16 DAS. The leaves were detached and the trichomes were counted from a disc of leaf (0.67 cm^2) under a dissecting microscope, and the average number of trichomes carrying a certain number of appendages was plotted. * = significant compared to wildtype, * = significant compared to *cpr5-2*; ** = significant compared to wildtype Col-0 and *cpr5-2*. nlsA = *nlsCPR5A*, nlsB = *nlsCPR5B*, nlsAB = *nlsCPR5AB*, nlsBC = *nlsCPR5BC*, and nlsABC = *nlsCPR5ABC*.

8.2.10 *nlsCPR5* transgenic plants confer resistance to *Pseudomonas syringae*

Pseudomonas syringae pv *DC3000* (*PstDC3000*) is a bacterial pathogen that infects tomato leaves (Clarke et al., 2000). Similar to tomato, *Arabidopsis thaliana* ecotype Columbia (Col-0) was also shown to be susceptible to this pathogen, whereas, *cpr5-2* mutant plants conferred resistance against this pathogen (Clarke et al., 2000). Being partially complemented, the *nlsCPR5* plants were expected to be resistant to the bacterial pathogen. As demonstrated in Figure 8.9, all of the *nlsCPR5* lines including *nlsCPR5A* and *nlsCPR5B*, which were found to have normal trichomes and no HR-lesions, conferred resistance to *PstDC3000* as compared to the wildtype (Figure 8.9). In conclusion, these results are in line with my hypothesis that *nlsCPR5* plants confer resistance due to partial complementation.

8.2.11 *nlsCPR5* mutants show elevated levels for *PR1* and *PR5* genes

PATHOGENESIS-RELATED-GENE1 and *5* (*PR1* and *PR5*) are antimicrobial genes that have been shown to be upregulated when a pathogen attacks a plant (Bowling et al., 1997; Clarke et al., 2001). Plants producing higher levels of Reactive Oxygen Species (ROS) or under oxidative stress conditions were also shown to have elevated levels of defence-related genes (Clarke et al., 2001). It has been established that *cpr5-2* plants show elevated levels of *PR1* and *PR5* (Bowling et al., 1998; Clarke et al., 2000). Based on complementation and resistance levels, it was assumed that *nlsCPR5* plants have upregulated levels of *PR1* and *PR5* genes. Relative transcript abundances of *PR1* and *PR5* were quantified from cDNAs prepared from *nlsCPR5* transgenic plants. As expected, *PR1* and *PR5* were found to be higher in *nlsCPR5* lines (Figure 8.10). Additionally, *PR1* and *PR5* levels vary among *nlsCPR5* transgenic lines. For instance, *nlsCPR5AB* and *nlsCPR5BC* show higher levels of *PR1* and *PR5* than *nlsCPR5A* and *nlsCPR5B*. In conclusion, *PR1* and *PR5* expression is induced in *nlsCPR5* transgenic plants.

8.2.12 *nlsCPR5* transgenic plants display downregulation of *PDF1.2*

PDF1.2, a *PLANT DEFENSIN* gene, is shown to be upregulated in response to wounding or insects (Clarke et al., 2001). In contrast to already published reports (Bowling et al., 1997; Clarke et al., 2001), the *PDF1.2* gene was shown to be downregulated in *cpr5-2* as well as in

SynCPR5 and *metCPR5* (Chapter 7). Thus, it was hypothesized that *nlsCPR5* transgenic plants will have downregulated levels of *PDF1.2*. Contrary to northern blotting carried out by Clarke et al., 2001, the relative transcript abundance of *PDF1.2* was measured using real-time qRT-PCR. *nlsCPR5* show *cpr5-2*-like or lower than *cpr5-2* levels of *PDF1.2* expression, compared to the wildtype (Figure 8.10). Amongst *nlsCPR5* mutants, *nlsCPR5AB* and *nlsCPR5BC* both showed *PDF1.2* transcripts significantly lower than the wildtype but these levels were insignificantly different from the levels of *PDF1.2* expression in *cpr5-2*. Strikingly, transcript abundance of *PDF1.2* from *nlsCPR5AB* and *nlsCPR5BC* were significantly higher compared to *nlsCPR5A* and *nlsCPR5B* levels. In summary, these results are consistent with the *PDF1.2* results shown previously in Chapter 7, where *PDF1.2* levels in *SynCPR5* and *metCPR5* were also shown to be downregulated as compared to the wildtype.

8.2.13 *nlsCPR5* mutant plants show reduced plant height

In additon to various phenotypes associated with a mutation(s) in At*CPR5*, shorter stature was one of the many unique features of *cpr5-2* plants when compared to the wild-type (Bowling et al., 1997; Jing et al., 2002). In addtion to intermediate leaf and cell sizes, *nlsCPR5* transgenic plants were hypothesised to have intermediate plant heights. In order to record plant height differences, the height of naturally scenesced *nlsCPR* transgenic plants was measured. The average height was estimated from 10 plants of the same independent lines. As shown, the *nlsCPR5A* and *nlsCPR5B* mutants were shown to be less affected by these mutations (Figure 8.11) and yielded plants with an average height of 38.5 cm, compared to the wildtype, which had an average height of 42.0 cm. Additionally, *nlsCPR5AB* and *nlsCPR5BC* also displayed further reduction in plant height, with an average plant height of 35.5 cm when compared to *nlsCPR5A* and *nlsCPR5B* plants (Figure 8.11). However, the *nlsCPR5ABC* lines exhibited an average height of 38.0 cm (Figure 8.11). To conclude, these results support our hypothesis that plant stature is also affected in *nlsCPR5* transgenic plants.

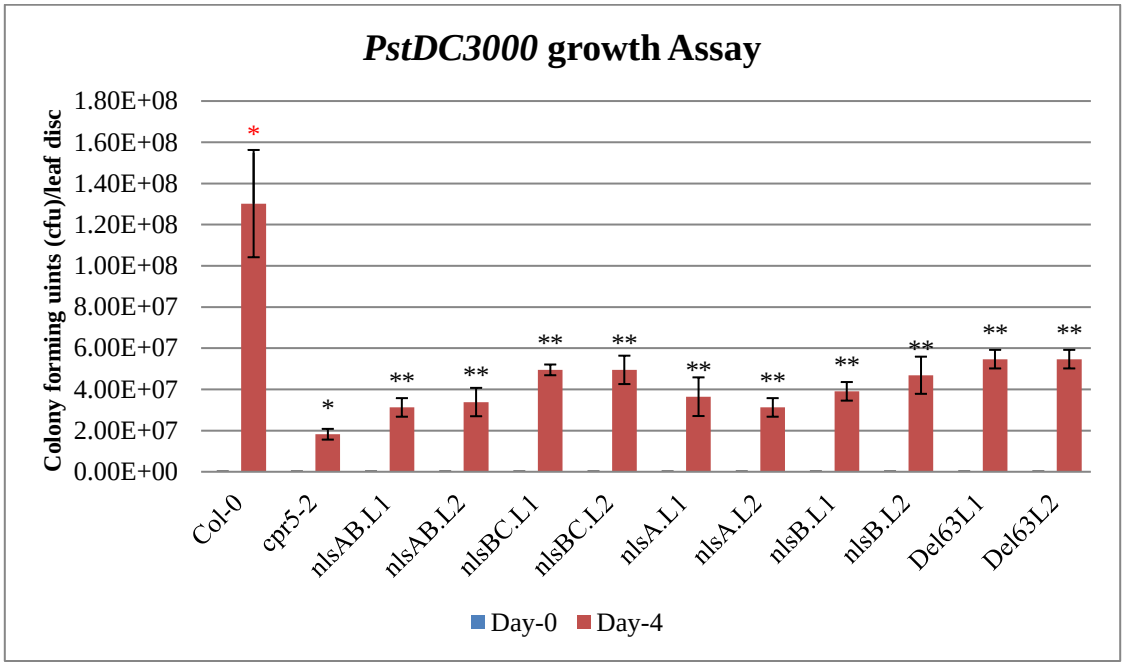


Figure 8.9 *Pseudomonas syringae* pv *DC3000* growth assay

This figure shows colony forming units (cfu) formed by bacteria extracted from leaf discs of each selected line as described in materials and methods. Leaves of 6 week old plants were infiltrated with ~10^5cells of *Pseudomonas syringae* pv *Dc3000* (*PstDc3000*). The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant compared to wildtype, * = significant compared to *cpr5-2*; ** = significant compared to wildtype Col-0 and *cpr5-2*. nlsA = *nlsCPR5A*, nlsB = *nlsCPR5B*, nlsAB = *nlsCPR5AB*, nlsBC = *nlsCPR5BC*, and *nlsABC* = *nlsCPR5ABC*, Del63= *Del63CPR5*.

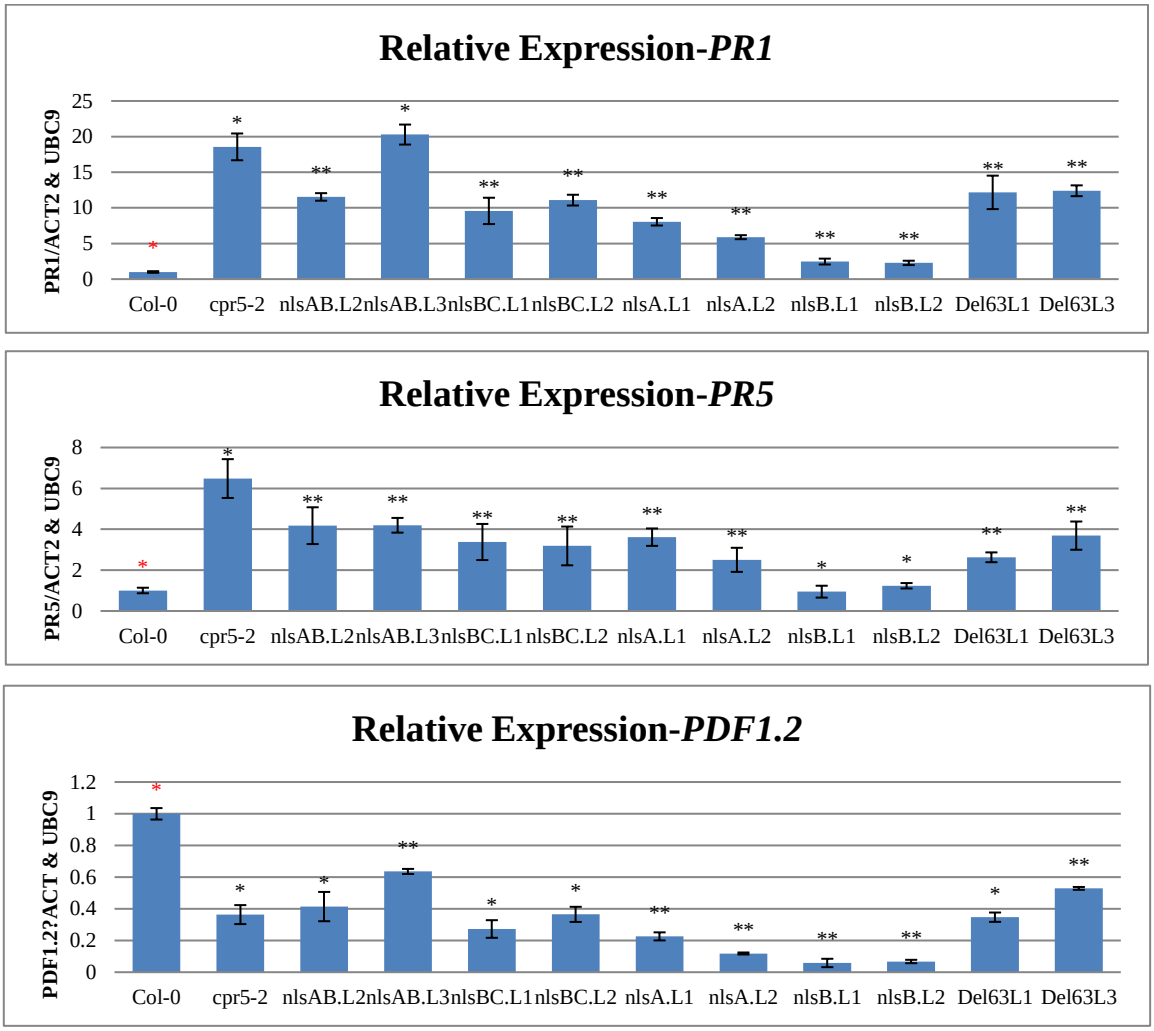


Figure 8.10 Quantification of transcript abundance of *PR1*, *PR5*, and *PDF1.2*

This figure shows relative levels of transcripts of defence-related genes, such as, *PR1*, *PR5*, and *PDF1.2* using real-time quantitative reverse transcriptase PCR. Total RNA was isolated from 31 day old plants, using the Plant miniprep kit from ZymoResearch. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant compared to wildtype, * = significant compared to *cpr5-2*; ** = significant compared to wildtype Col-0 and *cpr5-2*. nlsA = *nlsCPR5A*, nlsB = *nlsCPR5B*, nlsAB = *nlsCPR5AB*, nlsBC = *nlsCPR5BC*, and nlsABC = *nlsCPR5ABC*, Del63= *Del63CPR5*.

8.2.14 Collective mutations in NLS-encoding clusters results in early bolting

Bolting results from the transition of vegetative growth towards reproductive development that starts with the elongation of the first internode and initiation of the main stem or bolt (Pouteau and Albertini, 2009, Rosas et al., 2014). While charactising *nlsCPR5* plants, it was noticed that *nlsCPR5* plants bolt earlier than the wildtype and *cpr5* plants. Thus it was hypothesised that mutations in the NLS-encoding clusters have an impact on bolting in *nlsCPR5* transgenic plants. *nlsCPR5* transgenic plants were grown under short day conditions (12 hours of light and 12 hours of dark). Plants were observerd on a daily basis in order to count rossette leaves, and the number of days taken to bolt were recorded.

As shown in Figure 8.12, *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* plants display earleir bolting than the wildype, *cpr5-2*, *nlsCPR5A* and *nlsCPR5B*. Bolting was initiatied in *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* plants when they had 5-7 rosette leaves compared to wildtype plants which bolted when they had 8-10 rosette leaves. Additionally, bolting was initiated 4-5 days earlier in *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* compared to Col-0 and *cpr5-2* plants. *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* plants were observed to already have 5-6 flowers opened by the time bolting was initiated in the wildtype. The *nlsCPR5A* and *nlsCPR5B* transgenic plants bolted on the same day or slightly (~1-2 days) earlier than the wild-type and *cpr5-2* plants. In conclusion, these results indicate that collective mutations in the putative NLS-encoding clusters may have a role in the regulation of bolting in *Arabidopsis*. However, further investigation is needed.

8.2.15 Putative CPR5 NLS clusters and an intrinsically disordered region belong to the same region of CPR5 protein

As shown in Figure 8.1, the putative NLS clusters are shown to be predicted from 41-61 of the amino acids of CPR5. Concomitantly, the N-terminus (first 98 amino acids) of the CPR5 protein is also annotated to be an intrinsically disordered region (IDR; Chapter 4). Thus, it was hypothesised that NLS-encoding clusters could be part of the CPR5 IDR and *vice versa*. To test the relationship between the NLS and IDRs, the N-terminus of CPR5 was deleted in parts. Since the first NLS cluster is predicted at position 41-44 and there are three methionine residues

(38-40 aa) present immediately before the first NLS cluster, the N-terminal region comprising of the first 37 amino acids was deleted, and the construct was termed *Del37CPR5*. As mentioned above and shown in (Figure 8.1), annotated NLS clusters are present within the 41-61 amino acids of CPR5. Additionally, there is a methionine residue at position 64 of CPR5, thus the first 63 amino acids of CPR5 were deleted in *Del63CPR5* in order to exclude NLS-encoding clusters. The first 98 amino acids of CPR5 were deleted in *Del98CPR5* in order to see the effect of deletion of the whole putative IDR. To conclude, *Del37CPR5*, *Del63CPR5* and *Del98CPR5* would help to see the effect of deletion of IDRs on CPR5 functioning, and characterise the role and dependence of IDRs and the NLS on each other. Possibly, the deletion of the N-terminal region of CPR5 is expected to uncouple the role of each region, if any.

8.2.16 *Del37CPR5* and *Del63CPR5* complement HR-like lesions and early leaf yellowing

Since *DelCPR5* transgenes include deletions, it was hypothesised that *DelCPR5* transgenic plants will show aberrant phenotypes in a *cpr5-2* manner. In order to study IDRs and their role, *cpr5-2* plants were complemented with *Del37CPR5*, *Del63CPR5*, and *Del98CPR5* transgenes, and the homozygous plant lines of each transgene were grown in soil and investigated for the presence of HR-like lesions and early yellowing on their leaves.

As shown in Figure 8.3, HR-like lesions and yellowing on the cotyledons and first rosette leaves can be observed in *cpr5-2* plants as compared to the wildtype (Figure 8.3). However, *DelCPR5* (*Del37CPR5* and *Del63CPR5*) lines produced healthy leaves without any lesions or early yellowing (Figure 8.3). Only *DelCPR598* plants displayed *cpr5-2* like leaves i.e. leaves had lesions and yellowing (data not shown). Since *DelCPR598* transgenic plants displayed phenotypes insignificantly different from *cpr5-2* plants, *DelCPR598* transgenic lines were considered as non-complemented lines and were not carried over for further physiological characterisations. From these results, it can be concluded that the execution of HR-like lesions and early leaf yellowing is independent of the first 63 amino acids of the AtCPR5 protein, as HR-like lesions are complemented in both *Del37CPR5* and *Del63CPR5* plants, which lack the first 37 and 63 amino acids, respectively.

8.2.17 *Del63CPR5* transgenic lines display reduced transcript abundance

As shown in Figure 8.4, the majority of *nlsCPR5* lines have displayed reduced expression levels of their transgenes. Therefore, it was hypothesised that the *DelCPR5* transgene (*Del63CPR5*) will also have reduced expression levels of its transcript. To test this, the relative expression levels of the *Del63CPR5* transgene were measured in order to see the difference between the relative expression levels of *cpr5-2*, *CPR5* and *Del63CPR5*. As shown in Figure 8.4, *Del63CPR5* transgenic lines displayed expression levels in a *cpr5-2* manner. In conclusion, *Del63CPR5* reduced transcript abundance is consistent with my hypothesis.

8.2.18 Typical pattern of trichomes is restored by *Del37CPR5* but not by *Del63CPR5* and *Del98CPR5*

All the *cpr5* alleles produce aberrant trichomes on their leaves (Bowling et al., 1997, Boch et al., 1998, Kirik et al., 2001, Jing et al., 2002, Jing et al., 2007). Thus, it was hypothesised that the deletion(s) of putative intrinsically disordered region(s) will result in the accomplishment of abnormal trichomes on leaves from *DelCPR5* plants. To test this hypothesis, trichomes and their appendages were counted from the third and fourth rosette leaves under a dissecting microscope. As shown in Figure 8.13, despite some variations in their total numbers, wildtype-like trichomes and their appendages are observed on *Del37CPR5* lines. In contrast to *Del37CPR5* lines, the typical pattern of trichomes and their appendages was found to be missing on the same leaves (3^{rd} and 4^{th}) from *Del63CPR5* plants (Figure 8.13). Compared to the trichomes on wildtype leaves, the trichomes on *Del63CPR5* leaves were observed to be less in number (Figure 8.13) and predominantly present on the edges and tips of the leaves (data not shown). The trichomes patterns on *Del63CPR5* leaves were reminiscent of trichomes on *nlsCPR5AB*, *nlsCPR5BC* and *nlsCPR5ABC* (Figure 8.13). The third *IDR* mutant, *Del98CPR5*, displayed *cpr5-2* like trichome patterns (data not shown). In summary, the deletion of 37, 62 and 98 amino acids exerted none, moderate and/or severe effects on trichome density and appendages, respectively.

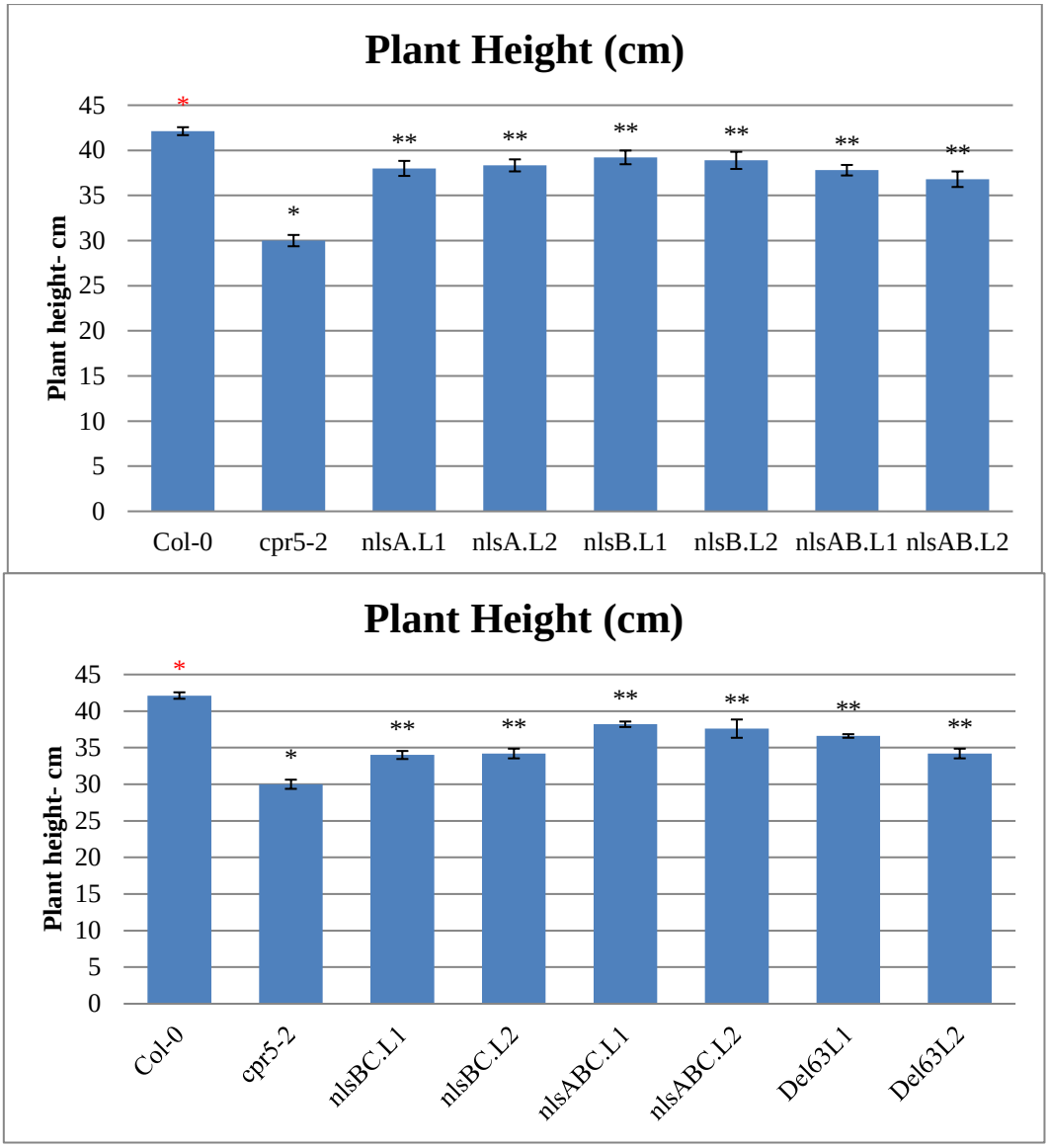
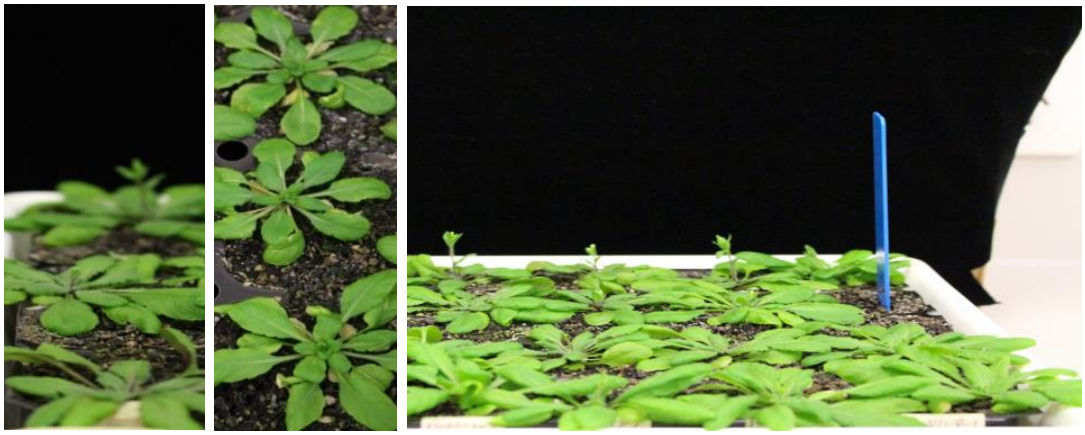


Figure 8.11 Plant height measurement

These figures show the average plant height of *NLS* mutant lines. Upon natural senescence, the plants were harvested and their heights were measured. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant compared to wildtype, * = significant compared to *cpr5-2*; ** = significant compared to wildtype Col-0 and *cpr5-2*. nlsA = *nlsCPR5A*, nlsB = *nlsCPR5B*, nlsAB = nls*CPR5AB*, nlsBC = *nlsCPR5BC*, and nlsABC = *nlsCPR5ABC*.


Col-0
cpr5-2
nlsA.L1
nlsA.L2
nlsA.L3
nlsA.L4


Col-0
cpr5-2
nlsB.L1
nlsB.L2
nlsB.L3
nlsB.L4


Col-0
cpr5-2
nlsAB.L1
nlsAB.L2
nlsAB.L3

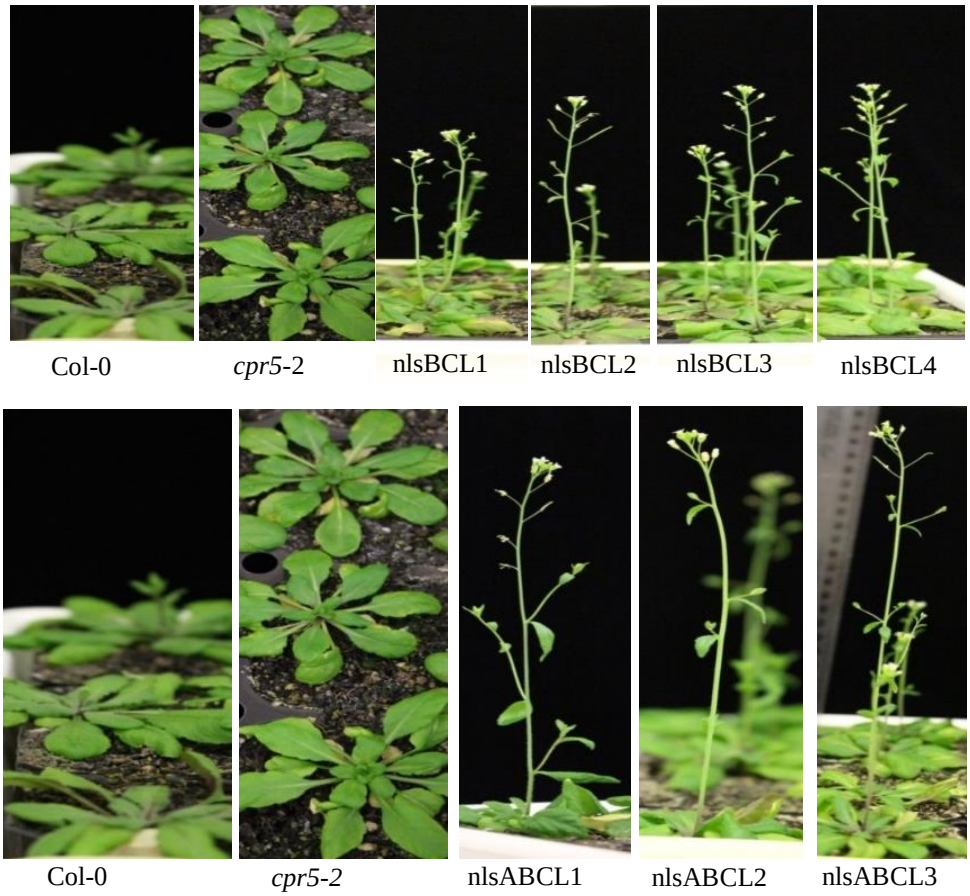


Figure 8.12 Plant growth and bolting of *nlsCPR5* transgenic lines at 27 DAS

These figures show the growth and difference in bolting of *nlsCPR5* transgenic lines compared to *cpr5-2* and the wildtype. Photos were taken at the same position by fixing the camera (Nikon D5300) on a stand. These plants were grown at 20°C under 12 hours light and 12 hours dark conditions. nlsA = *nlsCPR5A*, nlsB = *nlsCPR5B*, nlsAB = nls*CPR5AB*, nlsBC = *nlsCPR5BC*, and nlsABC = *nlsCPR5ABC*

8.2.19 *Del37CPR5* and *Del63CPR5* both produce leaves of intermediate sizes

As shown in Sections 7.2.3 and 8.2.6, the majority of the transgenic lines of *synCPR5G2* and *nlsCPR5* have shown smaller and intermediate-sized leaves than the wildtype, respectively. Thus, it is likely that *DelCPR5* transgenic plants also display leaves of intermediate sizes compared to the wildtype and *cpr5-2*. In order to quantify differences, the areas of the first four rosette leaf pairs of *DelCPR5* transgenic lines were measured. As expected, the leaves on the *Del37CPR5* and *Del63CPR5* plants were significantly smaller compared to the wildtype when measured at 23 DAS (Figure 8.14). In conclusion, the deletions of IDRs in *DelCPR5* transgenic plants somehow result in leaves of intermediate sizes, which is in line with my hypothesis.

8.2.20 Sizes of epidermal cells are consistent with leaf sizes of *DelCPR5* plants

Based on leaf sizes, the cells of *DelCPR5* transgenic plants were hypothesised to be smaller than the wildtype. To test this hypothesis, the size of epidermal pavement cells of *DelCPR5* transgenic lines was measured. As expected, the epidermal pavements cells of leaves 3 and 6 from *Del37CPR5* and *Del63CPR5* transgenic plants were intermediate in size when compared to the wildtype and *cpr5-2* cells. A closer examination of the cell size differences between the leaves of *Del37CPR5* and *Del63CPR5* revealed that epidermal cells from *Del37CPR5* are bigger than *Del63CPR5* cells (Figure 8.15). In conclusion, cell sizes of *DelCPR5* transgenic plants are in line with the hypothesis.

8.2.21 Nuclei from *DelCPR5* leaf cells show reduced ploidy levels

Del37CPR5 and *Del63CPR5* plants have shown smaller leaves and pavement cells for leaves 3 and 6, therefore their cells were hypothesised to have reduced ploidy levels. To test this hypothesis, the ploidy levels of *Del37CPR5* and *Del63CPR5* leaves were measured using flow cytometry, as described in materials and methods (Section 2.16). Compared to the wildtype, none of the *DelCPR5* lines show the presence of 32C as shown by *cpr5-2*. Additionally, there were reductions in 16C nuclei populations (Figure 8.16). To conclude, these results are consistent with the notion that *DelCPR5* transgenic plants have reduced ploidy levels in their leaf cells.

8.2.22 *Del63CPR5* confers resistance but *Del37CPR5* is susceptible to *PstDC3000*

As shown in Section 8.2.10, the majority of *nlsCPR5* transgenic lines have shown resistance to the bacterial pathogen, *PstDC3000*. Thus, it was hypothesised that *DelCPR5* transgenic plants will be resistant to the bacterial pathogen, *P. syringae*. To test this hypothesis, the leaves from 5-6 week old plants of *DelCPR5* (*Del37CPR5* and *Del63CPR5*) were challenged with the *P. syringae* (*PstDC3000*) infection, in order to record their resistance levels to the pathogen, as mentioned in materials and methods (Section 2.14). As shown in Figure 8.17, the *Del37CPR5* lines showed susceptibility to *PstDC3000* bacterial infection. Remarkably, the susceptibility levels of *Del37CPR5* lines were found to be significantly higher than the wildtype (Figure 8.17). In contrast, *Del63CPR5* lines conferred resistance to *Pseudomonas syringae*, *PstDC30000*. In conclusion, these results partly support our hypothesis since *Del37CPR5* were susceptible to *Pseudomonas syringae* pv *PstDC3000* in contrast to *Del63CPR5* plants, which were found to be resistant.

8.2.23 *PR1* levels are not restored in *Del63CPR5* transgenic plants

Since *DelCPR5* transgenes have been unable to complement some of the compromised phenotypes in *cpr5-2*, it was assumed that *DelCPR5* transgenic lines would have upregulated levels of *PATHOGENESIS-RELATED GENE1*, *PR1*. The relative transcript levels of the defence-related gene (*PR1*) were quantified only from *Del63CPR5* transgenic plants, due to the non-availability of cDNA from *Del37CPR5* independent lines. As shown in Figure 8.10, *Del63CPR5* transgenic plants displayed higher expression levels of *PR1*, which were similar to the *PR1* levels shown by *cpr5-2* plants. In summary, the upregulation of *PATHOGENESIS-RELATED GENE1* (*PR1*) is in line with my hypothesis.

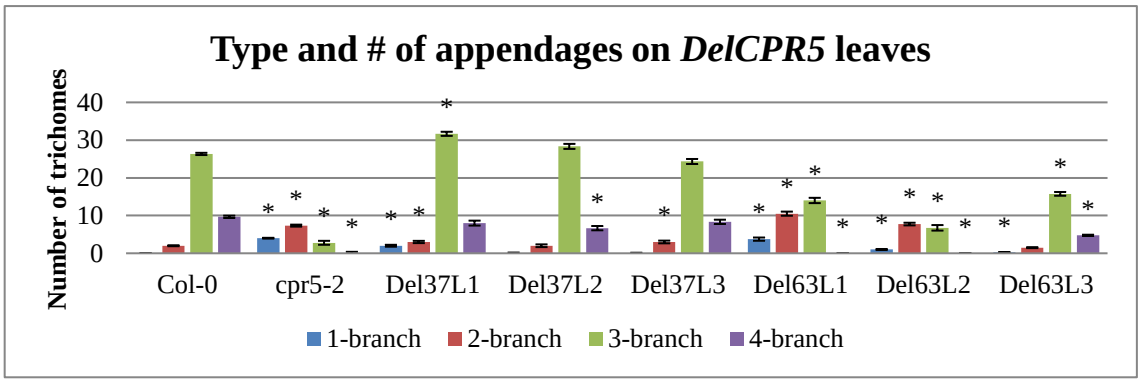


Figure 8.13 Type and average number of trichomes on *DelCPR5* leaves

This figure represents the average number of trichomes and their branching pattern on *Del37* and *Del63* lines. All these lines were grown in soil for 16:8 hours light:dark conditions. Del37 = *Del37CPR5* and Del63 = *Del63CPR5*.

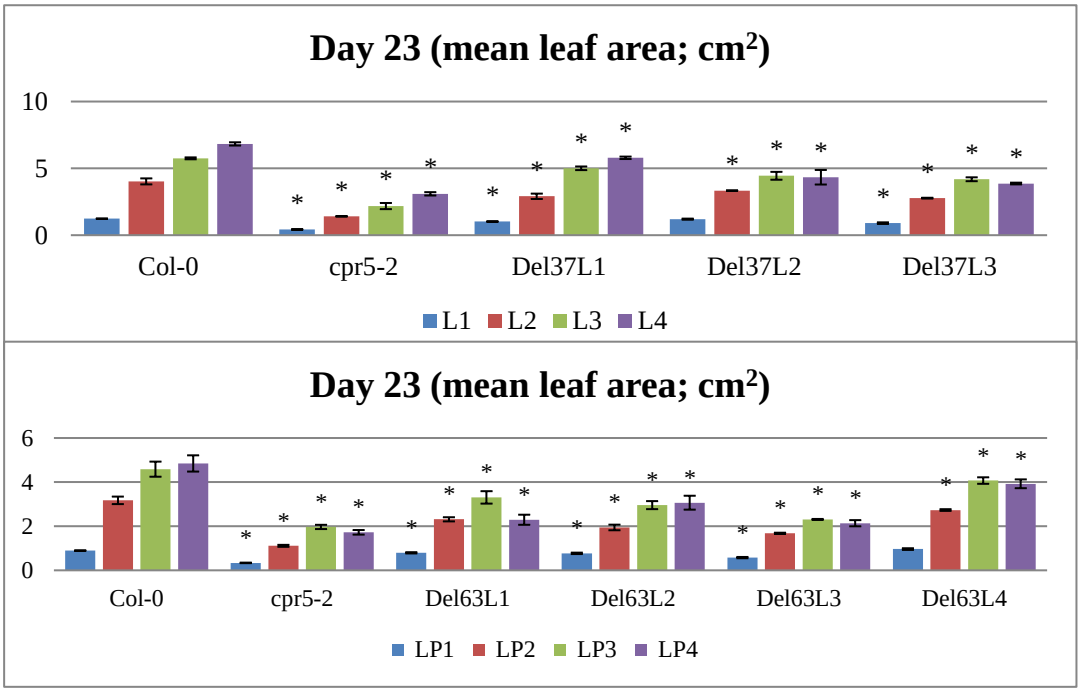


Figure 8.14 Mean leaf area of *DelCPR5*

This figure shows the mean leaf areas of cotyledons and first four rosette leaf pairs at 23 DAS. The leaves were removed from 3-5 different plants of the same line, and the areas were recorded using a leaf area machine. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test) compared to wildtype. Del37 = *Del37CPR5* and Del63 = *Del63CPR5*.

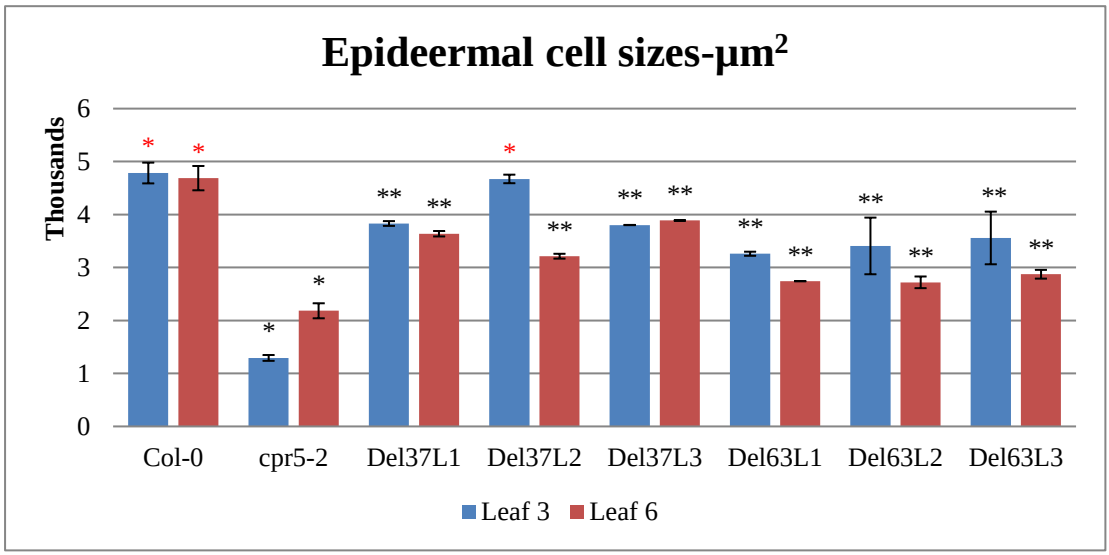


Figure 8.15 Mean area of epidermal pavement cells

This figure shows the average area of cells measured from the images taken by the scanning electron microscope, as described in materials and methods. Areas of the cells were measured which had complete boundaries. The bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant compared to wildtype, * = significant compared to *cpr5-2*; ** = significant compared to wildtype Col-0 and *cpr5-2*. Del37 = *Del37CPR5* and Del63 = *Del63CPR5*.

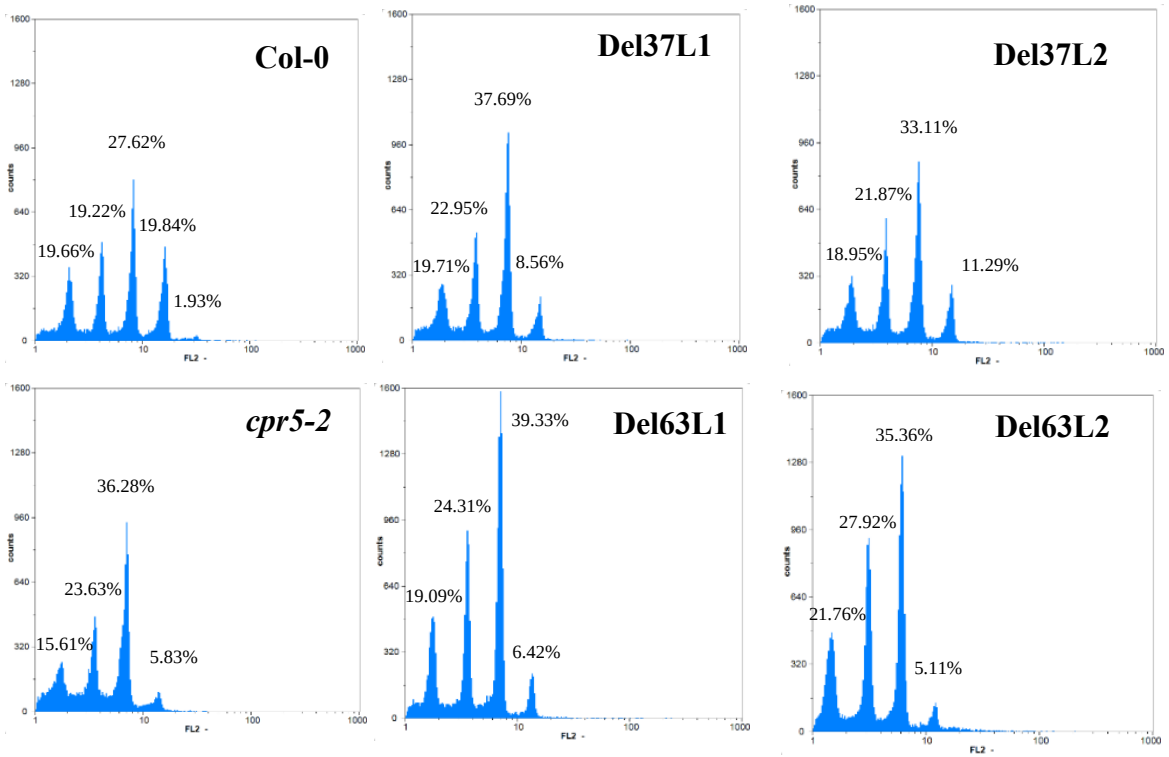


Figure 8.16 Ploidy levels measurement in leaves of *DelCPR5* mutants

This figure shows the proportion or percentages of nuclei populations in the leaf cells of Col-0, *cpr5-2* and *DelCPR5* mutants. The nuclei were extracted from the cells and stained with propidium iodide (PI) as described in materials and methods (Section 2.16). Del37 = *Del37CPR5* and Del63 = *Del63CPR5*.

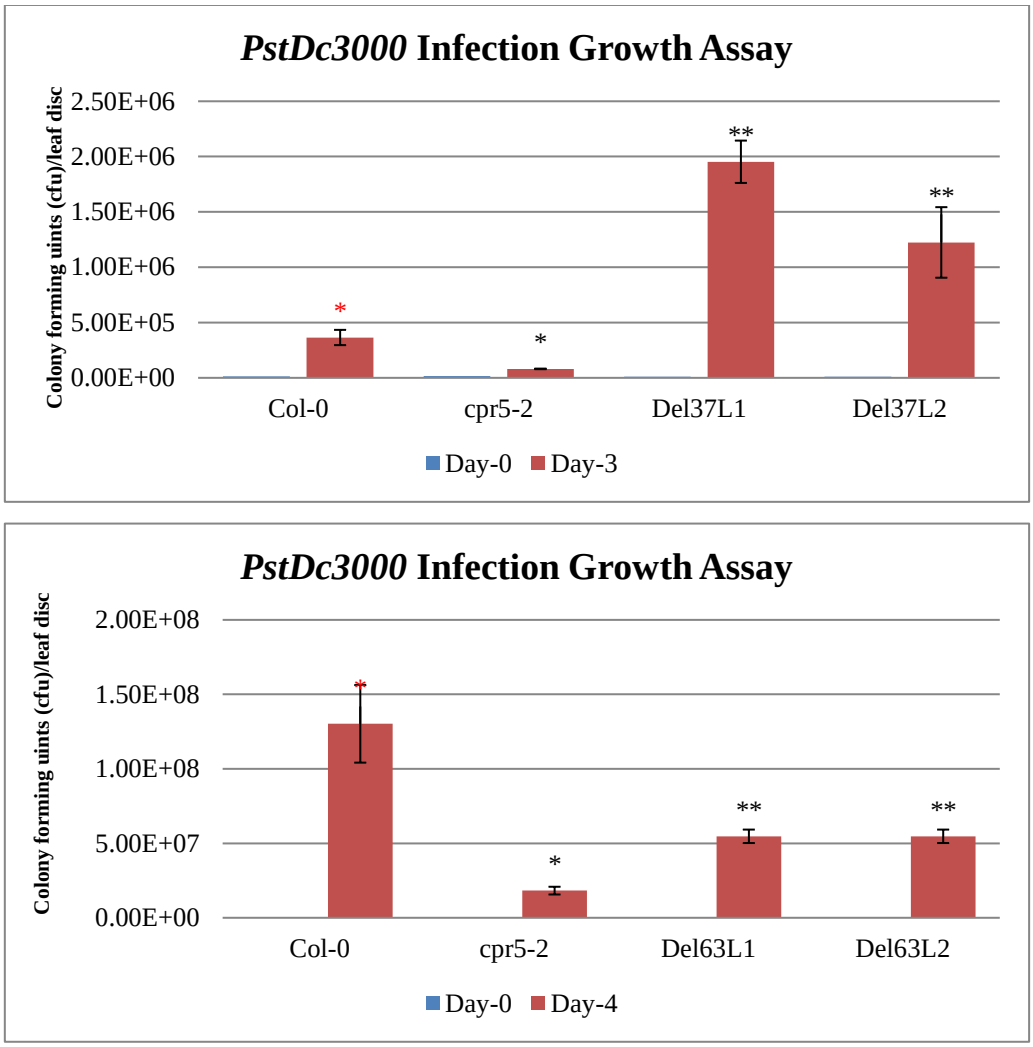


Figure 8.17 *Pseudomonas syringae* infection assay

Colony forming units (cfu) formed by bacteria extracted from leaf discs of each transgenic line were measured and results were reproduced in triplicate as described in materials and methods. Bars represent the standard error, and the asterisks indicate the level of statistical significance at $p < 0.05$ (Students' t-test). * = significant compared to wildtype, * = significant compared to *cpr5-2*; ** = significant compared to wildtype Col-0 and *cpr5-2*. Del37 = *Del37CPR5* and Del63 = *Del63CPR5*.

8.3 DISCUSSION

8.3.1 *In silico* studies suggest CPR5 to be a nuclear protein

CPR5 is believed to be a membrane protein since it was annotated to contain four or five transmembrane (TM) domains (Kirik et al., 2001). In addition to TM domains, a bipartite nuclear localisation signal (NLS) is predicted in the N-terminal part of the CPR5 protein (Kirik et al., 2001). It was proposed that CPR5 protein could embed in the outer nuclear membrane (Kirik et al., 2001; Yoshida et al., 2002). This hypothesis was found to be in line with Gao et al., (2011) results, where they found the CPR5 protein (fused to the GFP at its C-terminal) to be associated with the membrane and localised in the cytoplasm. However, the cytoplasmic localisation of the CPR5 protein was challenged in the same year by Perraza et al., (2011) who found full-length CPR5 protein (fused to the GFP at the N-terminal of the CPR5 protein) to be localised in the nuclear fraction. Recently, Wang et al., (2014) have also confirmed the presence of CPR5 protein in the nuclear fraction.

As mentioned in the introduction, proteins smaller than 45 kDa (Jans and Hubner, 1996) or 40-60 kDa (Breeuwer and Goldfarb, 1990, Jakel et al., 1999), depending on the size of the nuclear pore complex, can enter into the nucleus passively. Additionally, some of the proteins ranging in size between 20-30 kDa, failed to be imported via passive diffusion and required a functional NLS for their active transport (Breeuwer and Goldfarb, 1990). For CPR5, which has a molecular weight of 64 kDa and is believed to be membrane-bound protein, it is unlikely to be transported passively, unless cleaved by proteolytic cleavage and the transmembrane regions excluded (Kirik et al., 2001). Despite nuclear accumulation, none of the studies (Perraza et al., 2011, Wang et al., 2014) have characterised the role of the putative nuclear localisation signal (NLS) in mediating CPR5 nuclear transport. Therefore, the key objective of the current study aims at the elucidation of the role of the annotated NLS in nuclear import. In addition to the previous annotations containing a bipartite NLS (Kirik et al., 2001, Brininstool et al., 2008), the *in silico* and manual scanning of the CPR5 protein carried out in the current study showed that CPR5 contains three clusters enriched in NLS encoding amino acids (KKKKDTNPSNLE KRKLKGKKK). These three NLS-encoding clusters could act as mono, bi- or tripartite

form(s) of the NLS. Clusters A and C of the predicted CPR5 NLS resembles a classical example of the Simian virus 40 (SV-40) large T antigen (PKKKRKV) monopartite class of NLS (Dingwall and Laskey, 1991). At the same time, three clusters (KKKKDTNPSNLEKRKLK GKKK) mimic the classical example of either nucleoplasmin (KRPAATKKAGQAKKKK) (Chatterjee and Stochaj, 1998) or p53 (KRALPNNTSSSPQP KKK) (Shaulsky et al., 1990, Liang and Clarke, 1999) bipartite NLS.

The predicted AtCPR5 NLS also shows resemblance to a novel type of tripartite, EGFR NLS (RRRHIVRKRTLRR), in which all three clusters of EGFR NLS were found to be required for normal protein localisation (Hsu and Hung, 2007)). Based on *in silico* results, CPR5 NLS has a strong resemblance to a number of classical and functionally characterised NLSs. Therefore, either of the putative cluster(s) of CPR5 NLS could be involved in mediating CPR5 nuclear import. Thus, the AtCPR5 NLS could be functional and required for nuclear import. Hence, a number of *nlsCPR5* transgenes were developed. Since the key objective of the current study was aimed at the characterisation of the putative NLS-encoding clusters and their role in AtCPR5 nuclear import, the fusion of a localisation tag may influence the overall structure of *nlsCPR5* recombinant proteins. Similarly, low expression levels of CPR5 may also pose a challenge for protein quantification. Post thesis submission, attempts were made to quantify AtCPR5 from nuclear fractions using MRM-mass spectrometry and AtCPR5 antibody gifted by Wang et al., (2014). Unfortunately, AtCPR5 failed to be detected using both of these aforementioned methods. Keeping in view these limitations, the role of putative NLS-encoding residues was evaluated in the form of degree of complementation of the phenotypes compromised in the *cpr5-2* mutant as discussed below.

8.3.2 Some CPR5 phenotypes/functions appear to be dependent on *CPR5* expression

As shown in Chapter 7, the *CPR5* native promoter has resulted in variable *SynCPR5* expression levels for *SynCPR5* lines, which subsequently resulted in different phenotypes of *SynCPR5* lines. Similarly, the variable transcript abundance of *nlsCPR5* also resulted in differential phenotypes on *nlsCPR5* lines. Thus, it appears that some of CPR5 functions and phenotypes are dependent on CPR5 expression. For example, *nlsCPR5B* lines, which had wildtype-like

expression levels, displayed wildtype-like trichomes. Their leaves display no HR-lesions or early leaf yellowing. Similarly, *nlsCPR5A* lines, which had less transcript abundance compared to the wildtype and *nlsCPR5B*, showed *nlsCPR5B*-like phenotypes. In contrast to *nlsCPR5A* and *nlsCPR5B* lines, *nlsCPR5AB*, *nlsCPR5BC* and *Del63CPR5*, which had slightly higher than *cpr5-2* or *cpr5*-like expression levels, respectively, showed aberrant trichomes. As spontaneous lesions or early yellowing were not found on *nlsCPR5AB*, *nlsCPR5BC* and *Del63CPR5* leaves, these results indicate that trichome development appears to be dependent on expression levels, or that this difference in trichome development could be the result of mutations in clusters. On the other hand, smaller leaves, upregulated levels of resistance and *PR* genes are observed on all the *nlsCPR5* and *Del63CPR5* lines, which suggest that these phenotypes or functions are independent of *CPR5* expression or NLS clusters. Compared to *nlsCPR5A* and *nlsCPR5B*, *SynCPR5L4* and *SynCPR5L6*, despite having wildtype-like *SynCPR5* expression levels, displayed smaller leaves and less trichomes, which suggests that trichome development may also be independent of the *CPR5* expression level and could be the result of mutations. Intriguingly, early bolting, which was observed only in *nlsCPR5AB nlsCPR5BC* and *nlsCPR5ABC* lines separate themselves from *nlsCPR5A* and *nlsCPR5B* lines which bolted in the wildtype manner. Taken together, these results show that the proper execution or development of trichomes appears to be dependent on expression in *nlsCPR5* lines, but independent in *SynCPR5* lines. Thus, expression could play a role in trichome development. Additionally, despite having intermediate expression as compared to *nlsCPR5B* and *cpr5-2*, wildtype-like trichomes on *nlsCPR5A* indicate that a threshold of *CPR5* expression may be required for the normal execution of trichomes in *nlsCPR5*.

8.3.3 Putative NLS clusters appear to have limited function

Mutations in the nuclear localisation signal (NLS) of a protein, result in the failure of nuclear accumulation and loss of function of that protein (Lusk et al., 2007, Gardner et al., 2011, Lokareddy et al., 2015). It was hypothesised that mutations in the CPR5 putative NLS cluster(s) will result in *cpr5-2*-like phenotype(s). Thus, the *nlsCPR5* transgene(s) will be unable to complement *cpr5-2* phenotypes if the NLS cluster(s) has a role in CPR5 localisation and function. The wildtype-like leaves containing no lesions or yellowing and carrying wildtype-

like trichomes on *nlsCPR5A* and *nlsCPR5B* lines, show that *nlsCPR5A* and *nlsCPR5B* recombinant proteins were somehow able to be natively localised and/or accomplish some CPR5 functions similar to the wildtype. On the other hand, smaller leaves and upregulated levels of *PR1* and the resistance of *nlsCPR5*, suggest that nlsCPR5 proteins were unable to recover all CPR5 function, so NLS clusters may be required for the execution of the aforementioned phenotypes of CPR5 functions. As mentioned above in Section 8.3.2, the majority of the CPR5 phenotypes appear to be independent of CPR5 expression levels. Thus, the aberrant phenotypes on *nlsCPR5* plants could be the result of mutations in the putative NLS clusters. Consequently, the mutations in the NLS clusters may have affected CPR5 function either by CPR5 improper localisation (if the putative NLS was functional) or structural changes (if the NLS sites represent another important unknown structural motif), which could have a role in the regulation of leaf size, *PR1* and resistance. Since, the C-terminal half of CPR5 was found to be localised in a wildtype manner (Wang et al., 2014), the NLS clusters appears to have no, or a limited, role in CPR5 localisation. Therefore, it is likely that NLS clusters may represent another structural motif, be part of a structural motif or may have a dual role. In conclusion, *nlsCPR5* characterisation shows that NLS clusters have limited functions as an NLS. However, the evidences presented here are insufficient to conclude their role in CPR5 localisation.

8.3.4 First 37 amino acids of CPR5 are crucial for wildtype-like leaf sizes

As shown in Chapter 6, the disruption of putative start codons in *metCPR5b* plants managed to show bigger leaves than the wildtype. In addition to start codons, the same residues of AtCPR5 were also annotated to be part of an intrinsically disordered region (IDR; Chapters 3 and 4). Structurally, the *Del37CPR5* transgene includes the deletion of the first 37 amino acids of AtCPR5 in addition to the deletion of two of the three putative start codons, which were mutated in *metCPR*5. Phenotypically, *metCPR5b* and *Del37CPR5* both displayed leaves with wildtype-like trichomes and no spontaneous lesions or leaf yellowing. Additionally, both of these (*metCPR5b* and *Del37CPR5*) exhibited enhanced disease susceptibility. These results suggest that *metCPR5b* and *Del37CPR5* share some functions. Since mutations in the start codons in *metCPR5* lines resulted in bigger leaves, the deletion of the first 37 amino acids,

including start codon residues, in *Del37CPR5* was expected to give bigger leaves, if the putative start codons were the cause of leaf size increase in *metCPR5*. There could be a number of plausible explanations for this unexpected result. Firstly, a full-length CPR5 protein is required for normal leaf size in Arabidopsis and the first 37 amino acids are dedicated for this function. Secondly, the presence of a stem-loop structure and putative start codon or translation start site may hinder *CPR5* translation efficiency, which, when mutated in *metCPR5*, could have resulted in increased mRNA stability and concomitantly more protein translation. Consistently, *metCPR5* plants had bigger leaves and enhanced susceptibility. Thirdly, the development of normal trichomes and the absence of spontaneous lesions or early leaf senescence on *Del37CPR5* leaves, reinforces that the CPR5 protein can perform a number of functions except normal leaf size and resistance. Thus, the first 37 amino acids could be dedicated for leaf size and resistance in *Arabidopsis*. To conclude, these results propose a role for this part of AtCPR5 (the first 37 amino acids) in the regulation of leaf size, the exact mechanism of which is yet to be investigated.

8.3.5 *nlsCPR5*-like phenotypes on *DelCPR5* plants suggest NLS clusters to be part of IDRs

Compared to *Del37CPR5* plants, *Del63CPR5* transgenic plants exhibited aberrant trichomes, smaller leaves, smaller epidermal pavements cells, reduced ploidy levels, reduced *CPR5* expression levels and resistance to *Pseudomonas syringae*. These *Del63CPR5* phenotypes were found to be intermediate when compared to the wildtype and *cpr5-2* plants. Phenotypes, such as, aberrant trichomes, smaller leaves, smaller epidermal pavements cells, reduced ploidy levels, reduced *CPR5* transcript abundance and enhanced resistance to *Pseudomonas syringae* as shown by *Del63CPR5*, are identical to the phenotypes executed by *nlsCPR5AB, nlsCPR5BC* and *nlsCPR5ABC* transgenic lines.

Structurally, the whole region (1-98 amino acids) of AtCPR5 was annotated as an intrinsically disordered region (IDR). Specifically, the first 37 amino acids were predicted to contain no other structural motif apart from an IDR. In addition to being annotated as an IDR, a number of structural motifs, such as, three in-frame methionine, three NLS-encoding clusters and a

casein kinase (CKII) phosphorylation site were predicted between the 38-63 amino acids of CPR5 (Figure 8.18). Similarly, the region of CPR5 comprising of 64-98 amino acids was proposed to contain a CKI phosphorylation site and IDR. Thus, there could be four possible explanations for the compromise of phenotypes in *Del37CPR5*, *Del63CPR5* and *Del98CPR5*. First, the smaller leaf size in *Del37CPR5* plants could be the consequence of the loss of part of the AtCPR5 IDR, which could be specifically dedicated for leaf size regulation as discussed above. Second, the abnormal phenotypes could be a consequence of the deletion of predicted NLS-encoding cluster(s). However, the execution of intermediate phenotypes, instead of *cpr5-2*-like phenotypes on *nlsCPR5AB*, *nlsCPR5BC* and *nlsCPR5ABC* plants, negates this hypothesis. Third, the region between 38-98 residues of CPR5 contains a CKI phosphorylation site, which could potentially affect CPR5 function. Fourth, the whole region (1-98 aa) has been annotated as an intrinsically disordered region. Thus, the substitution of residues in *nlsCPR5* or deletion of parts of the disordered region as carried out in *Del37CPR5*, *Del63CPR5* and *Del98CPR5* could have resulted in the loss or change of the IDR and/or protein structure, which subsequently may have led to loss-of-function phenotypes.

Intrinsically disordered regions or proteins are known for their roles in multiple functions and pathways, such as, cell division regulation, cell signalling, ligand binding, and protein-protein interactions with multiple partners (Dunker et al., 2005, Dunker et al., 2001, Dyson and Wright, 2005, Tompa, 2012). Taking into account the flexible structural nature and the involvement of intrinsically disordered region(s) or protein(s) in various molecular events, such as, proteins, membranes, nucleic acid and small molecules (Tompa and Csermely, 2004, Uversky et al., 2005, Oldfield et al., 2008, Oldfield and Dunker, 2014), AtCPR5 appears to be an IDR in nature. Hence the defects in various biological processes in *cpr5* plants e.g. cell death regulation, senescence, oxidative stress, cell division, cell development and cell expansion, are in line with the multiple functions performed by AtCPR5.

In addition, disordered regions protect or safeguard some of the interacting partners (repressors and regulators) in order to control the expression of genes (Dunker et al., 2001). For example, a few chaperones were found to be safeguarding intrinsically disordered regions in order to avoid non-specific binding or interaction (Dunker et al., 2001). As AtCPR5 has been implicated

in the suppression of cell death regulation by binding and controlling cell death regulators (Boch et al., 1998; Kirik et al., 2001), it could be possible that AtCPR5 IDRs may be involved in the binding and safeguarding of some of the proteins that could either repress or regulate some other genes. This hypothetical role of AtCPR5 or AtCPR5 IDRs is in line with the model presented by Wang et al., (2014) and Gu et al., (2016) i.e. AtCPR5 embeds in the nuclear membrane and safeguards one of the CDKs. This notion needs further structural and interaction evidence. Thus, the structural analyses of AtCPR5 protein in the presence and/or absence of potential interacting partners using NMR could help to understand its binding, interaction and the role of disordered regions (IDRs) in various pathways.

In conclusion, evidence that NLS-encoding clusters are required for AtCPR5 proper functions and may have no role in AtCPR5 nuclear import, are presented in the current chapter. Thus, putative NLS-encoding clusters appear to have limited roles in CPR5 nuclear import. Moreover, this study reinforces and proposes AtCPR5 to be an intrinsically disordered protein, which is a novel observation.

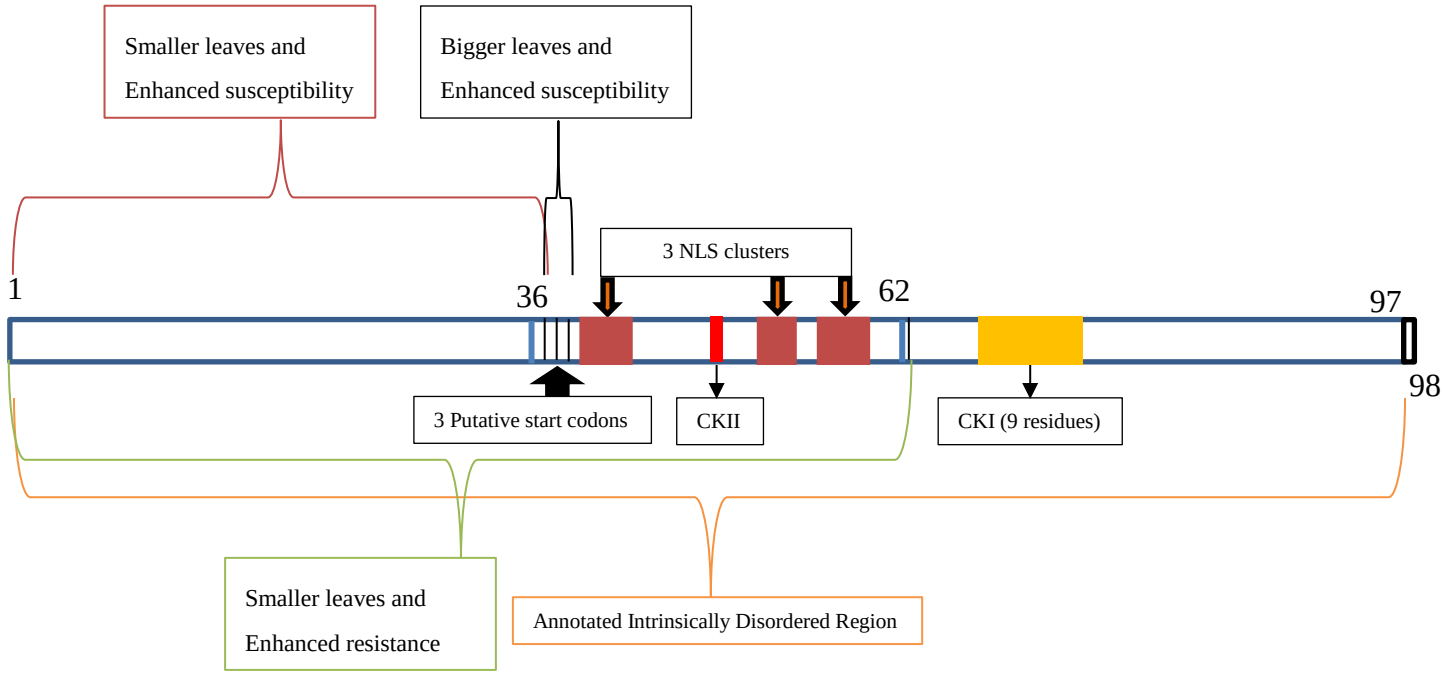


Figure 8.18 Position of structural motifs present in CPR5 first 98 amino acids

This figure shows the estimated (not scaled) positions of various predicted important structural motifs present in the N-terminal half (first 98 amino acids) of AtCPR5. The number depicts the position of amino acids, and similar colors represent the same structural motif. Arrows represent estimated positions. In-frame methionine residues reside at positions 1, 37, 38, 39, 63 and 98 of the AtCPR5 protein.

Chapter 9 Characterization of putative CPR5 leucine and glycine motifs

9.1 INTRODUCTION

Coiled-coil domains are important structural motifs present in many eukaryotic proteins (Lupas and Gruber, 2005). The proteins carrying coiled-coil motifs participate in different types of interactions, such as, protein-protein, DNA-proteins, or act as transcription factors, molecular motors, receptors, signalling molecules, and stalks of surface proteins (Vinson et al., 1989, Krylov and Vinson, 2001, Martin-Serrano et al., 2004, Lupas and Gruber, 2005, Chae et al., 2006, Chevalier et al., 2008). Structurally, coiled-coil domains are long, rigid segments or helices, which assemble into long and mechanically rigid structures, for instance, hair, horn, keratin, and fibrinogen. Similarly, coiled-coil domains also act as extracellular matrices, such as, laminin, and cytoskeletal networks e.g. intermediate filaments (Lupas et al., 2005). Due to their diverse nature, coiled-coil domains have been exploited in medicine. For instance, coiled coils were used for the targeting and suppression of the adenomatous polyposis coli protein; a colorectal cancer associated protein, (Woolfson, 2005). The coiled-coil domains were also used to target and inhibit the HIV virus (Sharma et al., 1998).

Coiled-coil domains consist of segments of 7 residues, which contain a periodic repetition of leucine residue at position *d* in each segment of the 7 residues, known as a *heptad* repeat (Lupas et al., 2005). The α-helices of two coiled coil domains twist around each other and form a zip like structure, in which the hydrophobic side-chains are packed inside the protein core, whereas, hydrophilic residues protrude outside to interact with partner(s). The coiled coils of a protein can homodimerize as well as hetrodimerize (Krylov and Vinson, 2001, Lupas and Gruber, 2005).

In addition to coiled-coil motifs or leucine zippers, glycine zipper or motif(s) are of central importance for proper folding and insertion into the membranes. The typical glycine sequence

patterns include, (G,A,S)XXXGXXXG or GXXXGXXX(G,A,S) where the position of glycine (G) is conserved in the motif. X can be any residue such as glycine, alanine or serine can present at the start or the end of the motif. In the glycine motif, GXXXG is the key pattern for the interaction and dimerization of transmembrane helices (Kim et al., 2005). Similar to leucine zippers, the glycine motif follows a conserved pattern and is majorly involved in a variety of oligomerizations. Therefore, it was termed the glycine zipper (Landschulz et al., 1988). In many studies, the substitution of glycine residue in the GXXXG motif resulted in loss-of-function phenotypes.

As mentioned in Chapter 3, CPR5 is predicted to contain leucine as well as glycine zippers, therefore transgenic lines were developed including substitutions of both types of zippers in order to characterize their role in conferring CPR5 function. The current chapter covers complementation of the *cpr5-2* plants transformed with *CPR5* transgenes, defective in both types of zippers.

9.2 RESULTS

9.2.1 AtCPR5 protein is predicted to contain 4 coiled-coil domains

As mentioned in Chapter 3, the CPR5 protein is predicted to contain coiled-coil motifs. According to the coiled-coil model, the presence of nonpolar (hydrophobic) residues in the first (*a*) and fourth (*d*) positions of the *heptad* (*a, b, c, d, e, f, g*) is found to be preferential in coiled-coil domains (Burkhard et al., 2001). Thus, it was hypothesised that leucine or either of the hydrophobic residues present at position *a* and *d*, are a structural part of coiled-coil domains in AtCPR5. To test, leucine residues of each *heptad* were changed by substitutions in *SynCPAR5* by GeneScript. As shown in Figure 9.1A, the predicted regions of the AtCPR5 protein contain leucine or hydrophobic residue(s), which are present in position *d* of the *heptad* in the predicted coiled-coil motifs. Therefore, the substitution of these predicted leucine residues (L) of each annotated *heptad* with asparagine (N) residues, allow the testing of their effect on CPR5 functioning. In this way, four different types of constructs were designed with the introduction of mutations in the leucine residues, individually as well as in different combinations (Table 6.3). Coiled coil transgenes were named as *ccdCPR5* (*ccd1CPR5*, *ccd2CPR5*, and *ccd3CPR5*). Successfully transformed lines of coiled-coil motifs were grown in the soil, screened for segregation and characterised in order to see the recovery of phenotypes compromised in *cpr5-2*. The disruption of functional leucine (a structural part of the *heptad*) is expected to show *cpr5-2*-like phenotypes.

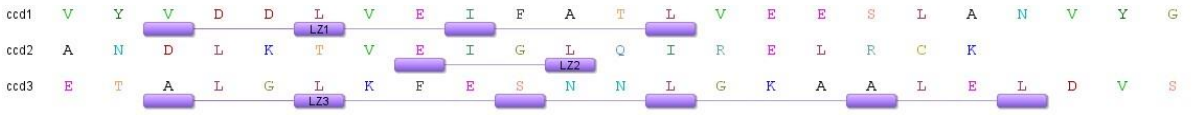
9.2.2 AtCPR5 protein is predicted to contain glycine motifs

In *cpr5-1* (Bowling et al., 1997) and *old1-1* (Jing et al., 2002) mutants, the glycine residues present at the 420 and 459 positions were found to be converted into aspartic acid (D) and serine (S), respectively. The physiological characterisation of *cpr5-1*, *old1-1* and *cpr5-2* show that conversion of glycine into aspartic acid (*cpr5-1*) and glycine into serine (*old1-1*) has a similar effect on plant phenotypes as shown by the mutation of glycine into a stop codon in *cpr5-2*. The amino acid sequence alignment of AtCPR5 and its homologue protein sequences revealed that the position of glycine residues (420 and 459 aa) appears to be conserved and

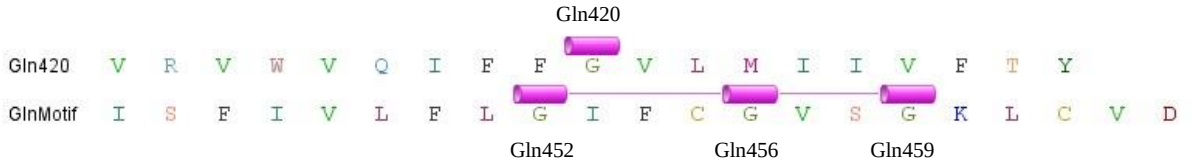
present in the majority of AtCPR5 homologues. Thus, it was hypothesised that glycine residues at positions 420 and 459 are part of a structural motif, and thus represent loss-of-function phenotypes upon mutation. To confirm, firstly the AtCPR5 protein sequence was analysed through a number of *in silico* tools, and literature was reviewed in order to find out a structural motif containing a conserved glycine residue. The *in silico* analyses of AtCPR5 protein sequence revealed that firstly, the regions containing glycine 420 and 459 residues are part of the predicted transmembrane regions of the CPR5 protein. For example, glycine 420 is present in the second TM domain, whereas, glycine 459 is part of the third TM domain. Secondly and most importantly, the region containing glycine 452, 456 and 459 predicts as a glycine motif, and glycine at 452 and 459 were predicted as integral residues of glycine motif (**G**IFC**G**VS**G**). This annotation excludes glycine 456 as part of the glycine motif as shown in Figure 9.1B.

As shown in Figure 9.1B, proposed glycine residues were mutated and a number of constructs were developed for glycine mutants in order to see their effects on plant phenotypes. These mutations were introduced by GeneScript and mutants were named as *Gln420CPR5*, *Gln452CPR5*, *Gln456CPR5*, and *Gln459CPR5*, depending on the position of glycine residues. Successfully transformed lines of glycine were grown in the soil, screened for segregation and characterised in order to see the recovery of phenotypes compromised in *cpr5-2* plants.

A.



B.



C.

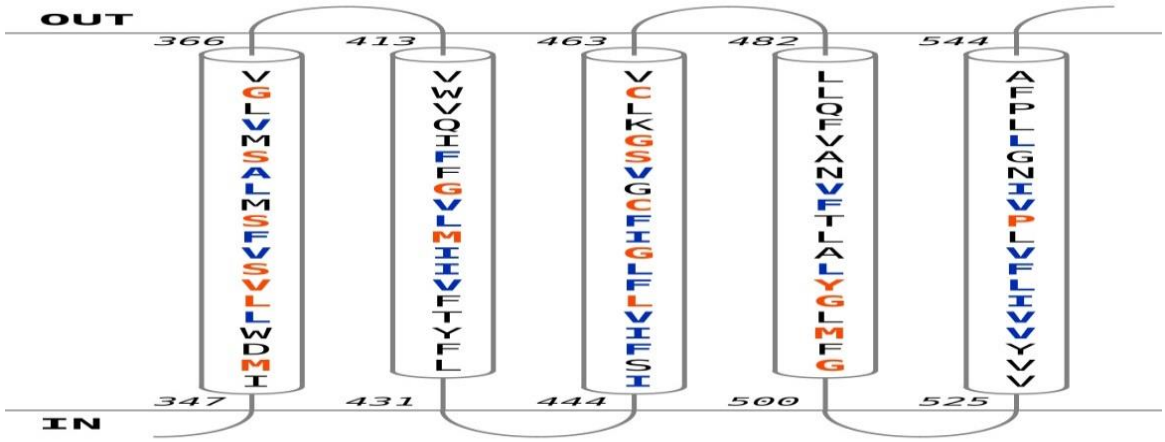


Figure 9.1 Position of putative leucine and glycine residues of leucine and glycine motifs

This figure shows the position of leucine (9.1A) and glycine (9.1B) residues of the predicted coiled-coil and glycine motifs. These figures were prepared using Geneious. The critical residues of the CPR5 transmembrane and glycine motifs, highlighted in orange font, are depicted in 9.1C.

9.2.3 *ccdCPR5* plants display *cpr5*-like phenotypes

As mentioned above, the position of leucine residues is conserved, and any mutation in their positions was expected to result in the loss of structure of these coiled-coil domains. Thus, *ccdCPR5* transgenic plants were hypothesised to display compromised phenotypes identical to *cpr5-2* plants. To test, *ccdCPR5* transgenes were transformed into *cpr5-2* plants and screened, unless homozygous plants of each line were achieved. As shown in Figure 9.2, none of the *ccdCPR5* transgenes managed to rescue *cpr5-2* compromised phenotypes. *ccdCPR5* transgenic plants displayed phenotypes identical to phenotypes on *cpr5-2*. For example, the leaves on the *ccdCPR5* transgenic lines were shown to possess HR-like lesions and displayed early leaf senescence for cotyledons and rosette leaves (Figure 9.2). Similarly, *ccdCPR5* transgenic plants produced trichomes in a similar fashion to *cpr5-2* plants i.e. trichome typical patterns and their appendages were missing on the leaves of these plants compared to wild-type Col-0 plants (data not shown). In conclusion, the inability of *ccdCPR5* to restore compromised phenotypes in *cpr5-2* plants suggest that the leucine residues of the each putative coiled-coil motif are likely to be a structural part of the coiled-coil motifs, and hence, their presence is essential for CPR5 proper functioning. Due to the accomplishment of *cpr5-2* phenotypes, the *ccdCPR5* transgenic lines were marked as non-complimented lines and were not further characterised at genetic or functional levels.

9.2.4 Glycine residues of the glycine motif show intolerance to mutations

As mentioned above and shown in Chapter 3, glycine residues at 420 452, 456 and 459 positions of CPR5 represent an integral part of the structural motifs of second and third TM domains of the AtCPR5 protein. Thus, the introduction of mutation(s) in the selected glycine residues will result in the loss-of-function phenotypes on *GlnCPR5* plants, which constitutes our hypothesis. To test, the *cpr5-2* plants were complemented with *GlnCPR5* transgenes in order to confirm the role of glycine residues. Contrary to the compromised phenotypes on the *cpr5-1* plants (Bowling et al., 1997), the *Gln420CPR5* transgene carrying the substitution of glycine 420 into alanine, managed to restore compromised phenotypes of *cpr5-2*. For example, *Gln420CPR5* transgenic plants displayed leaves without HR-like lesions and/or early leaf

senescence compared to CPR5 wildtype Col-0 (Figure 9.2). Like *Gln420CPR5*, *Gln456CPR5* transgenic plants (carrying substitution of glycine at position 456) also managed to rescue *cpr5-2* compromised phenotypes (Figure 9.2). Contrarily, none of the glycine transgenes (*Gln452CPR5* and *Gln459CPR5*), managed to recover *cpr5-2* compromised phenotypes. For example, *Gln452CPR5* and *Gln459CPR5* plants both exhibited leaves with HR-like lesions and leaf yellowing in a *cpr5-2*-like manner (Figure 9.2). In summary, glycine substitutions at positions 452 and 459 of the glycine motif were unable to complement *cpr5-2*, compared to the substitution of glycine residues present at positions 420 and 456 in CPR5.

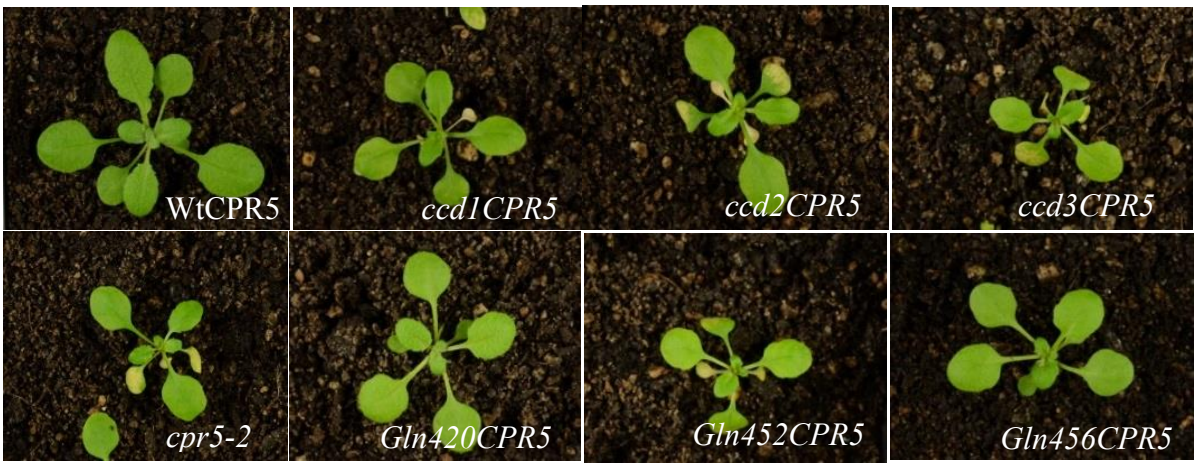


Figure 9.2 Plant growth and early leaf senescence on *ccdCPR5* and *GlnCPR5* leaves

This figure presents a comparison of plant growth and early leaf senescence on *ccdCPR5* and *GlnCPR5* leaves as compared to wildtype. Early leaf senescence or yellowing on the leaves of all of the coiled-coil domain (*ccdCPR5*) and the majority of *GlnCPR5* transgenic lines including, *cpr5-2* plants at 18 DAS can be observed. All the plants of *ccdCPR5* and *GlnCPR5* transgenic lines were grown at 22°C and 65% humidity in conditions of 16 hours light and 8 hours dark.

9.3 DISCUSSION

9.3.1 Leucine residues are crucial for CPR5 functioning

As mentioned in the introduction, coiled-coil domain(s) participate in many biological processes and play crucial roles. *In silico* studies predicted the presence of leucine zippers in AtCPR5. The position of leucine residues in the *heptad* is consistent with the typical leucine zipper *heptad* pattern (Ciani et al., 2010). In order to confirm the existence and role of coiled-coil domains, the typical pattern of the *heptad* is disturbed through the introduction of mutations that disrupt putative leucine residues present at the fourth position (*d*) of the *heptad*. For instance, the substitutions of leucine residues with proline (P) residues have abolished protein activity (Harbury et al., 1993, Cheng et al., 2001, Lupas and Gruber, 2005, Tanaka et al., 2007, Ciani et al., 2010). In some studies, leucine was mutated into alanine (A), which successfully resulted in the loss of function. However, the conversion of leucine into alanine in one study failed to disrupt the coiled-coil domain structure, since this leucine to alanine substitutions at both positions (*a* and *d*) of the *heptad*, displayed no difference in the helix stability (Zhou et al., 1992). In the current study, leucine was substituted with asparagine (N) residues, which successfully abolished protein function and resulted in the accomplishment of *cpr5-2* like phenotypes on the *ccdCPR5* transgenic lines, confirming that asparagine (N) is also a good candidate to mutate coiled-coil domain stability and activity. Moreover, these results strongly indicate an important role of leucine residues for AtCPR5 proper functions, possibly by taking part in the formation of coiled-coil domains.

9.3.2 Glycine 452 and 459 are also crucial for CPR5 functions

Like leucine zippers, glycine zippers or motifs are also an integral part of the transmembrane domains of several proteins, and mutation(s) in the glycine residues of glycine motifs are often deleterious (Kim et al., 2005). As shown by Bowling et al., (1997) and Jing et al., (2002), a single nonsense/non-synonymous mutation (G420D and G459S) led to *cpr5-2*-like phenotypes in *cpr5-1* and *old1-1*, respectively. Additionally, both the glycine residues are annotated to be part of the TM domains. The conversion of glycine (a neutral and the most flexible amino acid),

which generally buries inside TM domains, into aspartic acid (an acidic and negatively charged residue), could have resulted in conformational change of the AtCPR5 protein structure. This conformational change or failure of AtCPR5 to be embedded in the membrane caused loss-of-function of the AtCPR5 protein. However, the substitution of glycine with serine (G459S), and the execution of *cpr5*-like phenotypes on *old1-1* plants is the most interesting, since serine (S) is present in other AtCPR5 annotated TM domains, indicating that serine can be a part of the TM domain. On the other hand, the execution of wildtype-like phenotypes on *Gln420CPR5* (G420A) plants show that glycine can be swapped with alanine in the TM domain.

Recently, Venco et al., (2015) found that a glycine substitution with serine can abolish dimerization of a glycine zipper which, in turn, directs the protein of interest to the endoplasmic reticulum and mitochondrial-associated membrane. On the other hand, Kim et al., (2005) have shown that valine, leucine, Isoleucine, serine and threonine are not good substitutes for glycine motif disruptions, when compared to alanine, since alanine abolishes a stable helix formation. The accomplishment of *cpr5*-like phenotypes on *Gln452CPR5* and *Gln459CPR5* plants (Figure 9.2) reinforce Venco et al., (2015) results that alanine is one of the amino acids that can successfully disrupt the glycine motif. Thus, glycine present at 420, may not be part of a glycine motif compared to glycine present at 459. Hence *Gln420CPR5* plants managed to tolerate G420A substitution and displayed wildtype-like phenotypes, whereas, *Gln459CPR5* failed to compensate for this change (G459A) and exhibited *cpr5*-like phenotypes. These results are in line with predictions (Figure 9.1C) that suggest that no glycine motif in the second TM domain contains glycine 420, since the TM domain lacks the typical consensus sequence (GXXXGXXG) for a glycine motif. In contrast, the third TM domain (containing glycine 459) is shown to contain a typical consensus sequence (GXXXGXXG) and strongly annotates as glycine motif (Figure 9.1C). Therefore, glycine 420 may not be as structurally important in the same way as glycine 459.

Having established that glycine (420) is not part of a glycine motif and *Gln420CPR5* plants display wildtype-like phenotypes, it raises a question about how G420D (*cpr5-1*) conversion results in *cpr5*-like phenotypes. However, loss-of-function phenotypes of *cpr5-1* are in line with several studies where the activities of proteins were significantly reduced or abolished

when aspartic residue was introduced in transmembrane regions of the proteins (Mauzy and Hermodson, 1992, Haimeur et al., 2004, Freeman-Cook et al., 2004). The most plausible explanation could be the difference of charge, structure and propensity of aspartic acid and alanine residues, which could either hinder or contribute to the stable helix formation and/or membrane insertion. Structurally, glycine and alanine both are hydrophobic amino acids and are identical in size as compared to aspartic acid, which is hydrophilic in nature and bigger in size. Thus, the G420A hydrophobic substitution was tolerated as compared to G420D hydrophilic substitution. Alternatively, alanine could have promoted helix dimerization in the same way as it promoted glycine helices dimerization in (Harmeier et al., 2009) studies.

On the other hand, *cpr5*-like phenotypes on *Gln452CPR5* and *Gln459CPR5* transgenic plants are also consistent with glycine motif prediction (Figure 9.1C), which suggests glycine residues (452 and 459) as part of the glycine zipper in the third TM region. As shown in Figure 9.1C, this prediction excludes glycine 456 and suggests that only glycine residues present at 452 and 459 positions are involved in interacting and making interfaces between two helices in order to facilitate the dimerization of transmembrane helices of AtCPR5. The accomplishment of wildtype-like phenotypes on *Gln456CPR5* plants whereas *cpr5*-like phenotypes on *Gln452CPR5* and *Gln459CPR5* plants, is consistent with the idea that glycine 452 and 459 could be part of glycine zipper as predicted and shown in Figure 9.1C. Thus, these results indicate that glycine residues at 452 and 459 positions are crucial for AtCPR5 function, and that these glycine residues appear to be part of the glycine motif. In addition, substitutions in glycine residues (452 and 459) are detrimental, compared to the glycine (456) in the non-glycine motifs of the transmembrane regions.

In summary, current evidence suggests that leucine residues of the annotated coiled-coil domains and glycine residues of glycine motifs are crucial for AtCPR5 proper functions. Based on current information, leucine and glycine residues appear to be part of the said motifs, however, this needs further investigation.

Chapter 10 FINAL DISCUSSION AND FUTURE PROSPECTS

10.1 AtCPR5 is a nuclear membrane protein

The *in silico* studies carried out in the current thesis show that NLS-encoding clusters have a strong resemblance to a number of already characterised classical NLSs from multiple species (Chapter 8). The presence of three NLS clusters suggests that AtCPR5 NLS could act as a monopartite, bipartite or tripartite NLS (Chapter 8). It was assumed that if NLS-encoding cluster(s) were active and functional, the complementation of *cpr5-2* plants with *nlsCPR5* transgenes would result in *cpr5*-like phenotypes i.e. a failure to rescue the compromised phenotypes of *cpr5*. As shown in Chapter 8, the *nlsCPR5* transgenes were shown to partially complement *cpr5-2*; i.e. some of the phenotypes were fully recovered, whereas, others were partly rescued. For instance, none of the *nlsCPR5* transgenic lines displayed HR-like lesions or early leaf yellowing compared to *cpr5*, except the *nlsCPR5* transgenic lines, which were classified as non-complemented plants. Similarly, trichome aberration as shown by *cpr5*, was fully rescued in *nlsCPR5A* and *nlsCPR5*, compared to *nlsCPR5AB*, *nlsCPR5BC* and *nlsCPR5ABC* (Chapter 8). Contrarily, all of the *nlsCPR5* transgenic plants displayed intermediate-sized leaves when compared to the wildtype and *cpr5* leaves. Additionally, *nlsCPR5*, *PR1*, and *PR5* transcript abundances were affected to varying extents. For example, while position effects on *nlsCPR5* transgene expression cannot be excluded, the results indicate that *nlsCPR5B* plants displayed transcript abundance insignificantly different from the wildtype, whereas transcript abundance of *nlsCPR5A* was significantly less than the wildtype. Similarly, *nlsCPR5BC* transcript abundance was insignificant from *cpr5* but was significantly less than the wildtype. In contrast to *nlsCPR5BC*, transcript abundance of *nlsCPR5AB* was significantly higher than *cpr5* and less than the wildtype. *PR1* and *PR5* levels were also found to vary among *nlsCPR5* lines. Lastly, all the *nlsCPR5* lines showed increased resistance to virulent *P. syringae*, which was similar to *cpr5* mutants. Depending on the expression and mutation, these results demonstrate a degree of dependence of some of the phenotypes on the NLS cluster(s). Despite having a defective (*nlsCPR5*) or even no NLS (*Del63CPR5)*, how

nlsCPR5 recombinant proteins managed to execute functions or phenotypes to varying levels, is obscure. At the same time, these results suggest limited implications of NLS clusters in CPR5 localisation. Moreover, the localisation of a nuclear protein with a defective NLS should be non-native and the function of the protein should be abolished completely, which is not the case in the *nlsCPR5* and *Del63CPR5* lines.

A majority of integral inner nuclear membrane proteins are predicted to have NLS-like sequences, which help proteins to translocate to the nuclear membrane (Egea et al., 2005, Lusk et al., 2007). Since localisations of *nlsCPR5* transgenes were not shown in the current study, it is difficult to relate the effect of mutations in NLS clusters on AtCPR5 nuclear localisation. However, partial complementation (intermediate leaf size, the rescue of HR-lesions and early leaf yellowing) in *nlsCPR5ABC* and *Del63CPR5*, where all three of the NLS clusters were deleted, suggests that NLS clusters may not be required for AtCPR5 nuclear membrane localisation *per se*. These results also suggest that the AtCPR5 protein would have somehow managed to be natively localised to the NE (Nuclear Envelope) in *nlsCPR5ABC* and *Del63CPR5* lines, and hence rescued HR-like lesions and early leaf yellowing phenotypes. AtCPR5 localisation in the nuclear lemma is further supported by Gu et al., (2016), who have recently shown AtCPR5 to be localised in the nuclear envelope (NE) and proposed CPR5 to be a part of the nuclear pore complex (NPC). Despite having putative classical NLSs, the role of NLS clusters in the AtCPR5 nuclear import is further questioned by the localisation of the C-terminal half (275-564 aa) of AtCPR5 in the nuclear membrane, similar to the wildtype protein (Gu et al., 2016). Moreover, the presence of TM domains (Chapter 3) and functional glycine motif within the third TM domain (Chapter 9) of AtCPR5 reinforces AtCPR5 to be a nuclear envelope protein.

Having established that AtCPR5 is a nuclear membrane protein (Gu et al., 2016), how AtCPR5 exclusively targets the nuclear membrane compared to other membranes, and especially when AtCPR5 NLS clusters are deleted in *nlsCPR5*, poses an important question. Membrane proteins are concomitantly translated by translation machinery and targeted to its destined membrane depending on the signal peptide (Egea et al., 2005, White and von Heijne, 2008, Luirink et al., 2012). Following recognition of the membrane signal (by Alu domain), which is present just

before the start of the first transmembrane domain, by the signal recognition particle (SRP) and its receptor (SR), the translation of nascent protein is arrested, and nascent protein is then directed to endoplasmic reticulum due to the hydrophilic environment of the cytoplasm (Fluman et al., 2014). Subsequently, depending on the type of signal peptides (nuclear, mitochondrial, or chloroplast), protein is directed to its target membrane(s) (Fluman et al., 2014). Keeping in view these results, in contrast to intermediate phenotypes, the localisation and function of nlsCPR5 recombinant membrane proteins is expected to be non-native and non-functional respectively. The nuclear envelope (NE), especially the outer membrane, is a specialized continuation of the rough endoplasmic reticulum (Ribosomes-studded; Lusk et al., 2007). Similarly, as the perinuclear space between the inner and outer membranes is continuous with the ER lumen (Laba et al., 2014), it could be possible that following translation by ribosomes studded onto the outer nuclear membrane, AtCPR5 could become part of the nuclear membrane or translocate passively to the NE (Laba et al., 2014). Since CPR5 is reported to be localised in the endoplasmic reticulum (Gu et al., 2016), this indicates non-specific localisation and supports ER-mediated CPR5 translocation as well. Whereas, based on the native localisation of the C-terminus of CPR5 (Gu et al., 2016), AtCPR5 translocation to the nuclear membrane appears to be independent of NLS clusters. Thus, these results led me to propose that NLS clusters may not be required for AtCPR5 nuclear membrane targeting. However, they are crucial for AtCPR5 proper functioning and may contribute to efficient localisation, or provide a function that is not related to nuclear localisation. The quantifications and localisations of nlsCPR5 recombinant proteins need to be investigated in order to show reduction in protein localisation to the NE, due to mutation(s) in the putative NLS clusters, and to relate these reductions with phenotypes or functions.

10.2 CPR5 may exert pleiotropic effects by modifying NPC-mediated selectivity and entry of nuclear proteins

CPR5 is a part of the nuclear pore complex and could be one of the nucleoporins (Gu et al., 2016). Being a nucleoporin, CPR5 is shown to manipulate the opening and closing of NPC depending on the role of CPR5. For example, CPR5 is shown to restrict the entry of NPR1,

JAZ1 and ABI5 under normal conditions, since cytoplasmic retention of NPR1, JAZ1 and ABI5 was compromised in the *cpr5* allele *old1-3* (G120D) (Gu et al., 2016). Additionally, the N-terminus of AtCPR5 was shown to be implicated in CPR5 homomerization which is abolished in *old1-3*. Therefore, Gu et al., (2016) proposed that the NPC restrict the entry of the aforementioned proteins through CPR5-mediated (dimerization) gating. Moreover, since *cpr5-2* (G477*) and *old1-3* (G120D) plants show similar phenotypes, it could be possible that loss-of-function phenotypes shown by *cpr5-2* could also be a result of improper entry of NPR1, JAZ1 and ABI5 proteins in *cpr5*. Thus, keeping in view the results obtained by Gu et al., (2016) and shown in the current thesis, a tempting working model of CPR5 was constructed here. To understand this model, it is important to assume that 1) CPR5 is a nuclear membrane protein, which facilitates NPC for the selection of nuclear cargoes, and 2) CPR5 recombinant proteins are able to be localised in the nuclear membrane in a wildtype manner. Keeping in view the implication of the CPR5 N-terminus in dimerization and restriction on the entry of nuclear proteins, SynCPR5 recombinant protein being identical to the wildtype CPR5 protein, would behave in a wildtype manner and hence restrict or allow the same proteins (Figure 10.1A and 10.1C). Compared to wildtype, the selectivity of nuclear proteins is believed to be abolished in *cpr5* plants (Figure 10.1B) as proposed by Gu et al., (2016). Since, the cpr5-2 mutant protein is present in the majority of *CPR5* transgenic lines in the background, the cpr5-2 protein could exert a negative (dominant negative allele) effect on CPR5-mediated NPC gating. Therefore, the degree of effect on gating (closed, partly open or open) would largely depend on the relative abundance of each type of recombinant protein.

On the other hand, CPR5-mediated NPC selectivity is proposed to be variable depending on the effect of mutations on CPR5 structural changes, resulting in CPR5 recombinant proteins. The *nlsCPR5A* and *nlsCPR5B* lines have higher transcript abundances than *nlsCPR5AB*, *nlsCPR5BC* and *nlsCPR5ABC*, and rescue more phenotypes as compared to *nlsCPR5AB*, *nlsCPR5BC* and *nlsCPR5ABC*. It could be possible that a higher amount of nlsCPR5A and nlsCPR5B recombinant proteins would have a better or wildtype-like selectivity for some of the nuclear proteins (Figure 10.1D) as compared to nlsCPR5AB, nlsCPR5BC and nlsCPR5ABC. Hence, the *nlsCPR5A* and *nlsCPR5B* lines showed the majority of wildtype-

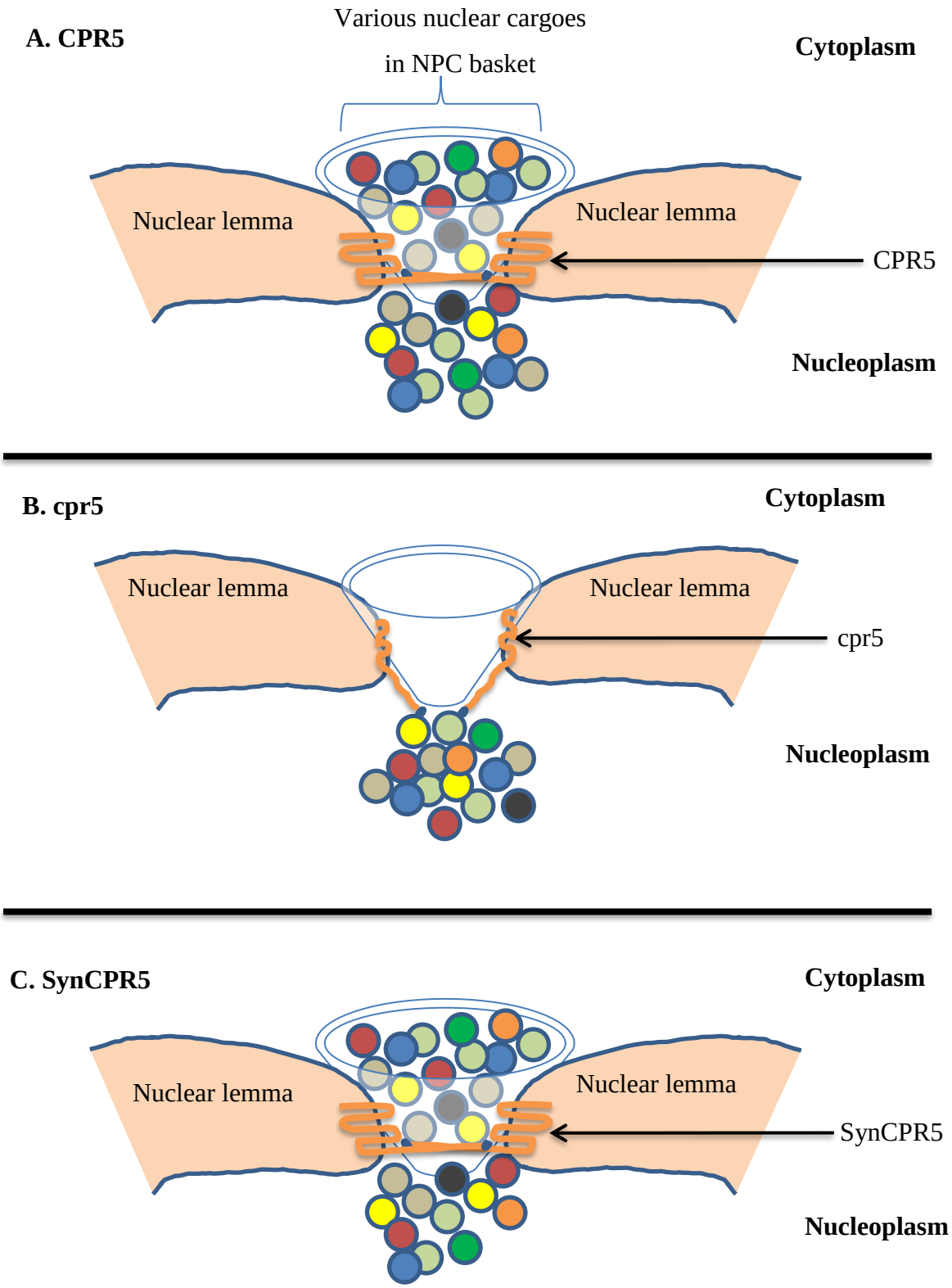
like phenotypes. The accomplishment of wildtype-like phenotypes, such as, normal trichomes, susceptibility to *PstDC3000* and healthy green leaves without spontaneous lesions or yellowing, in *Del37CPR5* plants show that the first 37 amino acids of AtCPR5 could have no effect on the selection and restriction of proteins required for the regulation of the said phenotypes. However, the restriction of proteins required for the regulation of resistance appears to be abolished in the *nlsCPR5A* and *nlsCPR5B* lines (Figure 10.1D). Additionally, intermediate (bigger than *cpr5* but smaller than the wildtype) sized leaves on *Del37CPR5* plants suggest the partial restriction of proteins required for normal leaf size. Similarly, intermediate-sized leaves shown by *Del63CPR5*, demonstrate that compared to *Del37CPR5* and *Del63CPR5*, restriction on the entry of leaf size proteins was stricter in *Del98CPR5* lines, since *Del98CPR5* lines displayed *cpr5*-like leaves. Likewise, the bigger leaves but enhanced susceptibility shown by *metCPR5* plants, can be the result of the complete restriction of proteins needed for the accomplishment of resistance. Whereas, the enhanced entry of growth promoting proteins could be established in *metCPR5* transgenic lines. Alternatively, mutations in the putative start codons could have enabled metCPR5 recombinant proteins to acquire an ability to completely restrict resistance related proteins, whereas, it would enhance the entry of growth promoting proteins (Figure 10.1E). In summary, these results allowed me to conclude that deletions or mutations in the N-terminus of CPR5 could have resulted in the altered CPR5-mediated NPC selectivity and gating. Thus, depending on the severity of structural loss, some of the proteins required for CPR5 normal functions were restricted in some of the *CPR5* transgenic lines, whereas, the same proteins were allowed to enter in other lines. Hence, CPR5-mediated NPC selectivity and nuclear entry allowed CPR5 to exert pleiotropic effects.

10.3 CPR5 IDRs and Coiled-coil-domains may be involved in CPR5-mediated NPC gating

Coiled-coil domains or leucine zippers of proteins are commonly involved in multiple interactions (protein-protein, protein-DNA or protein-RNA) (Nguyen and Modis, 2013, Truebestein and Leonard, 2016) and protein dimerization (Joslyn et al., 1993, Lu et al., 2013). As shown in Chapter 3, the AtCPR5 N-terminus is predicted to contain four coiled-coil

domains, which may allow the AtCPR5 N-terminus to be implicated in the dimerization of AtCPR5 proteins (homodimer) or with putative interacting partners (heterodimer). The execution of *cpr5*-like phenotypes by *ccdCPR5* transgenic lines strongly suggests the existence of functional coiled-coil domains and their role in AtCPR5 homomeric interaction (Chapter 9). Intrinsically disordered regions (IDRs) or proteins (IDPs) act as signalling hubs (Oldfield and Dunker, 2014). These hubs enable carrier protein(s) to bind to multiple partners simultaneously (Kriwacki et al., 1996, Dunker et al., 1998, Hasty and Collins, 2001) by rendering more than one binding site (Dunker et al., 2005, Xie et al., 2007, Oldfield et al., 2008, Uversky and Dunker, 2010, Hsu et al., 2013). Alternatively, IDRs/IDPs can bind to a single binding partner by transiting from a disordered-to-ordered shape (Oldfield *et al.*, 2008; Hsu, 2013). As discussed in Chapter 4, CPR5 IDRs show strong resemblance to the classical already characterised IDRs, and contain the majority of the characteristics of typical IDRs. Keeping in view these results, CPR5 IDR may also facilitate the CPR5 protein to interact with multiple partners. As established by Gu et al, (2016) and discussed above in Section 10.3, the CPR5 N-terminus is required for CPR5 homo-dimerization. Thus, coiled-coil (CC) domains and IDRs could allow CPR5 to interact and select multiple proteins based on the structural motifs present in its N-terminus, which would subsequently affect CPR5-mediated NPC selectivity.

Taken altogether, the results presented in the current thesis show that the NLS-encoding cluster may not be required for AtCPR5 nuclear membrane localisation. However, these clusters are essential for some AtCPR5 functions. Additionally, the present thesis establishes that the AtCPR5 N-terminus participates or coordinates in a number of distinct pathways as part of its role as a selective barrier of NPC. Moreover, quantification and localisation studies in the future would help to test the degree of effect of mutations on nlsCPR5 recombinant proteins' localisation, and subsequently on protein translocation.


A. CPR5
Various nuclear cargoes
in NPC basket
Cytoplasm
Nuclear lemma
Nuclear lemma
CPR5
Nucleoplasm
B. cpr5
Cytoplasm
Nuclear lemma
Nuclear lemma
cpr5
Nucleoplasm
C. SynCPR5
Cytoplasm
Nuclear lemma
Nuclear lemma
SynCPR5
Nucleoplasm

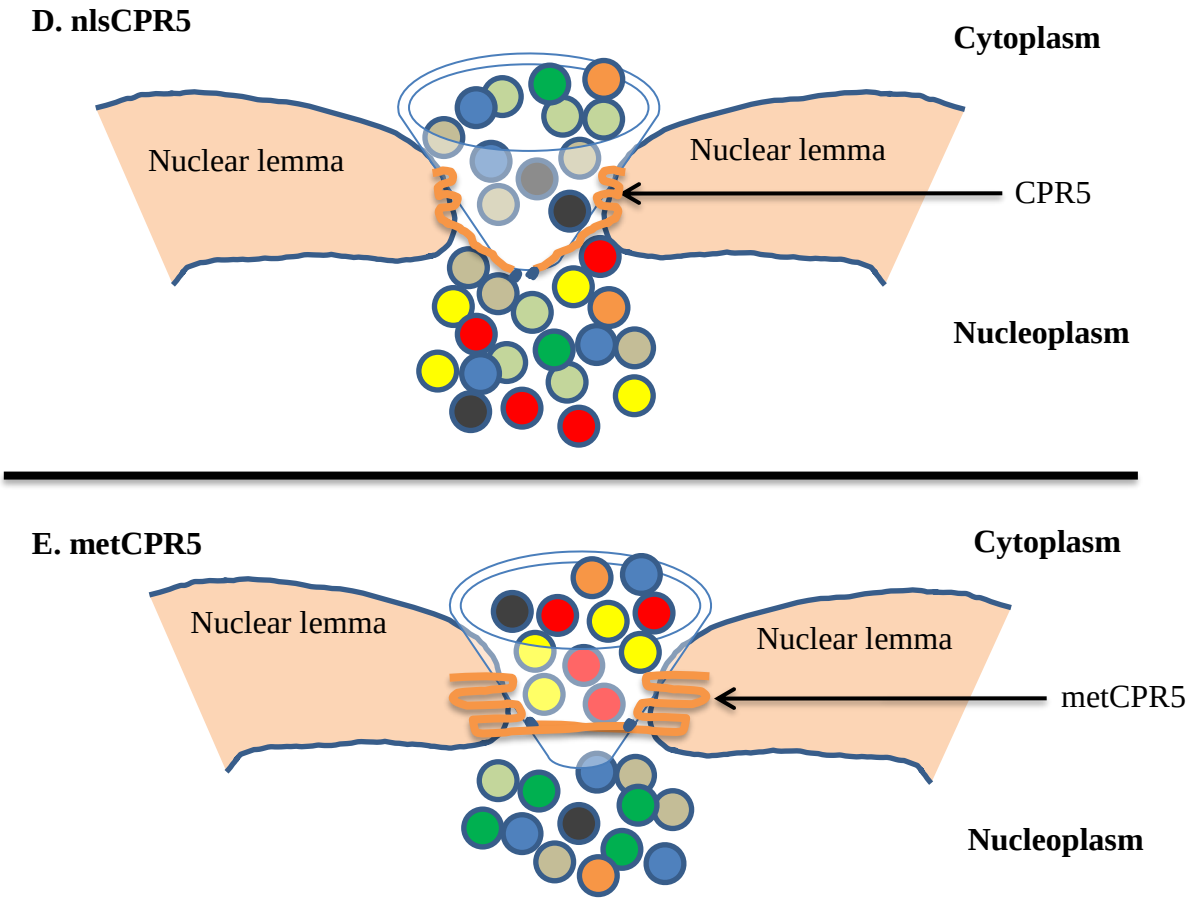


Figure 10.1 Proposed CPR5 schematic diagram in nuclear lemma and its working

This figure shows the proposed position and working of CPR5 in the nuclear membrane as one of the nucleoporins that controls the entry of an array of nuclear cargo. Under normal conditions, CPR5 keeps the balance of the entry of cargoes, especially growth and resistance proteins, by facilitating or modifying NPC gating through CPR5 N-terminal homomerization (**A and C**). A large number of deregulated nuclear cargoes enter into the nucleus as result of the removal of NPC restriction in *cpr5* (**B**). In CPR5 transgenic lines, some, or a varying number of, proteins could be restricted or allowed to enter into the nucleus in *nlsCPR5* lines depending on CPR5 expression levels (**D**). A majority of resistance related proteins could be restricted, whereas, growth promoting proteins are passed in *metCPR5* lines (**E**). The basket represents NPC and a blue spot on one end of CPR5, marks the N-terminus of CPR5 or CPR5 recombinant versions. This figure is constructed from Gu et al., (2016) results.

Chapter 11 APPENDICES

11.1 List of primers

Following is the list of primers which were used for the amplification of *CPR5* constructs to be used for structural studies. The lowercase letters represent restriction sites and stop codon.

11.1.1 Primers for the amplification of untagged *CPR5* constructs		
Construct	**Primer**	**Sequence**
A	Forward	ggaattccatATGGAAGCCCTCCTCCTC
	Reverse	ccgctcgagTCAAGCATAGTCAGACCCACCAT
B	Forward	catATGGAAGCCCTCCTCCTC
	Reverse	ggatccTTACTCTAGCGCCGCTTTCC
C	Forward	catATGGAAGCCCTCCTCCTC
	Reverse	ggatccTTAAGCTTTCGAAACATCCAACTCT
D	Forward	catATGGAAGCCCTCCTCCTC
	Reverse	ggatccTTAGCCTGAGTTAATCGATGAAACTT
E	Forward	catATGGAAGCCCTCCTCCTC
	Reverse	GGATCCTTAAACCCGGCAGATGAAGGTGT
F	Forward	ggaattccatATGGACAACGACGAAGCTTCT
	Reverse	ccgctcgagttaAGCTTTCGAAACATCCAACTCT
G	Forward	ggaattccatatgCATAGATTACGAAACCCTA
	Reverse	ccgctcgagttaAGCTTTCGAAACATCCAACTCT
H	Forward	ggaattccatatgCATAGATTACGAAACCCTA
	Reverse	ccgctcgagttaCTCTTCTTTTCTTGTATCTTCTAAT
11.1.2 List of primers for the amplification of *CPR5* constructs with N-ter His-tag		
A	Forward	cgggatccATGGAAGCCCTCCTCCTC
	Reverse	ccgctcgagTCAAGCATAGTCAGACCCACCAT
B	Forward	ggatccATGGAAGCCCTCCTCCTC
	Reverse	ggatccTTACTCTAGCGCCGCTTTCC
C	Forward	ggatccATGGAAGCCCTCCTCCTC
	Reverse	ggatccTTAAGCTTTCGAAACATCCAACTCT
D	Forward	ggatccATGGAAGCCCTCCTCCTC
	Reverse	ggatccTTAGCCTGAGTTAATCGATGAAACTT
E	Forward	ggatccATGGAAGCCCTCCTCCTC

	Reverse	ggatccTTAAACCCGGCAGATGAAGGTGT
F	Forward	cgggatccATGGACAACGACGAAGCTTCT
	Reverse	ccgctcgagttaAGCTTTCGAAACATCCAACTCT
G	Forward	cgggatccCATAGATTACGAAACCCTA
	Reverse	ccgctcgagttaAGCTTTCGAAACATCCAACTCT
H	Forward	cgggatccCATAGATTACGAAACCCTA
	Reverse	ccgctcgagttaCTCTTCTTTTCTTGTATCTTCTAAT

11.1.3 List of primers for the amplification of *CPR5* constructs with C-ter His-tag

A	Forward	ggaattccatATGGAAGCCCTCCTCCTC
	Reverse	ccgctcgagAGCATAGTCAGACCCACCATGAAG
B	Forward	catATGGAAGCCCTCCTCCTC
	Reverse	ctcgagCTCTAGCGCCGCTTTCCCC
C	Forward	catATGGAAGCCCTCCTCCTC
	Reverse	ctcgagAGCTTTCGAAACATCCAACTCT
D	Forward	catATGGAAGCCCTCCTCCTC
	Reverse	ctcgagGCCTGAGTTAATCGATGAAACTT
E	Forward	catatgGAAGCCCTCCTCCTC
	Reverse	ctcgagAACCCGGCAGATGAAGGTGT
F	Forward	ggaattccatATGGACAACGACGAAGCTTCT
	Reverse	ccgctcgagAGCTTTCGAAACATCCAACTCT
G	Forward	ggaattccatatgCATAGATTACGAAACCCTA
	Reverse	ccgctcgagAGCTTTCGAAACATCCAACTCT
H	Forward	ggaattccatatgCATAGATTACGAAACCCTA
	Reverse	ccgctcgagCTCTTCTTTTCTTGTATCTTCTAAT

11.1.4 List of primers used for real-time quantifications

Primer Name		Sequences
CPR5	Forward	GGTGATGGCAAACTCTGGCTAATAGTTCG
	Reverse	AAAGATGGCCTCCTCGTCTCAAGCA
CPR5-Opt	Forward	ACTTCCTCAATAAGCGCTCGAGT
	Reverse	ATTAATGGGCCCGAACATCAAT
cpr5-2	Forward	GGTGATGGCAAACTCTGGCTAATAGTTCG
	Reverse	CCCACCATGAAGGTGATACATAGCGAAG
PRI	Forward	ACACGTGCAATGGAGTTTGTGG
	Reverse	TTGGCACATCCGAGTCTCACTG
PR5	Forward	ATCACCCACAGCACAGAGACAC

	Reverse	AGCAATGCCGCTTGTGATGAAC
PDF1.2	Forward	CTTGTTCTCTTTGCTGCTTTCGAC
	Reverse	TTGGCTCCTTCAAGGTTAATGCAC
KRP2	Forward	TAGGAGATTATGGCGGCGGTTAGG
	Reverse	TTTCACCGTCGTCGTAACTC
CYCA2;3	Forward	TCTTGGGAGATCAGCTTCTACAGC
	Reverse	GGCATAGAGGCAGCACAGTAAAGG
ACT2	Forward	TCTTCCGCTCTTTCTTTCCAAGC
	Reverse	ACCATTGTCACACACGATTGGTTG
UBC9	Forward	TCACAATTTCCAAGGTGCTGC
	Reverse	TCATCTGGGTTTGGATCCGT

11.1.5 List of genotyping primers

Primer Name		Primer Sequences
CPR5	Forward	GGTGATGGCAAACTCTGGCTAATAGTTCG
	Reverse	AAAGATGGCCTCCTCGTCTCAAGCA
cpr5-2	Forward	GGTGATGGCAAACTCTGGCTAATAGTTCG
	Reverse	CCCACCATGAAGGTGATACATAGCGAAG
MF-53	Forward	TCAAATCACCAATCCGGCGA
	Reverse	GGCCCGAACATCAATCCGTA
metCPR5b	Forward	TACATAAAGATGAGACGCAAAGACA
	Reverse	GGCCCGAACATCAATCCGTA
metCPR5bc	Forward	CTTAAGGGAAAGAAGAAAGAGATTCAA
	Reverse	GGCCCGAACATCAATCCGTA
nlsCPR5A	Forward	GACGATGATGATGAAGCAACAA
	Reverse	GGCCCGAACATCAATCCGTA
nlsCPR5B	Forward	ATACGAATCCATCGAATTTGGAAC
	Reverse	GGCCCGAACATCAATCCGTA
nlsCPR5AB	Forward	GGATACGAATCCATCGAATTTGGAACAAC
	Reverse	GGCCCGAACATCAATCCGTA
nlsCPR5BC	Forward	CAACAACAGCTTAAGGGACAGCAGCAG
	Reverse	GGCCCGAACATCAATCCGTA
nlsCPR5ABC	Forward	GAATCCATCGAATTTGGAACAACAAC
	Reverse	GGCCCGAACATCAATCCGTA
Del37CPR5	Forward	ATGGAATAGAAGCTTAAGGGAAAG
	Reverse	GGCCCGAACATCAATCCGTA
Del63CPR5	Forward	ATGGAATAGAAGCTTGTCGACCTCTA
	Reverse	GGCCCGAACATCAATCCGTA
Delck1CPR5	Forward	TTCGTCCTATTGTTCTACATCGTC

	Reverse	GGCCCGAACATCAATCCGTA
Gln452CPR5	Forward	TTCCTGACACTCCACAGAATATCG
	Reverse	GGCCCGAACATCAATCCGTA
Gln456CPR5	Forward	AACGCAAAGTTTTCCTGACACCG
	Reverse	GGCCCGAACATCAATCCGTA

11.2 List of accession number of CPR5-like proteins

Name	Accession #		Name	Accession #
A. thalina	gi|194306663		*Medicago truncatula*	gi|657395169
A. lyrata	XP_002866642.1		*Musa acuminata*	gi|695024986
Amborella trichopoda	gi|586710323		*Nelumbo nucifera*	gi|720092824
Beta vulgaris	gi|731339884		*Nicotiana sylvestris*	gi|698479954
Brachypodium distachyon	gi|357143745		*Nicotiana tomentosiformis*	gi|697168875
Brassica napus	gi|674918646		*Oryza brachyantha*	gi|573920747
Brassica rapa	XP_009150570.1		*Oryza glaberrima*	gi|532130969
Camelina sativa	gi|727541991		*Oryza sativa Indica*	gi|218189614
Capsella rubella	gi|565432272		*Phaseolus vulgaris*	gi|593328354
Cicer arietinum	gi|502140031		*Phoenix dactylifera*	gi|672110551
Citrus clementina	gi|567870861		*Populus euphratica*	gi|743823318
Coffea canephora	gi|661883576		*Populus euphratica*	gi|743838012
Cucumis melo	gi|659080230		*Populus trichocarpa*	gi|566170215
Cucumis sativus	gi|778719342		*Prunus mume*	gi|645280405
Elaeis guineensis	gi|743804951		*Ricinus communis*	gi|255575531
Eucalyptus grandis	gi|702457680		*Sesamum indicum*	gi|747105235
Eutrema salsugineum	gi|567136927		*Setaria italica*	gi|514718676
Fragaria vesca	gi|764596945		*Solanum lycopersicum*	gi|460383456
Glycine max	gi|571460555		*Solanum tuberosum*	gi|565376439
Glycine soja	gi|734436958		*Sorghum bicolor*	gi|242066666
Gossypium arboreum	gi|728836204		*Tarenaya hassleriana*	gi|729361963
Gossypium hirsutum	XP_016740714.1		*Theobroma cacao*	gi|590707806
Gossypium raimondii	gi|763750924		*Vitis vinifera*	gi|225438581
Hordeum vulgare	gi|326533350		*Zea mays*	gi|670394607
Jatropha curcas	gi|802615483		*Zea mays*	gi|413951523

11.3 Students' t-test results

11.3.1 Expression levels: Statistical significance at p < 0.05 (Students' t-test)

	cpr5	*SynCPR5L1*	*SynCPR5L2*	*SynCPR5L3*	*SynCPR5L4*	*SynCPR5L5*	*SynCPR5L6*
Col0	**0.00**	0.22	**0.00**	**0.01**	0.28	**0.02**	0.15
cpr5		**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
SynCPR5L1			**0.00**	**0.01**	**0.04**	**0.00**	0.50
SynCPR5L2				**0.01**	**0.00**	**0.00**	**0.01**
SynCPR5L3					**0.00**	**0.00**	0.07
SynCPR5L4						**0.04**	**0.04**
SynCPR5L5							**0.01**

11.3.2 Trichomes with three appendages: Statistical significance at p < 0.05 (Students' t-test)

	cpr5	*SynCPR5L1*	*SynCPR5L2*	*SynCPR5L3*	*SynCPR5L4*	*SynCPR5L5*	*SynCPR5L6*
Col0	**0.00**	**0.02**	**0.03**	**0.00**	**0.00**	**0.00**	**0.00**
cpr5		**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
SynCPR5L1			0.81	**0.00**	**0.02**	**0.00**	**0.00**
SynCPR5L2				**0.00**	**0.02**	**0.00**	**0.00**
SynCPR5L3					**0.00**	0.77	**0.03**
SynCPR5L4						**0.00**	**0.00**
SynCPR5L5							**0.03**

11.3.3 Area of third leaf pair: Statistical significance at p < 0.05 (Students' t-test)

	cpr5	*SynCPR5L1*	*SynCPR5L2*	*SynCPR5L3*	*SynCPR5L4*	*SynCPR5L5*	*SynCPR5L6*	*SynCPR5L7*
Col0	**0.00**	0.06	0.86	0.47	**0.00**	**0.00**	**0.00**	**0.00**
cpr5		**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
SynCPR5L1			0.13	0.07	**0.00**	**0.00**	**0.00**	**0.00**
SynCPR5L2				0.66	**0.01**	**0.02**	**0.05**	**0.04**
SynCPR5L3					0.00	0.00	0.00	0.00
SynCPR5L4						0.01	0.00	0.00
SynCPR5L5							0.01	0.02
SynCPR5L6								0.29

11.3.4 Area of fourth leaf pair: Statistical significance at p < 0.05 (Students' t-test)

	cpr5	*SynCPR5L1*	*SynCPR5L2*	*SynCPR5L3*	*SynCPR5L4*	*SynCPR5L5*	*SynCPR5L6*	*SynCPR5L7*
Col0	**0.00**	0.70	0.80	0.80	**0.02**	**0.02**	0.12	**0.04**
cpr5		**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
SynCPR5L1			0.69	0.73	**0.00**	**0.00**	**0.01**	**0.00**
SynCPR5L2				1.00	**0.00**	**0.00**	**0.01**	**0.00**
SynCPR5L3					**0.00**	**0.00**	**0.01**	**0.00**
SynCPR5L4						0.64	0.01	0.08
SynCPR5L5							0.02	0.13
SynCPR5L6								0.12

11.3.5 *PR1* transcript abundance: Statistical significance at p < 0.05 (Students' t-test)

	cpr5	*SynCPR5L1*	*SynCPR5L2*	*SynCPR5L3*	*SynCPR5L4*	*SynCPR5L5*	*SynCPR5L6*
Col0	**0.00**	**0.01**	**0.02**	**0.00**	0.00	0.00	0.00
cpr5		**0.00**	**0.00**	**0.01**	**0.01**	**0.01**	**0.02**
SynCPR5L1			0.14	**0.00**	**0.00**	**0.00**	0.00
SynCPR5L2				**0.00**	**0.00**	**0.00**	**0.01**
SynCPR5L3					**0.05**	**0.03**	0.45
SynCPR5L4						**0.02**	0.05
SynCPR5L5							0.46

11.3.6 Expression levels: Level of statistical significance at p < 0.05 (Students' t-test)

	cpr5	*nlsCPR5ABL1*	*nlsCPR5ABL2*	*nlsCPR5BCL1*	*nlsCPR5BCL2*	*nlsCPR5AL1*	*nlsCPR5AL2*	*nlsCPR5BL1*	*nlsCPR5BL2*	*DelCPR563L1*	*DelCPR563L2*
Col0	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.03**	**0.02**	0.63	0.51	**0.00**	**0.00**
cpr5		**0.00**	**0.01**	0.06	0.06	**0.00**	**0.02**	**0.00**	**0.00**	0.26	0.31
nlsCPR5ABL1			**0.00**	**0.00**	**0.00**	0.17	0.54	**0.02**	**0.01**	**0.00**	**0.01**
nlsCPR5ABL2				**0.01**	**0.02**	**0.00**	0.10	**0.00**	**0.00**	**0.02**	0.45
nlsCPR5BCL1					0.47	**0.00**	**0.03**	**0.00**	**0.00**	0.58	0.66
nlsCPR5BCL2						**0.00**	**0.03**	**0.00**	**0.00**	0.86	0.52
nlsCPR5AL1							0.20	**0.04**	**0.02**	**0.00**	0.01
nlsCPR5AL2								**0.02**	**0.01**	**0.03**	0.09
nlsCPR5BL1									0.89	**0.00**	0.00
nlsCPR5BL2										**0.00**	0.00
DelCPR563L1											0.79

REFERENCES

ABELES, F. B., MORGAN, P. W. & SALTVEIT JR, M. E. 2012. *Ethylene in plant biology*, Academic press.

ACHARD, P., GUSTI, A., CHEMINANT, S., ALIOUA, M., DHONDT, S., COPPENS, F., BEEMSTER, G. T. S. & GENSCHIK, P. 2009. Gibberellin Signaling Controls Cell Proliferation Rate in Arabidopsis. *Current Biology,* 19**,** 1188-1193.

ACHARD, P., LIAO, L., JIANG, C., DESNOS, T., BARTLETT, J., FU, X. & HARBERD, N. P. 2007. DELLAs Contribute to Plant Photomorphogenesis. *Plant Physiology,* 143**,** 1163.

AKI, T., KONISHI, M., KIKUCHI, T., FUJIMORI, T., YONEYAMA, T. & YANAGISAWA, S. 2007. Distinct modulations of the hexokinase1-mediated glucose response and hexokinase1-independent processes by HYS1/CPR5 in Arabidopsis. *J Exp Bot,* 58**,** 3239-48.

ALBRECHT, C., BOUTROT, F., SEGONZAC, C., SCHWESSINGER, B., GIMENEZ-IBANEZ, S., CHINCHILLA, D., RATHJEN, J. P., DE VRIES, S. C. & ZIPFEL, C. 2012. Brassinosteroids inhibit pathogen-associated molecular pattern–triggered immune signaling independent of the receptor kinase BAK1. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 303-308.

ALVES, J. D. C., SOUZA, A. P. D., PÔRTO, M. L., ARRUDA, J. A. D., TOMPSON JÚNIOR, U. A., SILVA, G. B. D., ARAÚJO, R. D. C. & SANTOS, D. 2008. Absorção e distribuição de chumbo em plantas devetiver, jureminha e algaroba. *Revista Brasileira de Ciência do Solo,* 32**,** 1329-1336.

ANAND, A., UPPALAPATI, S. R., RYU, C. M., ALLEN, S. N., KANG, L., TANG, Y. & MYSORE, K. S. 2008. Salicylic acid and systemic acquired resistance play a role in attenuating crown gall disease caused by Agrobacterium tumefaciens. *Plant Physiol,* 146**,** 703-15.

ANDERSEN, S. O. 2011. Are structural proteins in insect cuticles dominated by intrinsically disordered regions? *Insect Biochem Mol Biol,* 41**,** 620-7.

ANDERSON, L. K. 2006. *CPR5 encodes a novel protein required for endoreduplication and cell expansion in plants*, ProQuest.

ANTIKAINEN, N. M. & MARTIN, S. F. 2005. Altering protein specificity: techniques and applications. *Bioorg Med Chem,* 13**,** 2701-16.

ARMENGAUD, P., BREITLING, R. & AMTMANN, A. 2004. The Potassium-Dependent Transcriptome of Arabidopsis Reveals a Prominent Role of Jasmonic Acid in Nutrient Signaling. *Plant Physiology,* 136**,** 2556-2576.

ARMENGAUD, P., SULPICE, R., MILLER, A. J., STITT, M., AMTMANN, A. & GIBON, Y. 2009. Multilevel Analysis of Primary Metabolism Provides New Insights into the

Role of Potassium Nutrition for Glycolysis and Nitrogen Assimilation in Arabidopsis Roots. *Plant Physiology,* 150**,** 772-785.

ARMSTRONG, W. 1980. Aeration in Higher Plants. *Advances in Botanical Research,* 7**,** 225-332.

ARNAU, J., LAURITZEN, C., PETERSEN, G. E. & PEDERSEN, J. 2006. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr Purif,* 48**,** 1-13.

ARTS, G. J., FORNEROD, M. & MATTAJ, I. W. 1998. Identification of a nuclear export receptor for tRNA. *Curr Biol,* 8**,** 305-14.

ARVIDSSON, S., KWASNIEWSKI, M., RIAÑO-PACHÓN, D. M. & MUELLER-ROEBER, B. 2008. QuantPrime – a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics,* 9**,** 465-465.

ASANO, K. 2014. Why is start codon selection so precise in eukaryotes? *Translation,* 2**,** e28387.

ASHIKARI, M., SAKAKIBARA, H., LIN, S., YAMAMOTO, T., TAKASHI, T., NISHIMURA, A., ANGELES, E. R., QIAN, Q., KITANO, H. & MATSUOKA, M. 2005. Cytokinin Oxidase Regulates Rice Grain Production. *Science,* 309**,** 741.

ASHLEY, M. K., GRANT, M. & GRABOV, A. 2006. Plant responses to potassium deficiencies: a role for potassium transport proteins. *Journal of Experimental Botany,* 57**,** 425-436.

BAJGUZ, A. & HAYAT, S. 2009. Effects of brassinosteroids on the plant responses to environmental stresses. *Plant Physiology and Biochemistry,* 47**,** 1-8.

BAKER, M. 2010. Making membrane proteins for structures: a trillion tiny tweaks. *Nat Meth,* 7**,** 429-434.

BALAZADEH, S., KWASNIEWSKI, M., CALDANA, C., MEHRNIA, M., ZANOR, M. I., XUE, G.-P. & MUELLER-ROEBER, B. 2011. ORS1, an H(2)O(2)-Responsive NAC Transcription Factor, Controls Senescence in Arabidopsis thaliana. *Molecular Plant,* 4**,** 346-360.

BALAZADEH, S., SIDDIQUI, H., ALLU, A. D., MATALLANA-RAMIREZ, L. P., CALDANA, C., MEHRNIA, M., ZANOR, M.-I., KÖHLER, B. & MUELLER-ROEBER, B. 2010a. A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. *The Plant Journal,* 62**,** 250-264.

BALAZADEH, S., WU, A. & MUELLER-ROEBER, B. 2010b. Salt-triggered expression of the ANAC092-dependent senescence regulon in Arabidopsis thaliana. *Plant Signaling & Behavior,* 5**,** 733-735.

BALDWIN, I. T. 1998. Jasmonate-induced responses are costly but benefit plants under attack in native populations. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 8113-8118.

BAO, Z. & HUA, J. 2014. Interaction of CPR5 with Cell Cycle Regulators UVI4 and OSD1 in Arabidopsis. *PLoS ONE,* 9**,** e100347.

BARDWELL, J. C. A. & JAKOB, U. 2012. Conditional disorder in chaperone action. *Trends in biochemical sciences,* 37**,** 517-525.

BARTESAGHI, A. & SUBRAMANIAM, S. 2009. Membrane protein structure determination using cryo-electron tomography and 3D image averaging. *Curr Opin Struct Biol,* 19**,** 402-7.

BECHTOLD, N. & PELLETIER, G. 1998. In planta Agrobacterium-mediated transformation of adult Arabidopsis thaliana plants by vacuum infiltration. *Methods Mol Biol,* 82**,** 259-66.

BELKHADIR, Y., JAILLAIS, Y., EPPLE, P., BALSEMÃO-PIRES, E., DANGL, J. L. & CHORY, J. 2012. Brassinosteroids modulate the efficiency of plant immune responses to microbe-associated molecular patterns. *Proceedings of the National Academy of Sciences,* 109**,** 297-302.

BELKHADIR, Y., NIMCHUK, Z., HUBERT, D. A., MACKEY, D. & DANGL, J. L. 2004. Arabidopsis RIN4 Negatively Regulates Disease Resistance Mediated by RPS2 and RPM1 Downstream or Independent of the NDR1 Signal Modulator and Is Not Required for the Virulence Functions of Bacterial Type III Effectors AvrRpt2 or AvrRpm1. *The Plant Cell,* 16**,** 2822.

BENT, A. F. 2000. Arabidopsis in Planta Transformation. Uses, Mechanisms, and Prospects for Transformation of Other Species. *Plant Physiology,* 124**,** 1540-1547.

BERNSEL, A., VIKLUND, H., HENNERDAL, A. & ELOFSSON, A. 2009. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Research,* 37**,** W465-W468.

BERROCAL-LOBO, M., MOLINA, A. & SOLANO, R. 2002. Constitutive expression of ETHYLENE-RESPONSE-FACTOR1 in Arabidopsis confers resistance to several necrotrophic fungi. *The Plant Journal,* 29**,** 23-32.

BIRCH, R. G. 1997. PLANT TRANSFORMATION: Problems and Strategies for Practical Application. *Annu Rev Plant Physiol Plant Mol Biol,* 48**,** 297-326.

BLATTNER, C., HAY, T., MEEK, D. W. & LANE, D. P. 2002. Hypophosphorylation of Mdm2 augments p53 stability. *Mol Cell Biol,* 22**,** 6170-82.

BOCH, J., VERBSKY, M. L., ROBERTSON, T. L., LARKIN, J. C. & KUNKEL, B. N. 1998. Analysis of resistance gene-mediated defense responses in Arabidopsis thaliana plants carrying a mutation in CPR5. *Molecular Plant-Microbe Interactions,* 11**,** 1196-1206.

BONKE, M., THITAMADEE, S., MAHONEN, A. P., HAUSER, M.-T. & HELARIUTTA, Y. 2003. APL regulates vascular tissue identity in Arabidopsis. *Nature,* 426**,** 181-186.

BONNER, W. M. 1978. *Protein migration and accumulation in nuclei.* , Academic Press, New York.

BORGHI, M., RUS, A. & SALT, D. E. 2011. Loss-of-Function of Constitutive Expresser of Pathogenesis Related Genes5 Affects Potassium Homeostasis in Arabidopsis thaliana. *PLoS ONE,* 6**,** e26360.

BOWLING, S. A., CLARKE, J. D., LIU, Y., KLESSIG, D. F. & DONG, X. 1997. The cpr5 mutant of Arabidopsis expresses both NPR1-dependent and NPR1-independent resistance. *Plant Cell,* 9**,** 1573-84.

BOWLING, S. A., GUO, A., CAO, H., GORDON, A. S., KLESSIG, D. F. & DONG, X. 1994. A mutation in Arabidopsis that leads to constitutive expression of systemic acquired resistance. *The Plant Cell,* 6**,** 1845-1857.

BRACKEN, C. 2001. NMR spin relaxation methods for characterization of disorder and folding in proteins. *J Mol Graph Model,* 19**,** 3-12.
BREEUWER, M. & GOLDFARB, D. S. 1990. Facilitated nuclear transport of histone H1 and other small nucleophilic proteins. *Cell,* 60**,** 999-1008.
BREUER, C., ISHIDA, T. & SUGIMOTO, K. 2010. Developmental control of endocycles and cell growth in plants. *Current Opinion in Plant Biology,* 13**,** 654-660.
BRININSTOOL, G., KASILI, R., SIMMONS, L. A., KIRIK, V., HULSKAMP, M. & LARKIN, J. C. 2008. Constitutive Expressor Of Pathogenesis-related Genes5 affects cell wall biogenesis and trichome development. *BMC Plant Biol,* 8**,** 58.
BROWN, C. J., JOHNSON, A. K. & DAUGHDRILL, G. W. 2010. Comparing models of evolution for ordered and disordered proteins. *Mol Biol Evol,* 27**,** 609-21.
BRUCH, T., GRAALFS, H., JACOB, L. & FRECH, C. 2009. Influence of surface modification on protein retention in ion-exchange chromatography. Evaluation using different retention models. *J Chromatogr A,* 1216**,** 919-26.
BRUGGEMAN, Q., RAYNAUD, C., BENHAMED, M. & DELARUE, M. 2015. To die or not to die? Lessons from lesion mimic mutants. *Frontiers in Plant Science,* 6**,** 24.
BUCHANAN-WOLLASTON, V. 1997. The molecular biology of leaf senescence. *Journal of Experimental Botany,* 48**,** 181-199.
BÜCHERL, C. A., VAN ESSE, G. W., KRUIS, A., LUCHTENBERG, J., WESTPHAL, A. H., AKER, J., VAN HOEK, A., ALBRECHT, C., BORST, J. W. & DE VRIES, S. C. 2013. Visualization of BRI1 and BAK1(SERK3) Membrane Receptor Heterooligomers during Brassinosteroid Signaling. *Plant Physiology,* 162**,** 1911.
BULJAN, M., CHALANCON, G., DUNKER, A. K., BATEMAN, A., BALAJI, S., FUXREITER, M. & BABU, M. M. 2013. Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr Opin Struct Biol,* 23**,** 443-50.
BULJAN, M., CHALANCON, G., EUSTERMANN, S., WAGNER, G. P., FUXREITER, M., BATEMAN, A. & BABU, M. M. 2012. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol Cell,* 46**,** 871-83.
BURKHARD, P., STETEFELD, J. & STRELKOV, S. V. 2001. Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol,* 11**,** 82-8.
BUSKE, P. J., MITTAL, A., PAPPU, R. V. & LEVIN, P. A. 2015. An intrinsically disordered linker plays a critical role in bacterial cell division. *Semin Cell Dev Biol,* 37**,** 3-10.
BUSTOS, VICTOR H., MARIN, O., MEGGIO, F., CESARO, L., ALLENDE, CATHERINE C., ALLENDE, JORGE E. & PINNA, LORENZO A. 2005. Generation of protein kinase Ck1α mutants which discriminate between canonical and non-canonical substrates. *Biochemical Journal,* 391**,** 417-424.
CAFFREY, M. 2003. Membrane protein crystallization. *J Struct Biol,* 142**,** 108-32.
CAMPEN, A., WILLIAMS, R. M., BROWN, C. J., MENG, J., UVERSKY, V. N. & DUNKER, A. K. 2008. TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett,* 15**,** 956-63.
CAO, H., BOWLING, S. A., GORDON, A. S. & DONG, X. 1994. Characterization of an Arabidopsis Mutant That Is Nonresponsive to Inducers of Systemic Acquired Resistance. *The Plant Cell,* 6**,** 1583.

CARPENTER, E. P., BEIS, K., CAMERON, A. D. & IWATA, S. 2008a. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol,* 18**,** 581-6.

CARPENTER, T., BOND, P. J., KHALID, S. & SANSOM, M. S. 2008b. Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel. *Biophys J,* 95**,** 3790-801.

CARTA, G. & JUNGBAUER, A. 2010. Downstream Processing of Biotechnology Products. *Protein Chromatography*. Wiley-VCH Verlag GmbH & Co. KGaA.

CHABREGAS, S. M., LUCHE, D. D., VAN SLUYS, M. A., MENCK, C. F. & SILVA-FILHO, M. C. 2003. Differential usage of two in-frame translational start codons regulates subcellular localization of Arabidopsis thaliana THI1. *J Cell Sci,* 116**,** 285-91.

CHAE, W. J., HENEGARIU, O., LEE, S. K. & BOTHWELL, A. L. 2006. The mutant leucine-zipper domain impairs both dimerization and suppressive function of Foxp3 in T cells. *Proc Natl Acad Sci U S A,* 103**,** 9631-6.

CHANDRAN, D., RICKERT, J., HUANG, Y., STEINWAND, MICHAEL A., MARR, SHARON K. & WILDERMUTH, MARY C. 2014. Atypical E2F Transcriptional Repressor DEL1 Acts at the Intersection of Plant Growth and Immunity by Controlling the Hormone Salicylic Acid. *Cell Host & Microbe,* 15**,** 506-513.

CHATTERJEE, S. & STOCHAJ, U. 1998. Diffusion of proteins across the nuclear envelope of HeLa cells. *Biotechniques,* 24**,** 668-74.

CHENG, H. Y., SCHIAVONE, A. P. & SMITHGALL, T. E. 2001. A point mutation in the N-terminal coiled-coil domain releases c-Fes tyrosine kinase activity and survival signaling in myeloid leukemia cells. *Mol Cell Biol,* 21**,** 6170-80.

CHEREZOV, V. & CAFFREY, M. 2007. Membrane protein crystallization in lipidic mesophases. A mechanism study using X-ray microdiffraction. *Faraday Discuss,* 136**,** 195-212; discussion 213-29.

CHEVALIER, C., NAFATI, M., MATHIEU-RIVET, E., BOURDON, M., FRANGNE, N., CHENICLET, C., RENAUDIN, J.-P., GÉVAUDANT, F. & HERNOULD, M. 2011. Elucidating the functional role of endoreduplication in tomato fruit development. *Annals of Botany,* 107**,** 1159-1169.

CHEVALIER, F., PERAZZA, D., LAPORTE, F., LE HENANFF, G., HORNITSCHEK, P., BONNEVILLE, J. M., HERZOG, M. & VACHON, G. 2008. GeBP and GeBP-like proteins are noncanonical leucine-zipper transcription factors that regulate cytokinin response in Arabidopsis. *Plant Physiol,* 146**,** 1142-54.

CHINCHILLA, D., ZIPFEL, C., ROBATZEK, S., KEMMERLING, B., NÜRNBERGER, T., JONES, J. D. G., FELIX, G. & BOLLER, T. 2007. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature,* 448**,** 497-500.

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.

CHOI, J., CHOI, D., LEE, S., RYU, C.-M. & HWANG, I. 2011. Cytokinins and plant immunity: old foes or new friends? *Trends in Plant Science,* 16**,** 388-394.

CHOI, J., HUH, S. U., KOJIMA, M., SAKAKIBARA, H., PAEK, K.-H. & HWANG, I. 2010. The Cytokinin-Activated Transcription Factor ARR2 Promotes Plant Immunity via

TGA3/NPR1-Dependent Salicylic Acid Signaling in Arabidopsis. *Developmental Cell,* 19**,** 284-295.
CHOI, J. & HWANG, I. 2007. Cytokinin: perception, signal transduction, and role in plant growth and development. *Journal of Plant Biology,* 50**,** 98-108.
CHOTHIA, C. 1976. The nature of the accessible and buried surfaces in proteins. *J Mol Biol,* 105**,** 1-12.
CHRISTENDAT, D., YEE, A., DHARAMSI, A., KLUGER, Y., SAVCHENKO, A., CORT, J. R., BOOTH, V., MACKERETH, C. D., SARIDAKIS, V., EKIEL, I., KOZLOV, G., MAXWELL, K. L., WU, N., MCINTOSH, L. P., GEHRING, K., KENNEDY, M. A., DAVIDSON, A. R., PAI, E. F., GERSTEIN, M., EDWARDS, A. M. & ARROWSMITH, C. H. 2000. Structural proteomics of an archaeon. *Nat Struct Biol,* 7**,** 903-9.
CHRISTENSEN, A. C., LYZNIK, A., MOHAMMED, S., ELOWSKY, C. G., ELO, A., YULE, R. & MACKENZIE, S. A. 2005. Dual-domain, dual-targeting organellar protein presequences in Arabidopsis can use non-AUG start codons. *Plant Cell,* 17**,** 2805-16.
CHRISTOFFERSON, D. E. & YUAN, J. 2010. Necroptosis as an alternative form of programmed cell death. *Current opinion in cell biology,* 22**,** 263-268.
CIANI, B., BJELIC, S., HONNAPPA, S., JAWHARI, H., JAUSSI, R., PAYAPILLY, A., JOWITT, T., STEINMETZ, M. O. & KAMMERER, R. A. 2010. Molecular basis of coiled-coil oligomerization-state specificity. *Proc Natl Acad Sci U S A,* 107**,** 19850-5.
CITOVSKY, V., KOZLOVSKY, S. V., LACROIX, B., ZALTSMAN, A., DAFNY-YELIN, M., VYAS, S., TOVKACH, A. & TZFIRA, T. 2007. Biological systems of the host cell involved in Agrobacterium infection. *Cell Microbiol,* 9**,** 9-20.
CLAEYS, H., DE BODT, S. & INZÉ, D. 2014. Gibberellins and DELLAs: central nodes in growth regulatory networks. *Trends in Plant Science,* 19**,** 231-239.
CLARKE, J. D., AARTS, N., FEYS, B. J., DONG, X. & PARKER, J. E. 2001. Constitutive disease resistance requires EDS1 in the Arabidopsis mutants cpr1 and cpr6 and is partially EDS1-dependent in cpr5. *Plant J,* 26**,** 409-20.
CLARKE, J. D., VOLKO, S. M., LEDFORD, H., AUSUBEL, F. M. & DONG, X. 2000. Roles of salicylic acid, jasmonic acid, and ethylene in cpr-induced resistance in arabidopsis. *Plant Cell,* 12**,** 2175-90.
CLOUGH, S. J. & BENT, A. F. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J,* 16**,** 735-43.
CLOUSE, S. D. 2011. Brassinosteroid Signal Transduction: From Receptor Kinase Activation to Transcriptional Networks Regulating Plant Development. *The Plant Cell,* 23**,** 1219.
COHEN, J. D. & MEUDT, W. J. 1983. Investigations on the Mechanism of the Brassinosteroid Response. *Plant Physiology,* 72**,** 691.
COLEY, P. D., BRYANT, J. P. & CHAPIN, F. S. 1985. Resource Availability and Plant Antiherbivore Defense. *Science,* 230**,** 895.
COLLAZO CORDERO, C., CHACON, O. & BORRÁS, O. 2006. *Programmed cell death in plants resembles apoptosis of animals*.
CRICK, F. 1970. Central Dogma of Molecular Biology. *Nature,* 227**,** 561-563.

CRICK, S. L., RUFF, K. M., GARAI, K., FRIEDEN, C. & PAPPU, R. V. 2013. Unmasking the roles of N- and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *Proc Natl Acad Sci U S A,* 110**,** 20075-80.

CRIVAT, G. & TARASKA, J. W. 2012. Imaging proteins inside cells with fluorescent tags. *Trends Biotechnol,* 30**,** 8-16.

CUMMINS, P. M., DOWLING, O. & O'CONNOR, B. F. 2011. Ion-exchange chromatography: basic principles and application to the partial purification of soluble mammalian prolyl oligopeptidase. *Methods Mol Biol,* 681**,** 215-28.

DANG, C. V. & LEE, W. M. 1988. Identification of the human c-myc protein nuclear translocation signal. *Mol Cell Biol,* 8**,** 4048-54.

DANGL, J. L., DIETRICH, R. A. & RICHBERG, M. H. 1996. Death Don't Have No Mercy: Cell Death Programs in Plant-Microbe Interactions. *The Plant Cell,* 8**,** 1793-1807.

DAUGHDRILL, G. W., NARAYANASWAMI, P., GILMORE, S. H., BELCZYK, A. & BROWN, C. J. 2007. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *J Mol Evol,* 65**,** 277-88.

DE GARA, L., DE PINTO, M. C. & TOMMASI, F. 2003. The antioxidant systems vis-à-vis reactive oxygen species during plant–pathogen interaction. *Plant Physiology and Biochemistry,* 41**,** 863-870.

DE LUCAS, M., DAVIERE, J.-M., RODRIGUEZ-FALCON, M., PONTIN, M., IGLESIAS-PEDRAZ, J. M., LORRAIN, S., FANKHAUSER, C., BLAZQUEZ, M. A., TITARENKO, E. & PRAT, S. 2008. A molecular framework for light and gibberellin control of cell elongation. *Nature,* 451**,** 480-484.

DE VEYLDER, L., LARKIN, J. C. & SCHNITTGER, A. 2011. Molecular control and function of endoreplication in development and physiology. *Trends in Plant Science,* 16**,** 624-634.

DEGTEREV, A. & YUAN, J. 2008. Expansion and evolution of cell death programmes. *Nat Rev Mol Cell Biol,* 9**,** 378-390.

DEKKERS, B. J. W. & BENTSINK, L. 2015. Regulation of seed dormancy by abscisic acid and DELAY OF GERMINATION 1. *Seed Science Research,* 25**,** 82-98.

DEKKERS, B. J. W., SCHUURMANS, J. A. M. J. & SMEEKENS, S. C. M. 2008. Interaction between sugar and abscisic acid signalling during early seedling development in Arabidopsis. *Plant Molecular Biology,* 67**,** 151-167.

DEL CARMEN RODRÍGUEZ-GACIO, M., MATILLA-VÁZQUEZ, M. A. & MATILLA, A. J. 2009. Seed dormancy and ABA signaling: The breakthrough goes on. *Plant Signaling & Behavior,* 4**,** 1035-1048.

DELANEY, T. P., FRIEDRICH, L. & RYALS, J. A. 1995. Arabidopsis signal transduction mutant defective in chemically and biologically induced disease resistance. *Proceedings of the National Academy of Sciences of the United States of America,* 92**,** 6602-6606.

DELLAPORTA, S. L. & CALDERON-URREA, A. 1994. The sex determination process in maize. *Science,* 266**,** 1501-5.

DENANCÉ, N., SÁNCHEZ-VALLET, A., GOFFNER, D. & MOLINA, A. 2013. Disease resistance or growth: the role of plant hormones in balancing immune responses and fitness costs. *Frontiers in Plant Science,* 4.

DESFEUX, C., CLOUGH, S. J. & BENT, A. F. 2000. Female reproductive tissues are the primary target of Agrobacterium-mediated transformation by the Arabidopsis floral-dip method. *Plant Physiol,* 123**,** 895-904.

DIETRICH, R. A., DELANEY, T. P., UKNES, S. J., WARD, E. R., RYALS, J. A. & DANGL, J. L. 1994. Arabidopsis mutants simulating disease resistance response. *Cell,* 77**,** 565-577.

DILL, K. A. 1990. Dominant forces in protein folding. *Biochemistry,* 29**,** 7133-55.

DINGWALL, C. & LASKEY, R. A. 1991. Nuclear targeting sequences--a consensus? *Trends Biochem Sci,* 16**,** 478-81.

DOSZTÁNYI, Z., CSIZMOK, V., TOMPA, P. & SIMON, I. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics,* 21**,** 3433-3434.

DRENKARD, E., RICHTER, B. G., ROZEN, S., STUTIUS, L. M., ANGELL, N. A., MINDRINOS, M., CHO, R. J., OEFNER, P. J., DAVIS, R. W. & AUSUBEL, F. M. 2000. A Simple Procedure for the Analysis of Single Nucleotide Polymorphisms Facilitates Map-Based Cloning in Arabidopsis. *Plant Physiology,* 124**,** 1483.

DROZDETSKIY, A., COLE, C., PROCTER, J. & BARTON, G. J. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res,* 43**,** W389-94.

DUNKER, A. K., BABU, M. M., BARBAR, E., BLACKLEDGE, M., BONDOS, S. E., DOSZTÁNYI, Z., DYSON, H. J., FORMAN-KAY, J., FUXREITER, M., GSPONER, J., HAN, K.-H., JONES, D. T., LONGHI, S., METALLO, S. J., NISHIKAWA, K., NUSSINOV, R., OBRADOVIC, Z., PAPPU, R. V., ROST, B., SELENKO, P., SUBRAMANIAM, V., SUSSMAN, J. L., TOMPA, P. & UVERSKY, V. N. 2013. What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disordered Proteins,* 1**,** e24157.

DUNKER, A. K., BONDOS, S. E., HUANG, F. & OLDFIELD, C. J. 2015. Intrinsically disordered proteins and multicellular organisms. *Semin Cell Dev Biol,* 37**,** 44-55.

DUNKER, A. K., BROWN, C. J. & OBRADOVIC, Z. 2002. Identification and functions of usefully disordered proteins. *Adv Protein Chem,* 62**,** 25-49.

DUNKER, A. K., CORTESE, M. S., ROMERO, P., IAKOUCHEVA, L. M. & UVERSKY, V. N. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J,* 272**,** 5129-48.

DUNKER, A. K., GARNER, E., GUILLIOT, S., ROMERO, P., ALBRECHT, K., HART, J., OBRADOVIC, Z., KISSINGER, C. & VILLAFRANCA, J. E. 1998. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput***,** 473-84.

DUNKER, A. K. & KRIWACKI, R. W. 2011. The orderly chaos of proteins. *Sci Am,* 304**,** 68-73.

DUNKER, A. K., LAWSON, J. D., BROWN, C. J., WILLIAMS, R. M., ROMERO, P., OH, J. S., OLDFIELD, C. J., CAMPEN, A. M., RATLIFF, C. M., HIPPS, K. W., AUSIO, J., NISSEN, M. S., REEVES, R., KANG, C., KISSINGER, C. R., BAILEY, R. W.,

GRISWOLD, M. D., CHIU, W., GARNER, E. C. & OBRADOVIC, Z. 2001. Intrinsically disordered protein. *J Mol Graph Model,* 19**,** 26-59.

DUNKER, A. K., OBRADOVIC, Z., ROMERO, P., GARNER, E. C. & BROWN, C. J. 2000. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform,* 11**,** 161-71.

DUTT, M., DHEKNEY, S. A., SORIANO, L., KANDEL, R. & GROSSER, J. W. 2014. Temporal and spatial control of gene expression in horticultural crops. *Horticulture Research,* 1**,** 14047.

DYSON, H. J. & WRIGHT, P. E. 2002. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol,* 12**,** 54-60.

DYSON, H. J. & WRIGHT, P. E. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol,* 6**,** 197-208.

EBERT, M. O., BAE, S. H., DYSON, H. J. & WRIGHT, P. E. 2008. NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR. *Biochemistry,* 47**,** 1299-308.

EDELHEIT, O., HANUKOGLU, A. & HANUKOGLU, I. 2009. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol,* 9**,** 61.

EDWARDS, A. M., ARROWSMITH, C. H., CHRISTENDAT, D., DHARAMSI, A., FRIESEN, J. D., GREENBLATT, J. F. & VEDADI, M. 2000. Protein production: feeding the crystallographers and NMR spectroscopists. *Nat Struct Biol,* 7 Suppl**,** 970-2.

EGEA, P. F., STROUD, R. M. & WALTER, P. 2005. Targeting proteins to membranes: structure of the signal recognition particle. *Current Opinion in Structural Biology,* 15**,** 213-220.

ELLIS, R. E., YUAN, J. & HORVITZ, H. R. 1991. Mechanisms and Functions of Cell Death. *Annual Review of Cell Biology,* 7**,** 663-698.

ELMORE, S. 2007. Apoptosis: A Review of Programmed Cell Death. *Toxicologic pathology,* 35**,** 495-516.

ERB, M., MELDAU, S. & HOWE, G. A. 2012. Role of phytohormones in insect-specific plant reactions. *Trends in Plant Science,* 17**,** 250-259.

EULGEM, T., RUSHTON, P. J., ROBATZEK, S. & SOMSSICH, I. E. 2000. The WRKY superfamily of plant transcription factors. *Trends in Plant Science,* 5**,** 199-206.

FABIAN, T., LORBIECKE, R., UMEDA, M. & SAUTER, M. 2000. The cell cycle genes cycA1;1 and cdc2Os-3 are coordinately regulated by gibberellin in planta. *Planta,* 211**,** 376-383.

FAGERSTROM, K. O. & SCHNEIDER, N. G. 1989. Measuring nicotine dependence: A review of the Fagerstrom Tolerance Questionnaire. *Journal of Behavioral Medicine,* 12**,** 159-182.

FAN, M., BAI, M.-Y., KIM, J.-G., WANG, T., OH, E., CHEN, L., PARK, C. H., SON, S.-H., KIM, S.-K., MUDGETT, M. B. & WANG, Z.-Y. 2014. The bHLH Transcription Factor HBI1 Mediates the Trade-Off between Growth and Pathogen-Associated Molecular Pattern–Triggered Immunity in <em>Arabidopsis</em&gt. *The Plant Cell,* 26**,** 828.

FENG, S., MARTINEZ, C., GUSMAROLI, G., WANG, Y., ZHOU, J., WANG, F., CHEN, L., YU, L., IGLESIAS-PEDRAZ, J. M., KIRCHER, S., SCHAFER, E., FU, X., FAN, L.-M. & DENG, X. W. 2008. Coordinated regulation of Arabidopsisthaliana development by light and gibberellins. *Nature,* 451**,** 475-479.

FESIK, S. W. 2000. Insights into Programmed Cell Death through Structural Biology. *Cell,* 103**,** 273-282.

FINKELSTEIN, R., REEVES, W., ARIIZUMI, T. & STEBER, C. 2008. Molecular Aspects of Seed Dormancy. *Annual Review of Plant Biology,* 59**,** 387-415.

FLOR, H. H. 1971. Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology,* 9**,** 275-296.

FLUMAN, N., NAVON, S., BIBI, E. & PILPEL, Y. 2014. mRNA-programmed translation pauses in the targeting of E. coli membrane proteins. *eLife,* 3**,** e03440.

FORMAN-KAY, J. D. & MITTAG, T. 2013. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure,* 21**,** 1492-9.

FRANKEL, A. D., BERG, J. M. & PABO, C. O. 1987. Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proc Natl Acad Sci U S A,* 84**,** 4841-5.

FREEMAN-COOK, L. L., DIXON, A. M., FRANK, J. B., XIA, Y., ELY, L., GERSTEIN, M., ENGELMAN, D. M. & DIMAIO, D. 2004. Selection and characterization of small random transmembrane proteins that bind and activate the platelet-derived growth factor beta receptor. *J Mol Biol,* 338**,** 907-20.

FRIDMAN, Y. & SAVALDI-GOLDSTEIN, S. 2013. Brassinosteroids in growth control: How, when and where. *Plant Science,* 209**,** 24-31.

FRITZ, S. J. & GJERDE, D. J. 2009. *Ion Chromatography*.

FU, J. & WANG, S. 2011. Insights into Auxin Signaling in Plant–Pathogen Interactions. *Frontiers in plant science,* 2**,** 74.

GÄBLER, K., BEHRMANN, I. & HAAN, C. 2013. JAK2 mutants (e.g., JAK2V617F) and their importance as drug targets in myeloproliferative neoplasms. *JAK-STAT,* 2**,** e25025.

GAMPALA, S. S., KIM, T.-W., HE, J.-X., TANG, W., DENG, Z., BAI, M.-Y., GUAN, S., LALONDE, S., SUN, Y., GENDRON, J. M., CHEN, H., SHIBAGAKI, N., FERL, R. J., EHRHARDT, D., CHONG, K., BURLINGAME, A. L. & WANG, Z.-Y. 2007. An Essential Role for 14-3-3 Proteins in Brassinosteroid Signal Transduction in Arabidopsis. *Developmental Cell,* 13**,** 177-189.

GAO, G., ZHANG, S., WANG, C., YANG, X., WANG, Y., SU, X., DU, J. & YANG, C. 2011. Arabidopsis CPR5 independently regulates seed germination and postgermination arrest of development through LOX pathway and ABA signaling. *PLoS One,* 6**,** e19406.

GAO, J. & XU, D. 2012. Correlation between posttranslational modification and intrinsic disorder in protein. *Pac Symp Biocomput***,** 94-103.

GARAPATI, P., XUE, G.-P., MUNNÉ-BOSCH, S. & BALAZADEH, S. 2015. Transcription Factor ATAF1 in Arabidopsis Promotes Senescence by Direct Regulation of Key Chloroplast Maintenance and Senescence Transcriptional Cascades. *Plant Physiology,* 168**,** 1122-1139.

GARDNER, J. M., SMOYER, C. J., STENSRUD, E. S., ALEXANDER, R., GOGOL, M., WIEGRAEBE, W. & JASPERSEN, S. L. 2011. Targeting of the SUN protein Mps3 to

the inner nuclear membrane by the histone variant H2A.Z. *The Journal of Cell Biology,* 193**,** 489-507.
GELVIN, S. B. 2003. Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev,* 67**,** 16-37, table of contents.
GELVIN, S. B. 2010. Plant proteins involved in Agrobacterium-mediated genetic transformation. *Annu Rev Phytopathol,* 48**,** 45-68.
GILCHRIST, D. G. 1998. PROGRAMMED CELL DEATH IN PLANT DISEASE: The Purpose and Promise of Cellular Suicide. *Annual Review of Phytopathology,* 36**,** 393-414.
GILL, S. S. & TUTEJA, N. 2010. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiology and Biochemistry,* 48**,** 909-930.
GIMENEZ-IBANEZ, S., HANN, D. R., NTOUKAKIS, V., PETUTSCHNIG, E., LIPKA, V. & RATHJEN, J. P. 2009. AvrPtoB Targets the LysM Receptor Kinase CERK1 to Promote Bacterial Virulence on Plants. *Current Biology,* 19**,** 423-429.
GINGOLD, H. & PILPEL, Y. 2011. Determinants of translation efficiency and accuracy. *Molecular Systems Biology,* 7**,** 481-481.
GIRON, M. D. & SALTO, R. 2011. From green to blue: site-directed mutagenesis of the green fluorescent protein to teach protein structure-function relationships. *Biochem Mol Biol Educ,* 39**,** 309-15.
GLICKMANN, E., GARDAN, L., JACQUET, S., HUSSAIN, S., ELASRI, M., PETIT, A. & DESSAUX, Y. 1998. Auxin Production Is a Common Feature of Most Pathovars of Pseudomonas syringae. *Molecular Plant-Microbe Interactions,* 11**,** 156-162.
GONZALEZ, N., DE BODT, S., SULPICE, R., JIKUMARU, Y., CHAE, E., DHONDT, S., VAN DAELE, T., DE MILDE, L., WEIGEL, D., KAMIYA, Y., STITT, M., BEEMSTER, G. T. S. & INZÉ, D. 2010. Increased Leaf Size: Different Means to an End. *Plant Physiology,* 153**,** 1261-1279.
GOOKIN, T. E. & ASSMANN, S. M. 2014. Significant reduction of BiFC non-specific assembly facilitates in planta assessment of heterotrimeric G-protein interactors. *The Plant Journal,* 80**,** 553-567.
GORLICH, D. & KUTAY, U. 1999. Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol,* 15**,** 607-60.
GREENBERG, J., ARNDT, J., SIMON, L., PYSZCZYNSKI, T. & SOLOMON, S. 2000. Proximal and Distal Defenses in Response to Reminders of One's Mortality: Evidence of a Temporal Sequence. *Personality and Social Psychology Bulletin,* 26**,** 91-99.
GREENBERG, J. T. 1996. Programmed cell death: a way of life for plants. *Proceedings of the National Academy of Sciences,* 93**,** 12094-12097.
GU, Y., ZEBELL, SOPHIA G., LIANG, Z., WANG, S., KANG, B.-H. & DONG, X. 2016. Nuclear Pore Permeabilization Is a Convergent Signaling Event in Effector-Triggered Immunity. *Cell,* 166**,** 1526-1538.e11.
GUO, J. 2014. Transcription: the epicenter of gene expression. *Journal of Zhejiang University. Science. B,* 15**,** 409-411.
GUO, Y. & GAN, S.-S. 2012. Convergence and divergence in gene expression profiles induced by leaf senescence and 27 senescence-promoting hormonal, pathological and environmental stress treatments. *Plant, Cell & Environment,* 35**,** 644-655.

HÄFFNER, E., KONIETZKI, S. & DIEDERICHSEN, E. 2015. Keeping Control: The Role of Senescence and Development in Plant Pathogenesis and Defense. *Plants,* 4**,** 449-488.

HAIMEUR, A., CONSEIL, G., DEELEY, R. G. & COLE, S. P. 2004. Mutations of charged amino acids in or near the transmembrane helices of the second membrane spanning domain differentially affect the substrate specificity and transport activity of the multidrug resistance protein MRP1 (ABCC1). *Mol Pharmacol,* 65**,** 1375-85.

HANSEN, G. & WRIGHT, M. S. 1999. Recent advances in the transformation of plants. *Trends Plant Sci,* 4**,** 226-231.

HARBURY, P. B., ZHANG, T., KIM, P. S. & ALBER, T. 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science,* 262**,** 1401-7.

HARMEIER, A., WOZNY, C., ROST, B. R., MUNTER, L. M., HUA, H., GEORGIEV, O., BEYERMANN, M., HILDEBRAND, P. W., WEISE, C., SCHAFFNER, W., SCHMITZ, D. & MULTHAUP, G. 2009. Role of amyloid-beta glycine 33 in oligomerization, toxicity, and neuronal plasticity. *J Neurosci,* 29**,** 7582-90.

HARREMAN, M. T., KLINE, T. M., MILFORD, H. G., HARBEN, M. B., HODEL, A. E. & CORBETT, A. H. 2004. Regulation of nuclear import by phosphorylation adjacent to nuclear localization signals. *J Biol Chem,* 279**,** 20613-21.

HARRISON, S. J., MOTT, E. K., PARSLEY, K., ASPINALL, S., GRAY, J. C. & COTTAGE, A. 2006. A rapid and robust method of identifying transformed Arabidopsis thaliana seedlings following floral dip transformation. *Plant Methods,* 2**,** 19.

HASE, Y., TRUNG, K. H., MATSUNAGA, T. & TANAKA, A. 2006. A mutation in the uvi4 gene promotes progression of endo-reduplication and confers increased tolerance towards ultraviolet B light. *The Plant Journal,* 46**,** 317-326.

HASTY, J. & COLLINS, J. J. 2001. Protein interactions. Unspinning the web. *Nature,* 411**,** 30-1.

HAYNES, C. & IAKOUCHEVA, L. M. 2006. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res,* 34**,** 305-12.

HE, Y., FUKUSHIGE, H., HILDEBRAND, D. F. & GAN, S. 2002. Evidence Supporting a Role of Jasmonic Acid in Arabidopsis Leaf Senescence. *Plant Physiology,* 128**,** 876-884.

HE, Y. & GAN, S. 2002. A Gene Encoding an Acyl Hydrolase Is Involved in Leaf Senescence in Arabidopsis. *The Plant Cell,* 14**,** 805-815.

HERBERS, K., MEUWLY, P., FROMMER, W. B., METRAUX, J. P. & SONNEWALD, U. 1996. Systemic Acquired Resistance Mediated by the Ectopic Expression of Invertase: Possible Hexose Sensing in the Secretory Pathway. *The Plant Cell,* 8**,** 793-803.

HERMS, D. A. & MATTSON, W. J. 1992. The Dilemma of Plants: To Grow or Defend. *The Quarterly Review of Biology,* 67**,** 283-335.

HEYMAN, J., VAN DEN DAELE, H., DE WIT, K., BOUDOLF, V., BERCKMANS, B., VERKEST, A., KAMEI, C. L. A., DE JAEGER, G., KONCZ, C. & DE VEYLDER, L. 2011. Arabidopsis ULTRAVIOLET-B-INSENSITIVE4 Maintains Cell Division Activity by Temporal Inhibition of the Anaphase-Promoting Complex/Cyclosome. *The Plant Cell,* 23**,** 4394-4410.

HILF, R. J. & DUTZLER, R. 2008. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature,* 452**,** 375-9.

HIROAKI, Y., TANI, K., KAMEGAWA, A., GYOBU, N., NISHIKAWA, K., SUZUKI, H., WALZ, T., SASAKI, S., MITSUOKA, K., KIMURA, K., MIZOGUCHI, A. & FUJIYOSHI, Y. 2006. Implications of the aquaporin-4 structure on array formation and cell adhesion. *J Mol Biol,* 355**,** 628-39.

HOFFLAND, E., PIETERSE, C. M. J., BIK, L. & VAN PELT, J. A. 1995. Induced systemic resistance in radish is not associated with accumulation of pathogenesis-related proteins. *Physiological and Molecular Plant Pathology,* 46**,** 309-320.

HOH, J. H. 1998. Functional protein domains from the thermally driven motion of polypeptide chains: a proposal. *Proteins,* 32**,** 223-8.

HONG-YING, H., WEN-LING, L., LI-YING, S., ZHAN, S., CHENG-WEI, Y. & CHANG-LIAN, P. 2010. Increased resistance of Arabidopsis cpr5 mutant to H 2 O 2-induced photooxidation. *Pak J Bot,* 42**,** 1041-1049.

HONIG, B. & YANG, A. S. 1995. Free energy balance in protein folding. *Adv Protein Chem,* 46**,** 27-58.

HSU, S. C. & HUNG, M. C. 2007. Characterization of a novel tripartite nuclear localization sequence in the EGFR family. *J Biol Chem,* 282**,** 10432-40.

HSU, W. L., OLDFIELD, C. J., XUE, B., MENG, J., HUANG, F., ROMERO, P., UVERSKY, V. N. & DUNKER, A. K. 2013. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci,* 22**,** 258-73.

HUANG, H. & SARAI, A. 2012. Analysis of the relationships between evolvability, thermodynamics, and the functions of intrinsically disordered proteins/regions. *Comput Biol Chem,* 41**,** 51-7.

HUANG, Y. J., ACTON, T. B. & MONTELIONE, G. T. 2014. DisMeta – a Meta Server for Construct Design and Optimization. *Methods in molecular biology (Clifton, N.J.),* 1091**,** 3-16.

HUOT, B., YAO, J., MONTGOMERY, B. L. & HE, S. Y. 2014. Growth–Defense Tradeoffs in Plants: A Balancing Act to Optimize Fitness. *Molecular Plant (Oxford University Press / USA),* 7**,** 1-1.

HWANG, I. & SAKAKIBARA, H. 2006. *Cytokinin biosynthesis and perception*.

IGARI, K., ENDO, S., HIBARA, K.-I., AIDA, M., SAKAKIBARA, H., KAWASAKI, T. & TASAKA, M. 2008. Constitutive activation of a CC-NB-LRR protein alters morphogenesis through the cytokinin pathway in Arabidopsis. *The Plant Journal,* 55**,** 14-27.

INABA, J.-I., KIM, B. M., SHIMURA, H. & MASUTA, C. 2011. Virus-Induced Necrosis Is a Consequence of Direct Protein-Protein Interaction between a Viral RNA-Silencing Suppressor and a Host Catalase. *Plant Physiology,* 156**,** 2026-2036.

INAGAKI, S. & UMEDA, M. 2011. Cell-Cycle Control and Plant Development. *International Review of Cell and Molecular Biology,* 291**,** 227-261.

INGOLIA, N. T., GHAEMMAGHAMI, S., NEWMAN, J. R. S. & WEISSMAN, J. S. 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science (New York, N.Y.),* 324**,** 218-223.

INOUE, H., NOJIMA, H. & OKAYAMA, H. 1990. High efficiency transformation of Escherichia coli with plasmids. *Gene,* 96**,** 23-28.

IQBAL, N., KHAN, N. A., FERRANTE, A., TRIVELLINI, A., FRANCINI, A. & KHAN, M. I. R. 2017. Ethylene Role in Plant Growth, Development and Senescence: Interaction with Other Phytohormones. *Frontiers in Plant Science,* 8**,** 475.

ISHIDA, T. & KINOSHITA, K. 2008. Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics,* 24**,** 1344-1348.

ISHIHARA, T., KADOYA, T. & YAMAMOTO, S. 2007. Application of a chromatography model with linear gradient elution experimental data to the rapid scale-up in ion-exchange process chromatography of proteins. *Journal of Chromatography A,* 1162**,** 34-40.

IWATA, E., IKEDA, S., MATSUNAGA, S., KURATA, M., YOSHIOKA, Y., CRIQUI, M.-C., GENSCHIK, P. & ITO, M. 2011. GIGAS CELL1, a Novel Negative Regulator of the Anaphase-Promoting Complex/Cyclosome, Is Required for Proper Mitotic Progression and Cell Fate Determination in Arabidopsis. *The Plant Cell,* 23**,** 4382-4393.

JABS, T., TSCHÖPE, M., COLLING, C., HAHLBROCK, K. & SCHEEL, D. 1997. Elicitor-stimulated ion fluxes and O(2)(−) from the oxidative burst are essential components in triggering defense gene activation and phytoalexin synthesis in parsley. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 4800-4805.

JACKSON, M. B., FENNING, T. M., DREW, M. C. & SAKER, L. R. 1985. Stimulation of ethylene production and gas-space (aerenchyma) formation in adventitious roots of Zea mays L. by small partial pressures of oxygen. *Planta,* 165**,** 486-492.

JACOBSON, M. D., WEIL, M. & RAFF, M. C. 1997. Programmed Cell Death in Animal Development. *Cell,* 88**,** 347-354.

JAKEL, S., ALBIG, W., KUTAY, U., BISCHOFF, F. R., SCHWAMBORN, K., DOENECKE, D. & GORLICH, D. 1999. The importin beta/importin 7 heterodimer is a functional nuclear import receptor for histone H1. *EMBO J,* 18**,** 2411-23.

JANS, D. A. & HUBNER, S. 1996. Regulation of protein transport to the nucleus: central role of phosphorylation. *Physiol Rev,* 76**,** 651-85.

JASTI, J., FURUKAWA, H., GONZALES, E. B. & GOUAUX, E. 2007. Structure of acid-sensing ion channel 1 at 1.9 A resolution and low pH. *Nature,* 449**,** 316-23.

JEWORUTZKI, E., ROELFSEMA, M. R. G., ANSCHÜTZ, U., KROL, E., ELZENGA, J. T. M., FELIX, G., BOLLER, T., HEDRICH, R. & BECKER, D. 2010. Early signaling through the Arabidopsis pattern recognition receptors FLS2 and EFR involves Ca2+-associated opening of plasma membrane anion channels. *The Plant Journal,* 62**,** 367-378.

JING, H. C., ANDERSON, L., STURRE, M. J., HILLE, J. & DIJKWEL, P. P. 2007. Arabidopsis CPR5 is a senescence-regulatory gene with pleiotropic functions as predicted by the evolutionary theory of senescence. *J Exp Bot,* 58**,** 3885-94.

JING, H. C. & DIJKWEL, P. P. 2008. CPR5: A Jack of all trades in plants. *Plant Signal Behav,* 3**,** 562-3.

JING, H. C., HEBELER, R., OELJEKLAUS, S., SITEK, B., STÜHLER, K., MEYER, H. E., STURRE, M. J. G., HILLE, J., WARSCHEID, B. & DIJKWEL, P. P. 2008. Early leaf senescence is associated with an altered cellular redox balance in Arabidopsis cpr5/old1 mutants. *Plant Biology,* 10**,** 85-98.

JING, H. C., HILLE, J. & DIJKWEL, P. P. 2003. Ageing in Plants: Conserved Strategies and Novel Pathways. *Plant Biology,* 5**,** 455-464.
JING, H. C., STURRE, M. J., HILLE, J. & DIJKWEL, P. P. 2002. Arabidopsis onset of leaf death mutants identify a regulatory pathway controlling leaf senescence. *Plant J,* 32**,** 51-63.
JING, Y., CUI, D., BAO, F., HU, Z., QIN, Z. & HU, Y. 2009. Tryptophan deficiency affects organ growth by retarding cell expansion in Arabidopsis. *The Plant Journal,* 57**,** 511-521.
JIRAGE, D., ZHOU, N., COOPER, B., CLARKE, J. D., DONG, X. & GLAZEBROOK, J. 2001. Constitutive salicylic acid-dependent signaling in cpr1 and cpr6 mutants requires PAD4. *Plant J,* 26**,** 395-407.
JIRGENSONS, B. 1966. Classification of proteins according to conformation. *Die Makromolekulare Chemie,* 91**,** 74-86.
JONES, J. D. G. & DANGL, J. L. 2006. The plant immune system. *Nature,* 444**,** 323-329.
JOSLYN, G., RICHARDSON, D. S., WHITE, R. & ALBER, T. 1993. Dimer formation by an N-terminal coiled coil in the APC protein. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 11109-11113.
JOURNOT-CATALINO, N., SOMSSICH, I. E., ROBY, D. & KROJ, T. 2006. The Transcription Factors WRKY11 and WRKY17 Act as Negative Regulators of Basal Resistance in Arabidopsis thaliana. *The Plant Cell,* 18**,** 3289-3302.
JUNGE, F., SCHNEIDER, B., RECKEL, S., SCHWARZ, D., DÖTSCH, V. & BERNHARD, F. 2008. Large-scale production of functional membrane proteins. *Cellular and Molecular Life Sciences,* 65**,** 1729-1755.
KALAB, P., WEIS, K. & HEALD, R. 2002. Visualization of a Ran-GTP gradient in interphase and mitotic Xenopus egg extracts. *Science,* 295**,** 2452-6.
KALDERON, D., RICHARDSON, W. D., MARKHAM, A. F. & SMITH, A. E. 1984. Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature,* 311**,** 33-8.
KASILI, R., WALKER, J. D., SIMMONS, L. A., ZHOU, J., DE VEYLDER, L. & LARKIN, J. C. 2010. SIAMESE Cooperates With the CDH1-like Protein CCS52A1 to Establish Endoreplication in Arabidopsis thaliana Trichomes. *Genetics,* 185**,** 257-268.
KELLEY, L. A. & STERNBERG, M. J. E. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protocols,* 4**,** 363-371.
KERR, J. F. R., WYLLIE, A. H. & CURRIE, A. R. 1972. Apoptosis: A Basic Biological Phenomenon with Wide-ranging Implications in Tissue Kinetics. *British Journal of Cancer,* 26**,** 239-257.
KHAN, T., DOUGLAS, G. M., PATEL, P., NGUYEN BA, A. N. & MOSES, A. M. 2015. Polymorphism Analysis Reveals Reduced Negative Selection and Elevated Rate of Insertions and Deletions in Intrinsically Disordered Protein Regions. *Genome Biol Evol,* 7**,** 1815-26.
KHURANA, S. M. P., PANDEY, S. K., SARKAR, D. & CHANEMOUGASOUNDHARAM, A. 2005. Apoptosis in plant disease response: A close encounter of the pathogen kind. *Current Science,* 88**,** 740-752.

KIM, H. J., RYU, H., HONG, S. H., WOO, H. R., LIM, P. O., LEE, I. C., SHEEN, J., NAM, H. G. & HWANG, I. 2006a. Cytokinin-mediated control of leaf longevity by AHK3 through phosphorylation of ARR2 in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 814-819.

KIM, J., CHANG, C. & TUCKER, M. L. 2015. To grow old: regulatory role of ethylene and jasmonic acid in senescence. *Frontiers in Plant Science,* 6**,** 20.

KIM, K.-C., FAN, B. & CHEN, Z. 2006b. Pathogen-Induced Arabidopsis WRKY7 Is a Transcriptional Repressor and Enhances Plant Susceptibility to Pseudomonas syringae. *Plant Physiology,* 142**,** 1180-1192.

KIM, S., JEON, T. J., OBERAI, A., YANG, D., SCHMIDT, J. J. & BOWIE, J. U. 2005. Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proc Natl Acad Sci U S A,* 102**,** 14278-83.

KIMCHI-SARFATY, C., OH, J. M., KIM, I.-W., SAUNA, Z. E., CALCAGNO, A. M., AMBUDKAR, S. V. & GOTTESMAN, M. M. 2007. A "Silent" Polymorphism in the MDR1 Gene Changes Substrate Specificity. *Science,* 315**,** 525.

KIRIK, V., BOUYER, D., SCHOBINGER, U., BECHTOLD, N., HERZOG, M., BONNEVILLE, J. M. & HULSKAMP, M. 2001. CPR5 is involved in cell proliferation and cell death control and encodes a novel transmembrane protein. *Curr Biol,* 11**,** 1891-5.

KOMAR, A. A., LESNIK, T. & REISS, C. 1999. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Letters,* 462**,** 387-391.

KOSUGI, S., HASEBE, M., TOMITA, M. & YANAGAWA, H. 2009. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A,* 106**,** 10171-6.

KOTTA-LOIZOU, I., TSAOUSIS, G. N. & HAMODRAKAS, S. J. 2013. Analysis of Molecular Recognition Features (MoRFs) in membrane proteins. *Biochimica et biophysica acta,* 1834**,** 798-807.

KOVALCHUK, I., KOVALCHUK, O., KALCK, V., BOYKO, V., FILKOWSKI, J., HEINLEIN, M. & HOHN, B. 2003. Pathogen-induced systemic plant signal triggers DNA rearrangements. *Nature,* 423**,** 760-762.

KOZAK, M. 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem,* 266**,** 19867-70.

KOZAK, M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene,* 234**,** 187-208.

KRAMER, RYAN M., SHENDE, VARAD R., MOTL, N., PACE, C N. & SCHOLTZ, J M. 2012. Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility. *Biophysical Journal,* 102**,** 1907-1915.

KRIWACKI, R. W., HENGST, L., TENNANT, L., REED, S. I. & WRIGHT, P. E. 1996. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 11504-11509.

KRYLOV, D. & VINSON, C. R. 2001. *Leucine Zipper,* Nature Publishing Group.

KUCERA, B., COHN, M. A. & LEUBNER-METZGER, G. 2007. Plant hormone interactions during seed dormancy release and germination. *Seed Science Research,* 15**,** 281-307.

KUDLA, G., MURRAY, A. W., TOLLERVEY, D. & PLOTKIN, J. B. 2009. Coding-sequence determinants of gene expression in Escherichia coli. *Science (New York, N.Y.),* 324**,** 255-258.

KUO, A., CAPPELLUTI, S., CERVANTES-CERVANTES, M., RODRIGUEZ, M. & BUSH, D. S. 1996. Okadaic acid, a protein phosphatase inhibitor, blocks calcium changes, gene expression, and cell death induced by gibberellin in wheat aleurone cells. *The Plant Cell,* 8**,** 259-269.

KURAKAWA, T., UEDA, N., MAEKAWA, M., KOBAYASHI, K., KOJIMA, M., NAGATO, Y., SAKAKIBARA, H. & KYOZUKA, J. 2007. Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature,* 445**,** 652-655.

KURGAN, L., RAZIB, A. A., AGHAKHANI, S., DICK, S., MIZIANTY, M. & JAHANDIDEH, S. 2009. CRYSTALP2: sequence-based protein crystallization propensity prediction. *Bmc Structural Biology,* 9.

KURISU, G., ZHANG, H., SMITH, J. L. & CRAMER, W. A. 2003. Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. *Science,* 302**,** 1009-14.

KUTAY, U., LIPOWSKY, G., IZAURRALDE, E., BISCHOFF, F. R., SCHWARZMAIER, P., HARTMANN, E. & GORLICH, D. 1998. Identification of a tRNA-specific nuclear export receptor. *Mol Cell,* 1**,** 359-69.

LABA, J. K., STEEN, A. & VEENHOFF, L. M. 2014. Traffic to the inner membrane of the nuclear envelope. *Current Opinion in Cell Biology,* 28**,** 36-45.

LAMB, C. & DIXON, R. A. 1997. THE OXIDATIVE BURST IN PLANT DISEASE RESISTANCE. *Annual Review of Plant Physiology and Plant Molecular Biology,* 48**,** 251-275.

LAMMENS, T., BOUDOLF, V., KHEIBARSHEKAN, L., PANAGIOTIS ZALMAS, L., GAAMOUCHE, T., MAES, S., VANSTRAELEN, M., KONDOROSI, E., LA THANGUE, N. B., GOVAERTS, W., INZÉ, D. & DE VEYLDER, L. 2008. Atypical E2F activity restrains APC/C(CCS52A2) function obligatory for endocycle onset. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 14721-14726.

LANDSCHULZ, W. H., JOHNSON, P. F. & MCKNIGHT, S. L. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science,* 240**,** 1759-1764.

LANFORD, R. E. & BUTEL, J. S. 1984. Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. *Cell,* 37**,** 801-13.

LANGE, A., MILLS, R. E., LANGE, C. J., STEWART, M., DEVINE, S. E. & CORBETT, A. H. 2007. Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J Biol Chem,* 282**,** 5101-5.

LARSON-RABIN, Z., LI, Z., MASSON, P. H. & DAY, C. D. 2009. FZR2/CCS52A1 Expression Is a Determinant of Endoreduplication and Cell Expansion in Arabidopsis. *Plant Physiology,* 149**,** 874-884.

LEE, B. J., CANSIZOGLU, A. E., SUEL, K. E., LOUIS, T. H., ZHANG, Z. & CHOOK, Y. M. 2006. Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell,* 126**,** 543-58.

LEE, C., KALMAR, L., XUE, B., TOMPA, P., DAUGHDRILL, G. W., UVERSKY, V. N. & HAN, K. H. 2014. Contribution of proline to the pre-structuring tendency of transient helical secondary structure elements in intrinsically disordered proteins. *Biochim Biophys Acta,* 1840**,** 993-1003.

LEE, C. W., EFETOVA, M., ENGELMANN, J. C., KRAMELL, R., WASTERNACK, C., LUDWIG-MULLER, J., HEDRICH, R. & DEEKEN, R. 2009. Agrobacterium tumefaciens promotes tumor induction by modulating pathogen defense in Arabidopsis thaliana. *Plant Cell,* 21**,** 2948-62.

LETUNIC, I., DOERKS, T. & BORK, P. 2009. SMART 6: recent updates and new developments. *Nucleic Acids Research,* 37**,** D229-D232.

LEVINE, A., PENNELL, R. I., ALVAREZ, M. E., PALMER, R. & LAMB, C. 1996. Calcium-mediated apoptosis in a plant hypersensitive disease resistance response. *Current Biology,* 6**,** 427-437.

LEVINE, A., TENHAKEN, R., DIXON, R. & LAMB, C. 1994. H2O2 from the oxidative burst orchestrates the plant hypersensitive disease resistance response. *Cell,* 79**,** 583-593.

LEVISON, P. R. 2003. Large-scale ion-exchange column chromatography of proteins. Comparison of different formats. *J Chromatogr B Analyt Technol Biomed Life Sci,* 790**,** 17-33.

LI, J., BRADER, G., KARIOLA, T. & TAPIO PALVA, E. 2006. WRKY70 modulates the selection of signaling pathways in plant defense. *The Plant Journal,* 46**,** 477-491.

LI, J., BRADER, G. & PALVA, E. T. 2004. The WRKY70 Transcription Factor: A Node of Convergence for Jasmonate-Mediated and Salicylate-Mediated Signals in Plant Defense. *The Plant Cell,* 16**,** 319-331.

LI, J. F., PARK, E., VON ARNIM, A. G. & NEBENFUHR, A. 2009. The FAST technique: a simplified Agrobacterium-based transformation method for transient gene expression analysis in seedlings of Arabidopsis and other plant species. *Plant Methods,* 5**,** 6.

LIANG, S. H. & CLARKE, M. F. 1999. The nuclear import of p53 is determined by the presence of a basic domain and its relative position to the nuclear localization signal. *Oncogene,* 18**,** 2163-6.

LIGHT, S. & ELOFSSON, A. 2013. The impact of splicing on protein domain architecture. *Curr Opin Struct Biol,* 23**,** 451-8.

LIGHT, S., SAGIT, R., EKMAN, D. & ELOFSSON, A. 2013. Long indels are disordered: a study of disorder and indels in homologous eukaryotic proteins. *Biochim Biophys Acta,* 1834**,** 890-7.

LIN, J.-R. & HU, J. 2013. SeqNLS: Nuclear Localization Signal Prediction Based on Frequent Pattern Mining and Linear Motif Scoring. *PLoS ONE,* 8**,** e76864.

LINDING, R., JENSEN, L. J., DIELLA, F., BORK, P., GIBSON, T. J. & RUSSELL, R. B. 2003. Protein disorder prediction: implications for structural proteomics. *Structure,* 11**,** 1453-9.

LING, M. M. & ROBINSON, B. H. 1997. Approaches to DNA mutagenesis: an overview. *Anal Biochem,* 254**,** 157-78.

LIU, W. & STEWART JR, C. N. 2016. Plant synthetic promoters and transcription factors. *Current Opinion in Biotechnology,* 37**,** 36-44.

LOGEMANN, E., BIRKENBIHL, R. P., ULKER, B. & SOMSSICH, I. E. 2006. An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol. *Plant Methods,* 2**,** 16.

LOKAREDDY, R. K., HAPSARI, R. A., VAN RHEENEN, M., PUMROY, R. A., BHARDWAJ, A., STEEN, A., VEENHOFF, L. M. & CINGOLANI, G. 2015. Distinctive properties of the Nuclear Localization Signals of Inner Nuclear Membrane proteins Heh1 and Heh2. *Structure (London, England : 1993),* 23**,** 1305-1316.

LONG, S. B., CAMPBELL, E. B. & MACKINNON, R. 2005. Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. *Science,* 309**,** 897-903.

LONG, S. B., TAO, X., CAMPBELL, E. B. & MACKINNON, R. 2007. Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment. *Nature,* 450**,** 376-82.

LORRAIN, S., VAILLEAU, F., BALAGUÉ, C. & ROBY, D. 2003. Lesion mimic mutants: keys for deciphering cell death and defense pathways in plants? *Trends in Plant Science,* 8**,** 263-271.

LOUGHERY, J., COX, M., SMITH, L. M. & MEEK, D. W. 2014. Critical role for p53-serine 15 phosphorylation in stimulating transactivation at p53-responsive promoters. *Nucleic Acids Res,* 42**,** 7666-80.

LOVE, A. J., MILNER, J. J. & SADANANDOM, A. 2008. Timing is everything: regulatory overlap in plant cell death. *Trends in Plant Science,* 13**,** 589-595.

LOZANO-DURÁN, R., MACHO, A. P., BOUTROT, F., SEGONZAC, C., SOMSSICH, I. E. & ZIPFEL, C. 2013. The transcriptional regulator BZR1 mediates trade-off between plant innate immunity and growth. *Elife,* 2**,** e00983.

LOZANO-DURÁN, R. & ZIPFEL, C. 2015. Trade-off between growth and immunity: role of brassinosteroids. *Trends in Plant Science,* 20**,** 12-19.

LU, T. T., ONIZAWA, M., HAMMER, G., TURER, E. E., YIN, Q., DAMKO, E., AGELIDES, A., SHIFRIN, N., ADVINCULA, R., BARRERA, J., MALYNN, B. A., WU, H. & MA, A. 2013. Dimerization and Ubiquitin Mediated Recruitment of A20, a Complex Deubiquitinating Enzyme. *Immunity,* 38**,** 896-905.

LÜ, X., DE LA PEÑA, L., BARKER, C., CAMPHAUSEN, K. & TOFILON, P. J. 2006. Radiation-Induced Changes in Gene Expression Involve Recruitment of Existing Messenger RNAs to and away from Polysomes. *Cancer Research,* 66**,** 1052.

LUIRINK, J., YU, Z., WAGNER, S. & DE GIER, J.-W. 2012. Biogenesis of inner membrane proteins in Escherichia coli. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 1817**,** 965-976.

LUPAS, A. N. & GRUBER, M. 2005. The structure of alpha-helical coiled coils. *Adv Protein Chem,* 70**,** 37-78.

LUSK, C. P., BLOBEL, G. & KING, M. C. 2007. Highway to the inner nuclear membrane: rules for the road. *Nat Rev Mol Cell Biol,* 8**,** 414-420.

MALINOVSKY, F. G., BATOUX, M., SCHWESSINGER, B., YOUN, J. H., STRANSFELD, L., WIN, J., KIM, S.-K. & ZIPFEL, C. 2014. Antagonistic Regulation of Growth and Immunity by the Arabidopsis Basic Helix-Loop-Helix Transcription Factor

HOMOLOG OF BRASSINOSTEROID ENHANCED EXPRESSION2 INTERACTING WITH INCREASED LEAF INCLINATION1 BINDING bHLH1. *Plant Physiology,* 164**,** 1443.

MALLET, P. L. & BACHAND, F. 2013. A proline-tyrosine nuclear localization signal (PY-NLS) is required for the nuclear import of fission yeast PAB2, but not of human PABPN1. *Traffic,* 14**,** 282-94.

MANULIS, S., HAVIV-CHESNER, A., BRANDL, M. T., LINDOW, S. E. & BARASH, I. 1998. Differential Involvement of Indole-3-Acetic Acid Biosynthetic Pathways in Pathogenicity and Epiphytic Fitness of Erwinia herbicola pv. gypsophilae. *Molecular Plant-Microbe Interactions,* 11**,** 634-642.

MARROCCO, K., BERGDOLL, M., ACHARD, P., CRIQUI, M.-C. & GENSCHIK, P. 2010. Selective proteolysis sets the tempo of the cell cycle. *Current Opinion in Plant Biology,* 13**,** 631-639.

MARTIN-SERRANO, J., PEREZ-CABALLERO, D. & BIENIASZ, P. D. 2004. Context-dependent effects of L domains and ubiquitination on viral budding. *J Virol,* 78**,** 5554-63.

MARTÍNEZ, C., PONS, E., PRATS, G. & LEÓN, J. 2004. Salicylic acid regulates flowering time and links defence responses and reproductive development. *The Plant Journal,* 37**,** 209-217.

MASON, A. B., HE, Q. Y., ADAMS, T. E., GUMEROV, D. R., KALTASHOV, I. A., NGUYEN, V. & MACGILLIVRAY, R. T. 2001. Expression, purification, and characterization of recombinant nonglycosylated human serum transferrin containing a C-terminal hexahistidine tag. *Protein Expr Purif,* 23**,** 142-50.

MASON, A. B., HE, Q. Y., HALBROOKS, P. J., EVERSE, S. J., GUMEROV, D. R., KALTASHOV, I. A., SMITH, V. C., HEWITT, J. & MACGILLIVRAY, R. T. 2002. Differential effect of a his tag at the N- and C-termini: functional studies with recombinant human serum transferrin. *Biochemistry,* 41**,** 9448-54.

MASOOD, A., IQBAL, N. & KHAN, N. A. 2012. Role of ethylene in alleviation of cadmium-induced photosynthetic capacity inhibition by sulphur in mustard. *Plant, Cell & Environment,* 35**,** 524-533.

MATSUMOTO-KITANO, M., KUSUMOTO, T., TARKOWSKI, P., KINOSHITA-TSUJIMURA, K., VÁCLAVÍKOVÁ, K., MIYAWAKI, K. & KAKIMOTO, T. 2008. Cytokinins are central regulators of cambial activity. *Proceedings of the National Academy of Sciences,* 105**,** 20027-20031.

MATTAJ, I. W. & ENGLMEIER, L. 1998. Nucleocytoplasmic transport: the soluble phase. *Annu Rev Biochem,* 67**,** 265-306.

MAUZY, C. A. & HERMODSON, M. A. 1992. Structural homology between rbs repressor and ribose binding protein implies functional similarity. *Protein Sci,* 1**,** 843-9.

MCCABE, P. F., LEVINE, A., MEIJER, P.-J., TAPON, N. A. & PENNELL, R. I. 1997. A programmed cell death pathway activated in carrot cells cultured at low cell density. *The Plant Journal,* 12**,** 267-280.

MCSTEEN, P. & ZHAO, Y. 2008. Plant Hormones and Signaling: Common Themes and New Developments. *Developmental Cell,* 14**,** 467-473.

MIAO, Y., LAUN, T., ZIMMERMANN, P. & ZENTGRAF, U. 2004. Targets of the WRKY53 transcription factor and its role during leaf senescence in Arabidopsis. *Plant Molecular Biology,* 55**,** 853-867.

MILLER, A., TSAI, C. H., HEMPHILL, D., ENDRES, M., RODERMEL, S. & SPALDING, M. 1997. Elevated CO2 Effects during Leaf Ontogeny (A New Perspective on Acclimation). *Plant Physiology,* 115**,** 1195-1200.

MILNE, J. J. 2011. Scale-up of protein purification: downstream processing issues. *Methods Mol Biol,* 681**,** 73-85.

MIZUNO, N., HOSOGI, N., PARK, P. & TAKUMI, S. 2010. Hypersensitive Response-Like Reaction Is Associated with Hybrid Necrosis in Interspecific Crosses between Tetraploid Wheat and Aegilops tauschii Coss. *PLoS ONE,* 5**,** e11326.

MOEDER, W. & YOSHIOKA, K. 2008. Lesion mimic mutants: A classical, yet still fundamental approach to study programmed cell death. *Plant Signaling & Behavior,* 3**,** 764-767.

MOHAN, A., OLDFIELD, C. J., RADIVOJAC, P., VACIC, V., CORTESE, M. S., DUNKER, A. K. & UVERSKY, V. N. 2006. Analysis of molecular recognition features (MoRFs). *J Mol Biol,* 362**,** 1043-59.

MOLINA, G. A., WATKINS, S. C. & TSANG, M. 2007. Generation of FGF reporter transgenic zebrafish and their utility in chemical screens. *BMC Dev Biol,* 7**,** 62.

MÖLLER, S., CRONING, M. D. R. & APWEILER, R. 2001. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics,* 17**,** 646-653.

MONTELIONE, G. T., ZHENG, D., HUANG, Y. J., GUNSALUS, K. C. & SZYPERSKI, T. 2000. Protein NMR spectroscopy in structural genomics. *Nat Struct Biol,* 7 Suppl**,** 982-5.

MOROIANU, J. 1999. Nuclear import and export pathways. *J Cell Biochem,* Suppl 32-33**,** 76-83.

MORRIS, K., MACKERNESS, S. A. H., PAGE, T., JOHN, C. F., MURPHY, A. M., CARR, J. P. & BUCHANAN-WOLLASTON, V. 2000. Salicylic acid has a role in regulating gene expression during leaf senescence. *The Plant Journal,* 23**,** 677-685.

MOUSTAFA, Y. M. & MORSI, R. E. 2013. Ion exchange chromatography-An overview. *Column Chromatography.* InTech.

MUR, L. A. J., KENTON, P., LLOYD, A. J., OUGHAM, H. & PRATS, E. 2008. The hypersensitive response; the centenary is upon us but how much do we know? *Journal of Experimental Botany,* 59**,** 501-520.

MURZIN, A. G., BRENNER, S. E., HUBBARD, T. & CHOTHIA, C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol,* 247**,** 536-40.

NAKASHITA, H., YASUDA, M., NITTA, T., ASAMI, T., FUJIOKA, S., ARAI, Y., SEKIMATA, K., TAKATSUTO, S., YAMAGUCHI, I. & YOSHIDA, S. 2003. Brassinosteroid functions in a broad range of disease resistance in tobacco and rice. *The Plant Journal,* 33**,** 887-898.

NAVABPOUR, S., MORRIS, K., ALLEN, R., HARRISON, E., A-H-MACKERNESS, S. & BUCHANAN-WOLLASTON, V. 2003. Expression of senescence-enhanced genes in response to oxidative stress. *Journal of Experimental Botany,* 54**,** 2285-2292.

NAVARRO, L., BARI, R., ACHARD, P., LISÓN, P., NEMRI, A., HARBERD, N. P. & JONES, J. D. G. 2008. DELLAs Control Plant Immune Responses by Modulating the Balance of Jasmonic Acid and Salicylic Acid Signaling. *Current Biology,* 18**,** 650-655.

NAVARRO, L., DUNOYER, P., JAY, F., ARNOLD, B., DHARMASIRI, N., ESTELLE, M., VOINNET, O. & JONES, J. D. G. 2006. A Plant miRNA Contributes to Antibacterial Resistance by Repressing Auxin Signaling. *Science,* 312**,** 436.

NAWRATH, C. & MÉTRAUX, J. P. 1999. Salicylic acid induction-deficient mutants of Arabidopsis express PR-2 and PR-5 and accumulate high levels of camalexin after pathogen inoculation. *The Plant Cell,* 11**,** 1393-1404.

NAZAR, R., KHAN, M. I. R., IQBAL, N., MASOOD, A. & KHAN, N. A. 2014. Involvement of ethylene in reversal of salt-inhibited photosynthesis by sulfur in mustard. *Physiologia Plantarum,* 152**,** 331-344.

NGUYEN, J. B. & MODIS, Y. 2013. Crystal structure of the dimeric coiled-coil domain of the cytosolic nucleic acid sensor LRRFIP1. *Journal of structural biology,* 181**,** 82-88.

NIKLAS, K. J., BONDOS, S. E., DUNKER, A. K. & NEWMAN, S. A. 2015. Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. *Frontiers in Cell and Developmental Biology,* 3**,** 8.

NILSSON, J., GRAHN, M. & WRIGHT, A. P. 2011. Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. *Genome Biol,* 12**,** R65.

NONAKA, S., YUHASHI, K., TAKADA, K., SUGAWARE, M., MINAMISAWA, K. & EZURA, H. 2008. Ethylene production in plants during transformation suppresses vir gene expression in Agrobacterium tumefaciens. *New Phytol,* 178**,** 647-56.

NORBURY, C. J. & HICKSON, I. D. 2001. Cellular responses to DNA damage. *Annu Rev Pharmacol Toxicol,* 41**,** 367-401.

NUGENT, T. & JONES, D. T. 2009. Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics,* 10**,** 159-159.

OATES, M. E., ROMERO, P., ISHIDA, T., GHALWASH, M., MIZIANTY, M. J., XUE, B., DOSZTANYI, Z., UVERSKY, V. N., OBRADOVIC, Z., KURGAN, L., DUNKER, A. K. & GOUGH, J. 2013. D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res,* 41**,** D508-16.

OGAWA, M., HANADA, A., YAMAUCHI, Y., KUWAHARA, A., KAMIYA, Y. & YAMAGUCHI, S. 2003. Gibberellin Biosynthesis and Response during Arabidopsis Seed Germination. *The Plant Cell,* 15**,** 1591.

OKADA, H. & MAK, T. W. 2004. Pathways of apoptotic and non-apoptotic death in tumour cells. *Nat Rev Cancer,* 4**,** 592-603.

OKADA, T. 1998. Interpretation of ion-exchange chromatographic retention based on an electrical double-layer model. *Anal Chem,* 70**,** 1692-700.

OKAMURA, H., GARCIA-RODRIGUEZ, C., MARTINSON, H., QIN, J., VIRSHUP, D. M. & RAO, A. 2004. A conserved docking motif for CK1 binding controls the nuclear localization of NFAT1. *Mol Cell Biol,* 24**,** 4184-95.

OLDFIELD, C., MENG, J., YANG, J., YANG, M., UVERSKY, V. & DUNKER, A. 2008. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics,* 9**,** 1-20.

OLDFIELD, C. J., CHENG, Y., CORTESE, M. S., BROWN, C. J., UVERSKY, V. N. & DUNKER, A. K. 2005. Comparing and combining predictors of mostly disordered proteins. *Biochemistry,* 44**,** 1989-2000.

OLDFIELD, C. J. & DUNKER, A. K. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem,* 83**,** 553-84.

ORENGO, C. A., MICHIE, A. D., JONES, S., JONES, D. T., SWINDELLS, M. B. & THORNTON, J. M. 1997. CATH – a hierarchic classification of protein domain structures. *Structure,* 5**,** 1093-1109.

OVERTON, I. M., PADOVANI, G., GIROLAMI, M. A. & BARTON, G. J. 2008. ParCrys: a Parzen window density estimation approach to protein crystallization propensity prediction. *Bioinformatics,* 24**,** 901-907.

OVERTON, I. M., VAN NIEKERK, C. A. & BARTON, G. J. 2011. XANNpred: neural nets that predict the propensity of a protein to yield diffraction-quality crystals. *Proteins,* 79**,** 1027-33.

PALLAVI, S. K., HO, D. M., HICKS, C., MIELE, L. & ARTAVANIS-TSAKONAS, S. 2012. Notch and Mef2 synergize to promote proliferation and metastasis through JNK signal activation in Drosophila. *The EMBO Journal,* 31**,** 2895-2907.

PALMA, K., THORGRIMSEN, S., MALINOVSKY, F. G., FIIL, B. K., NIELSEN, H. B., BRODERSEN, P., HOFIUS, D., PETERSEN, M. & MUNDY, J. 2010. Autoimmunity in Arabidopsis acd11 Is Mediated by Epigenetic Regulation of an Immune Receptor. *PLoS Pathogens,* 6**,** e1001137.

PARK, C. Y., LEE, J. H., YOO, J. H., MOON, B. C., CHOI, M. S., KANG, Y. H., LEE, S. M., KIM, H. S., KANG, K. Y., CHUNG, W. S., LIM, C. O. & CHO, M. J. 2005. WRKY group IId transcription factors interact with calmodulin. *FEBS Letters,* 579**,** 1545-1550.

PASTORI, G. M. & DEL RIO, L. A. 1997. Natural Senescence of Pea Leaves (An Activated Oxygen-Mediated Function for Peroxisomes). *Plant Physiology,* 113**,** 411-418.

PEACH, C. & VELTEN, J. 1991. Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. *Plant Molecular Biology,* 17**,** 49-60.

PEDERSEN, B. P., BUCH-PEDERSEN, M. J., MORTH, J. P., PALMGREN, M. G. & NISSEN, P. 2007. Crystal structure of the plasma membrane proton pump. *Nature,* 450**,** 1111-4.

PEGADARAJU, V., KNEPPER, C., REESE, J. & SHAH, J. 2005. Premature Leaf Senescence Modulated by the Arabidopsis PHYTOALEXIN DEFICIENT4 Gene Is Associated with Defense against the Phloem-Feeding Green Peach Aphid. *Plant Physiology,* 139**,** 1927.

PEMBERTON, L. F., BLOBEL, G. & ROSENBLUM, J. S. 1998. Transport routes through the nuclear pore complex. *Curr Opin Cell Biol,* 10**,** 392-9.

PEMBERTON, L. F. & PASCHAL, B. M. 2005. Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic,* 6**,** 187-98.

PENG, Z. & KURGAN, L. 2015. High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. *Nucleic Acids Research*.

PENNELL, R. I. & LAMB, C. 1997. Programmed Cell Death in Plants. *The Plant Cell,* 9**,** 1157-1168.

PENNINCKX, I. A., THOMMA, B. P., BUCHALA, A., METRAUX, J. P. & BROEKAERT, W. F. 1998. Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defensin gene in Arabidopsis. *Plant Cell,* 10**,** 2103-13.

PERAZZA, D., LAPORTE, F., BALAGUE, C., CHEVALIER, F., REMO, S., BOURGE, M., LARKIN, J., HERZOG, M. & VACHON, G. 2011. GeBP/GPL transcription factors regulate a subset of CPR5-dependent processes. *Plant Physiol,* 157**,** 1232-42.

PERTRY, I., VÁCLAVÍKOVÁ, K., GEMROTOVÁ, M., SPÍCHAL, L., GALUSZKA, P., DEPUYDT, S., TEMMERMAN, W., STES, E., DE KEYSER, A., RIEFLER, M., BIONDI, S., NOVÁK, O., SCHMÜLLING, T., STRNAD, M., TARKOWSKI, P., HOLSTERS, M. & VEREECKE, D. 2010. Rhodococcus fascians Impacts Plant Development Through the Dynamic Fas-Mediated Production of a Cytokinin Mix. *Molecular Plant-Microbe Interactions,* 23**,** 1164-1174.

PETERS, J.-M. 2006. *Peters JM.. The anaphase promoting complex/cyclosome: a machine designed to destroy. Nat Rev Mol Cell Biol 7: 644-656.*

PFAFFL, M. W., TICHOPAD, A., PRGOMET, C. & NEUVIANS, T. P. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnology Letters,* 26**,** 509-515.

PIETERSE, C. M. J., VAN DER DOES, D., ZAMIOUDIS, C., LEON-REYES, A. & VAN WEES, S. C. M. 2012. Hormonal Modulation of Plant Immunity. *Annual Review of Cell and Developmental Biology,* 28**,** 489-521.

PINA, A. S., LOWE, C. R. & ROQUE, A. C. 2014. Challenges and opportunities in the purification of recombinant tagged proteins. *Biotechnol Adv,* 32**,** 366-81.

PINO, M.-T., SKINNER, J. S., PARK, E.-J., JEKNIĆ, Z., HAYES, P. M., THOMASHOW, M. F. & CHEN, T. H. H. 2007a. Use of a stress inducible promoter to drive ectopic AtCBF expression improves potato freezing tolerance while minimizing negative effects on tuber yield. *Plant Biotechnology Journal,* 5**,** 591-604.

PINO, P., FOTH, B. J., KWOK, L.-Y., SHEINER, L., SCHEPERS, R., SOLDATI, T. & SOLDATI-FAVRE, D. 2007b. Dual Targeting of Antioxidant and Metabolic Enzymes to the Mitochondrion and the Apicoplast of Toxoplasma gondii. *PLoS Pathogens,* 3**,** e115.

PITZSCHKE, A. & HIRT, H. 2010. New insights into an old story: Agrobacterium-induced tumour formation in plants by plant transformation. *EMBO J,* 29**,** 1021-32.

POUTEAU, S. & ALBERTINI, C. 2009. The significance of bolting and floral transitions as indicators of reproductive phase change in Arabidopsis. *J Exp Bot,* 60**,** 3367-77.

PTITSYN, O. B. 1995. Molten globule and protein folding. *Adv Protein Chem,* 47**,** 83-229.

PURRINGTON, C. B. 2000. Costs of resistance. *Current Opinion in Plant Biology,* 3**,** 305-308.

QIAN, Y. Q., OTTING, G., FURUKUBO-TOKUNAGA, K., AFFOLTER, M., GEHRING, W. J. & WUTHRICH, K. 1992. NMR structure determination reveals that the

homeodomain is connected through a flexible linker to the main body in the Drosophila Antennapedia protein. *Proc Natl Acad Sci U S A,* 89**,** 10738-42.

RADIVOJAC, P., IAKOUCHEVA, L. M., OLDFIELD, C. J., OBRADOVIC, Z., UVERSKY, V. N. & DUNKER, A. K. 2007. Intrinsic disorder and functional proteomics. *Biophys J,* 92**,** 1439-56.

RANF, S., ESCHEN-LIPPOLD, L., PECHER, P., LEE, J. & SCHEEL, D. 2011. Interplay between calcium signalling and early signalling elements during defence responses to microbe- or damage-associated molecular patterns. *The Plant Journal,* 68**,** 100-113.

RASMUSSEN, S. G., CHOI, H. J., ROSENBAUM, D. M., KOBILKA, T. S., THIAN, F. S., EDWARDS, P. C., BURGHAMMER, M., RATNALA, V. R., SANISHVILI, R., FISCHETTI, R. F., SCHERTLER, G. F., WEIS, W. I. & KOBILKA, B. K. 2007. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature,* 450**,** 383-7.

REDMAN, A. M., DONALD, F. C., JR. & SCHULTZ, J. C. 2001. Fitness Costs of Jasmonic Acid-Induced Defense in Tomato, Lycopersicon esculentum. *Oecologia,* 126**,** 380-385.

REED, B. J., LOCKE, M. N. & GARDNER, R. G. 2015. A Conserved Deubiquitinating Enzyme Uses Intrinsically Disordered Regions to Scaffold Multiple Protein Interaction Sites. *J Biol Chem,* 290**,** 20601-12.

REVA, B., ANTIPIN, Y. & SANDER, C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res,* 39**,** e118.

RIVAS-SAN VICENTE, M. & PLASENCIA, J. 2011. *Salicylic acid beyond defence: Its role in plant growth and development*.

RIVERO, R. M., KOJIMA, M., GEPSTEIN, A., SAKAKIBARA, H., MITTLER, R., GEPSTEIN, S. & BLUMWALD, E. 2007. Delayed leaf senescence induces extreme drought tolerance in a flowering plant. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 19631-19636.

ROBBINS, J., DILWORTH, S. M., LASKEY, R. A. & DINGWALL, C. 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell,* 64**,** 615-23.

ROBERT-SEILANIANTZ, A., NAVARRO, L., BARI, R. & JONES, J. D. G. 2007. Pathological hormone imbalances. *Current Opinion in Plant Biology,* 10**,** 372-379.

ROGERS, E. E. & AUSUBEL, F. M. 1997. Arabidopsis enhanced disease susceptibility mutants exhibit enhanced susceptibility to several bacterial pathogens and alterations in PR-1 gene expression. *The Plant Cell,* 9**,** 305-316.

ROMERO, P., OBRADOVIC, Z., KISSINGER, C. R., VILLAFRANCA, J. E., GARNER, E., GUILLIOT, S. & DUNKER, A. K. 1998. Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput***,** 437-48.

ROMERO, P., OBRADOVIC, Z., LI, X., GARNER, E. C., BROWN, C. J. & DUNKER, A. K. 2001. Sequence complexity of disordered protein. *Proteins,* 42**,** 38-48.

ROMERO, P. R., ZAIDI, S., FANG, Y. Y., UVERSKY, V. N., RADIVOJAC, P., OLDFIELD, C. J., CORTESE, M. S., SICKMEIER, M., LEGALL, T., OBRADOVIC, Z. & DUNKER, A. K. 2006. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci U S A,* 103**,** 8390-5.

ROSANO, G. L. & CECCARELLI, E. A. 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology,* 5**,** 172.

ROSAS, U., MEI, Y., XIE, Q., BANTA, J. A., ZHOU, R. W., SEUFFERHELD, G., GERARD, S., CHOU, L., BHAMBHRA, N., PARKS, J. D., FLOWERS, J. M., MCCLUNG, C. R., HANZAWA, Y. & PURUGGANAN, M. D. 2014. Variation in Arabidopsis flowering time associated with cis-regulatory variation in CONSTANS. *Nat Commun,* 5.

ROY, A., KUCUKURAL, A. & ZHANG, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols,* 5**,** 725-738.

RUAL, J.-F., VENKATESAN, K., HAO, T., HIROZANE-KISHIKAWA, T., DRICOT, A., LI, N., BERRIZ, G. F., GIBBONS, F. D., DREZE, M., AYIVI-GUEDEHOUSSOU, N., KLITGORD, N., SIMON, C., BOXEM, M., MILSTEIN, S., ROSENBERG, J., GOLDBERG, D. S., ZHANG, L. V., WONG, S. L., FRANKLIN, G., LI, S., ALBALA, J. S., LIM, J., FRAUGHTON, C., LLAMOSAS, E., CEVIK, S., BEX, C., LAMESCH, P., SIKORSKI, R. S., VANDENHAUTE, J., ZOGHBI, H. Y., SMOLYAR, A., BOSAK, S., SEQUERRA, R., DOUCETTE-STAMM, L., CUSICK, M. E., HILL, D. E., ROTH, F. P. & VIDAL, M. 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature,* 437**,** 1173-1178.

RUIJTER, J. M., RAMAKERS, C., HOOGAARS, W. M. H., KARLEN, Y., BAKKER, O., VAN DEN HOFF, M. J. B. & MOORMAN, A. F. M. 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Research,* 37**,** e45-e45.

RYALS, J. A., NEUENSCHWANDER, U. H., WILLITS, M. G., MOLINA, A., STEINER, H. Y. & HUNT, M. D. 1996. Systemic Acquired Resistance. *The Plant Cell,* 8**,** 1809-1819.

RYU, H., KIM, K., CHO, H., PARK, J., CHOE, S. & HWANG, I. 2007. Nucleocytoplasmic Shuttling of BZR1 Mediated by Phosphorylation Is Essential in Arabidopsis Brassinosteroid Signaling. *The Plant Cell,* 19**,** 2749.

SAIDA, F., UZAN, M., ODAERT, B. & BONTEMS, F. 2006. Expression of highly toxic genes in E. coli: special strategies and genetic tools. *Current Protein and Peptide Science,* 7**,** 47-56.

SAKAKIBARA, H. 2006. *Cytokinins: Activity, biosynthesis, and translocation. Annu Rev Plant Biol*.

SALZMAN, R. A., TIKHONOVA, I., BORDELON, B. P., HASEGAWA, P. M. & BRESSAN, R. A. 1998. Coordinate Accumulation of Antifungal Proteins and Hexoses Constitutes a Developmentally Controlled Defense Response during Fruit Ripening in Grape. *Plant Physiology,* 117**,** 465-472.

SCHMIESING, A., EMONET, A., GOUHIER-DARIMONT, C. & REYMOND, P. 2016. Arabidopsis MYC Transcription Factors Are the Target of Hormonal Salicylic Acid/Jasmonic Acid Cross Talk in Response to Pieris brassicae Egg Extract. *Plant Physiology,* 170**,** 2432-2443.

SCHNELL, D. J. & HEBERT, D. N. 2003. Protein translocons: multifunctional mediators of protein translocation across membranes. *Cell,* 112**,** 491-505.

SHABALA, S. 2007. Transport from Root to Shoot. *Plant Solute Transport.* Blackwell Publishing Ltd.

SHAH, J., KACHROO, P. & KLESSIG, D. F. 1999. The Arabidopsis ssi1 mutation restores pathogenesis-related gene expression in npr1 plants and renders defensin gene expression salicylic acid dependent. *The Plant Cell,* 11**,** 191-206.

SHAN, L., HE, P., LI, J., HEESE, A., PECK, S. C., NÜRNBERGER, T., MARTIN, G. B. & SHEEN, J. 2008. Bacterial Effectors Target the Common Signaling Partner BAK1 to Disrupt Multiple MAMP Receptor-Signaling Complexes and Impede Plant Immunity. *Cell Host & Microbe,* 4**,** 17-27.

SHARMA, V. A., LOGAN, J., KING, D. S., WHITE, R. & ALBER, T. 1998. Sequence-based design of a peptide probe for the APC tumor suppressor protein. *Curr Biol,* 8**,** 823-30.

SHAULSKY, G., GOLDFINGER, N., BEN-ZE'EV, A. & ROTTER, V. 1990. Nuclear accumulation of p53 protein is mediated by several nuclear localization signals and plays a role in tumorigenesis. *Mol Cell Biol,* 10**,** 6565-77.

SHELUDKO, Y. V. 2008. Agrobacterium-mediated transient expression as an approach to production of recombinant proteins in plants. *Recent Pat Biotechnol,* 2**,** 198-208.

SHERMAN, F. 2005. The importance of mutation, then and now: studies with yeast cytochrome c. *Mutation Research/Reviews in Mutation Research,* 589**,** 1-16.

SHIM, J. S., JUNG, C., LEE, S., MIN, K., LEE, Y.-W., CHOI, Y., LEE, J. S., SONG, J. T., KIM, J.-K. & CHOI, Y. D. 2013. AtMYB44 regulates WRKY70 expression and modulates antagonistic interaction between salicylic acid and jasmonic acid signaling. *The Plant Journal,* 73**,** 483-495.

SHIN, H. Y. & REICH, N. C. 2013. Dynamic trafficking of STAT5 depends on an unconventional nuclear localization signal. *J Cell Sci,* 126**,** 3333-43.

SICKMEIER, M., HAMILTON, J. A., LEGALL, T., VACIC, V., CORTESE, M. S., TANTOS, A., SZABO, B., TOMPA, P., CHEN, J., UVERSKY, V. N., OBRADOVIC, Z. & DUNKER, A. K. 2007. DisProt: the Database of Disordered Proteins. *Nucleic Acids Research,* 35**,** D786-D793.

SIMMS, E. L. & RAUSHER, M. D. 1987. Costs and Benefits of Plant Resistance to Herbivory. *The American Naturalist,* 130**,** 570-581.

SLABINSKI, L., JAROSZEWSKI, L., RODRIGUES, A. P., RYCHLEWSKI, L., WILSON, I. A., LESLEY, S. A. & GODZIK, A. 2007. The challenge of protein structure determination--lessons from structural genomics. *Protein Sci,* 16**,** 2472-82.

SLOSS, C. M., CADALBERT, L., FINN, S. G., FULLER, S. J. & PLEVIN, R. 2005. Disruption of two putative nuclear localization sequences is required for cytosolic localization of mitogen-activated protein kinase phosphatase-2. *Cell Signal,* 17**,** 709-16.

SMEDEGAARD-PETERSEN, V. & TOLSTRUP, K. 1985. The Limiting Effect of Disease Resistance on Yield. *Annual Review of Phytopathology,* 23**,** 475-490.

SMIALOWSKI, P., SCHMIDT, T., COX, J., KIRSCHNER, A. & FRISHMAN, D. 2006. Will my protein crystallize? A sequence-based predictor. *Proteins,* 62**,** 343-55.

SONG, S., HUANG, H., GAO, H., WANG, J., WU, D., LIU, X., YANG, S., ZHAI, Q., LI, C., QI, T. & XIE, D. 2014. Interaction between MYC2 and ETHYLENE INSENSITIVE3

Modulates Antagonism between Jasmonate and Ethylene Signaling in Arabidopsis. *The Plant Cell*, 26**,** 263-279.

SØRENSEN, M. A., KURLAND, C. G. & PEDERSEN, S. 1989. Codon usage determines translation rate in Escherichia coli. *Journal of Molecular Biology*, 207**,** 365-377.

SPAEPEN, S. & VANDERLEYDEN, J. 2011. Auxin and Plant-Microbe Interactions. *Cold Spring Harbor Perspectives in Biology*, 3**,** a001438.

STANTON, P. 2004. Ion-exchange chromatography. *Methods Mol Biol*, 251**,** 23-44.

STELZL, U., WORM, U., LALOWSKI, M., HAENIG, C., BREMBECK, F. H., GOEHLER, H., STROEDICKE, M., ZENKNER, M., SCHOENHERR, A., KOEPPEN, S., TIMM, J., MINTZLAFF, S., ABRAHAM, C., BOCK, N., KIETZMANN, S., GOEDDE, A., TOKSOZ, E., DROEGE, A., KROBITSCH, S., KORN, B., BIRCHMEIER, W., LEHRACH, H. & WANKER, E. E. 2005. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 122**,** 957-68.

SUGIMOTO-SHIRASU, K. & ROBERTS, K. 2003. "Big it up": endoreduplication and cell-size control in plants. *Curr Opin Plant Biol*, 6**,** 544-53.

SUN, T.-P. & GUBLER, F. 2004. MOLECULAR MECHANISM OF GIBBERELLIN SIGNALING IN PLANTS. *Annual Review of Plant Biology*, 55**,** 197-223.

SUN, X., GREENWOOD, D. R., TEMPLETON, M. D., LIBICH, D. S., MCGHIE, T. K., XUE, B., YOON, M., CUI, W., KIRK, C. A., JONES, W. T., UVERSKY, V. N. & RIKKERINK, E. H. 2014. The intrinsically disordered structural platform of the plant defence hub protein RPM1-interacting protein 4 provides insights into its mode of action in the host-pathogen interface and evolution of the nitrate-induced domain protein family. *FEBS J*, 281**,** 3955-79.

SUN, X., JONES, W. T., HARVEY, D., EDWARDS, P. J., PASCAL, S. M., KIRK, C., CONSIDINE, T., SHEERIN, D. J., RAKONJAC, J., OLDFIELD, C. J., XUE, B., DUNKER, A. K. & UVERSKY, V. N. 2010a. N-terminal domains of DELLA proteins are intrinsically unstructured in the absence of interaction with GID1/gibberellic acid receptors. *J Biol Chem*, 285**,** 11557-71.

SUN, Y., FAN, X.-Y., CAO, D.-M., TANG, W., HE, K., ZHU, J.-Y., HE, J.-X., BAI, M.-Y., ZHU, S., OH, E., PATIL, S., KIM, T.-W., JI, H., WONG, W. H., RHEE, S. Y. & WANG, Z.-Y. 2010b. Integration of Brassinosteroid Signal Transduction with the Transcription Network for Plant Growth Regulation in Arabidopsis. *Developmental Cell*, 19**,** 765-777.

SUN, Y., HAN, Z., TANG, J., HU, Z., CHAI, C., ZHOU, B. & CHAI, J. 2013a. Structure reveals that BAK1 as a co-receptor recognizes the BRI1-bound brassinolide. *Cell Res*, 23**,** 1326-1329.

SUN, Y., LI, L., MACHO, A. P., HAN, Z., HU, Z., ZIPFEL, C., ZHOU, J.-M. & CHAI, J. 2013b. Structural Basis for flg22-Induced Activation of the Arabidopsis FLS2-BAK1 Immune Complex. *Science*, 342**,** 624.

SUNDERLAND, P. A., WEST, C. E., WATERWORTH, W. M. & BRAY, C. M. 2004. Choice of a start codon in a single transcript determines DNA ligase 1 isoform production and intracellular targeting in Arabidopsis thaliana. *Biochem Soc Trans*, 32**,** 614-6.

SUNDERLAND, P. A., WEST, C. E., WATERWORTH, W. M. & BRAY, C. M. 2006. An evolutionarily conserved translation initiation mechanism regulates nuclear or mitochondrial targeting of DNA ligase 1 in Arabidopsis thaliana. *Plant J,* 47**,** 356-67.

TAKAGI, M., ABSALON, M. J., MCLURE, K. G. & KASTAN, M. B. 2005. Regulation of p53 Translation and Induction after DNA Damage by Ribosomal Protein L26 and Nucleolin. *Cell,* 123**,** 49-63.

TAKEI, Y., YAMAMOTO, K. & TSUJIMOTO, G. 1999. Identification of the sequence responsible for the nuclear localization of human Cdc6. *FEBS Letters,* 447**,** 292-296.

TANAKA, K., ASAMI, T., YOSHIDA, S., NAKAMURA, Y., MATSUO, T. & OKAMOTO, S. 2005. Brassinosteroid Homeostasis in Arabidopsis Is Ensured by Feedback Expressions of Multiple Genes Involved in Its Metabolism. *Plant Physiology,* 138**,** 1117-1125.

TANAKA, S., MIURA, I., YOSHIKI, A., KATO, Y., YOKOYAMA, H., SHINOGI, A., MASUYA, H., WAKANA, S., TAMURA, M. & SHIROISHI, T. 2007. Mutations in the helix termination motif of mouse type I IRS keratin genes impair the assembly of keratin intermediate filament. *Genomics,* 90**,** 703-11.

TANG, D., CHRISTIANSEN, K. M. & INNES, R. W. 2005. Regulation of Plant Disease Resistance, Stress Responses, Cell Death, and Ethylene Signaling in Arabidopsis by the EDR1 Protein Kinase. *Plant Physiology,* 138**,** 1018-1026.

TANG, W., YUAN, M., WANG, R., YANG, Y., WANG, C., OSES-PRIETO, J. A., KIM, T.-W., ZHOU, H.-W., DENG, Z., GAMPALA, S. S., GENDRON, J. M., JONASSEN, E. M., LILLO, C., DELONG, A., BURLINGAME, A. L., SUN, Y. & WANG, Z.-Y. 2011. PP2A activates brassinosteroid-responsive gene expression and plant growth by dephosphorylating BZR1. *Nat Cell Biol,* 13**,** 124-131.

THEILLET, F.-X., KALMAR, L., TOMPA, P., HAN, K.-H., SELENKO, P., DUNKER, A. K., DAUGHDRILL, G. W. & UVERSKY, V. N. 2013. The alphabet of intrinsic disorder. *Intrinsically Disordered Proteins,* 1**,** e24360.

THOMAS, S. G. & FRANKLIN-TONG, V. E. 2004. Self-incompatibility triggers programmed cell death in Papaver pollen. *Nature,* 429**,** 305-309.

TOMPA, P. 2012. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci,* 37**,** 509-16.

TOMPA, P. & CSERMELY, P. 2004. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J,* 18**,** 1169-75.

TOMPA, P., SZASZ, C. & BUDAY, L. 2005. Structural disorder throws new light on moonlighting. *Trends Biochem Sci,* 30**,** 484-9.

TORNROTH-HORSEFIELD, S., WANG, Y., HEDFALK, K., JOHANSON, U., KARLSSON, M., TAJKHORSHID, E., NEUTZE, R. & KJELLBOM, P. 2006. Structural mechanism of plant aquaporin gating. *Nature,* 439**,** 688-94.

TRIPATHY, B. C. & OELMÜLLER, R. 2012. Reactive oxygen species generation and signaling in plants. *Plant Signaling & Behavior,* 7**,** 1621-1633.

TROMBITAS, K., GREASER, M., LABEIT, S., JIN, J. P., KELLERMAYER, M., HELMES, M. & GRANZIER, H. 1998. Titin extensibility in situ: entropic elasticity of permanently folded and permanently unfolded molecular segments. *J Cell Biol,* 140**,** 853-9.

TRUEBESTEIN, L. & LEONARD, T. A. 2016. Coiled-coils: The long and short of it. *Bioessays,* 38**,** 903-916.
TSUDA, K., QI, Y., NGUYEN LE, V., BETHKE, G., TSUDA, Y., GLAZEBROOK, J. & KATAGIRI, F. 2012. An efficient Agrobacterium-mediated transient transformation of Arabidopsis. *Plant J,* 69**,** 713-9.
UEDA, J. & KATO, J. 1980. Isolation and Identification of a Senescence-promoting Substance from Wormwood (Artemisia absinthium L.). *Plant Physiology,* 66**,** 246-249.
UVERSKY, V. N. & DUNKER, A. K. 2010. Understanding Protein Non-Folding. *Biochimica et biophysica acta,* 1804**,** 1231-1264.
UVERSKY, V. N., GILLESPIE, J. R. & FINK, A. L. 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins,* 41**,** 415-27.
UVERSKY, V. N., NARIZHNEVA, N. V., IVANOVA, T. V., KIRKITADZE, M. D. & TOMASHEVSKI, A. 1997. Ligand-free form of human alpha-fetoprotein: evidence for the molten globule state. *FEBS Lett,* 410**,** 280-4.
UVERSKY, V. N., OLDFIELD, C. J. & DUNKER, A. K. 2005. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit,* 18**,** 343-84.
UVERSKY, V. N. & PTITSYN, O. B. 1996. Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *J Mol Biol,* 255**,** 215-28.
VACIC, V., OLDFIELD, C. J., MOHAN, A., RADIVOJAC, P., CORTESE, M. S., UVERSKY, V. N. & DUNKER, A. K. 2007. Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res,* 6**,** 2351-66.
VAN BREUSEGEM, F. & DAT, J. F. 2006. Reactive Oxygen Species in Plant Cell Death. *Plant Physiology,* 141**,** 384-390.
VAN DER LEE, R., BULJAN, M., LANG, B., WEATHERITT, R. J., DAUGHDRILL, G. W., DUNKER, A. K., FUXREITER, M., GOUGH, J., GSPONER, J., JONES, D. T., KIM, P. M., KRIWACKI, R. W., OLDFIELD, C. J., PAPPU, R. V., TOMPA, P., UVERSKY, V. N., WRIGHT, P. E. & BABU, M. M. 2014. Classification of intrinsically disordered regions and proteins. *Chem Rev,* 114**,** 6589-631.
VAN DOORN, W. G. 2011. Classes of programmed cell death in plants, compared to those in animals. *Journal of Experimental Botany,* 62**,** 4749-4761.
VAN LOOCK, B., MARKAKIS, M. N., VERBELEN, J. P. & VISSENBERG, K. 2010. High-throughput transient transformation of Arabidopsis roots enables systematic colocalization analysis of GFP-tagged proteins. *Plant Signal Behav,* 5**,** 261-3.
VAN ROEY, K., GIBSON, T. J. & DAVEY, N. E. 2012. Motif switches: decision-making in cell regulation. *Curr Opin Struct Biol,* 22**,** 378-85.
VAN DER LEE, R., LANG, B., KRUSE, K., GSPONER, J., SÁNCHEZ DE GROOT, N., HUYNEN, MARTIJN A., MATOUSCHEK, A., FUXREITER, M. & BABU, M. M. 2014. Intrinsically Disordered Segments Affect Protein Half-Life in the Cell and during Evolution. *Cell Reports,* 8**,** 1832-1844.
VANACKER, H., SANDALIO, L. M., JIMÉNEZ, A., PALMA, J. M., CORPAS, F. J., MESEGUER, V., GÓMEZ, M., SEVILLA, F., LETERRIER, M., FOYER, C. H. & DEL RÍO, L. A. 2006. Roles for redox regulation in leaf senescence of pea plants grown

on different sources of nitrogen nutrition. *Journal of Experimental Botany,* 57**,** 1735-1745.

VANDEPUTTE, O., ÖDEN, S., MOL, A., VEREECKE, D., GOETHALS, K., EL JAZIRI, M. & PRINSEN, E. 2005. Biosynthesis of Auxin by the Gram-Positive Phytopathogen Rhodococcus fascians Is Controlled by Compounds Specific to Infected Plant Tissues. *Applied and Environmental Microbiology,* 71**,** 1169-1177.

VANSTRAELEN, M., BALOBAN, M., DA INES, O., CULTRONE, A., LAMMENS, T., BOUDOLF, V., BROWN, S. C., DE VEYLDER, L., MERGAERT, P. & KONDOROSI, E. 2009. APC/C(CCS52A) complexes control meristem maintenance in the Arabidopsis root. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 11806-11811.

VARENNE, S., BUC, J., LLOUBES, R. & LAZDUNSKI, C. 1984. Translation is a non-uniform process. *Journal of Molecular Biology,* 180**,** 549-576.

VECINO, A. J., SEGURA, R. D. L., DE LA ARADA, I., DE LA CRUZ, F., GOÑI, F. M., ARRONDO, J. L. & ALKORTA, I. 2012. Deletion of a single helix from the transmembrane domain causes large changes in membrane insertion properties and secondary structure of the bacterial conjugation protein TrwB. *Biochimica et Biophysica Acta (BBA) - Biomembranes,* 1818**,** 3158-3166.

VENCO, P., BONORA, M., GIORGI, C., PAPALEO, E., IUSO, A., PROKISCH, H., PINTON, P. & TIRANTI, V. 2015. Mutations of C19orf12, coding for a transmembrane glycine zipper containing mitochondrial protein, cause mis-localization of the protein, inability to respond to oxidative stress and increased mitochondrial Ca(2+). *Frontiers in Genetics,* 6**,** 185.

VERT, G. & CHORY, J. 2006. Downstream nuclear events in brassinosteroid signalling. *Nature,* 441**,** 96-100.

VINSON, C. R., SIGLER, P. B. & MCKNIGHT, S. L. 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science,* 246**,** 911-6.

VUCETIC, S., BROWN, C. J., DUNKER, A. K. & OBRADOVIC, Z. 2003. Flavors of protein disorder. *Proteins,* 52**,** 573-84.

WALLS, D., MCGRATH, R. & LOUGHRAN, S. T. 2011. A digest of protein purification. *Methods Mol Biol,* 681**,** 3-23.

WALSH, G. 2003. *The drug manufacturing proces,* John Wiley, L. Sons (Eds.), .

WALTER, J. & GOTTSCHALK, U. 2010. Concepts for Disposables in Biopharmaceutical Manufacture. *In:* SHIRE, S. J., GOMBOTZ, W., BECHTOLD-PETERS, K. & ANDYA, J. (eds.) *Current Trends in Monoclonal Antibody Development and Manufacturing.* Springer New York.

WALTERS, D. & HEIL, M. 2007. Costs and trade-offs associated with induced resistance. *Physiological and Molecular Plant Pathology,* 71**,** 3-17.

WAMBOLDT, Y., MOHAMMED, S., ELOWSKY, C., WITTGREN, C., DE PAULA, W. B. & MACKENZIE, S. A. 2009. Participation of leaky ribosome scanning in protein dual targeting by alternative translation initiation in higher plants. *Plant Cell,* 21**,** 157-67.

WANG, D., PAJEROWSKA-MUKHTAR, K., CULLER, A. H. & DONG, X. 2007. Salicylic Acid Inhibits Pathogen Growth in Plants through Repression of the Auxin Signaling Pathway. *Current Biology,* 17**,** 1784-1790.

WANG, L., CHU, H., LI, Z., WANG, J., LI, J., QIAO, Y., FU, Y., MOU, T., CHEN, C. & XU, J. 2014a. Origin and Development of the Root Cap in Rice. *Plant Physiology,* 166**,** 603.
WANG, S., GU, Y., ZEBELL, S. G., ANDERSON, L. K., WANG, W., MOHAN, R. & DONG, X. 2014b. A noncanonical role for the CKI-RB-E2F cell-cycle signaling pathway in plant effector-triggered immunity. *Cell Host Microbe,* 16**,** 787-94.
WANG, Z.-Y. 2012. Brassinosteroids modulate plant immunity at multiple levels. *Proceedings of the National Academy of Sciences,* 109**,** 7-8.
WANG, Z.-Y., NAKANO, T., GENDRON, J., HE, J., CHEN, M., VAFEADOS, D., YANG, Y., FUJIOKA, S., YOSHIDA, S., ASAMI, T. & CHORY, J. 2002. Nuclear-Localized BZR1 Mediates Brassinosteroid-Induced Growth and Feedback Suppression of Brassinosteroid Biosynthesis. *Developmental Cell,* 2**,** 505-513.
WAR, A. R., PAULRAJ, M. G., AHMAD, T., BUHROO, A. A., HUSSAIN, B., IGNACIMUTHU, S. & SHARMA, H. C. 2012. Mechanisms of plant defense against insect herbivores. *Plant Signaling & Behavior,* 7**,** 1306-1320.
WARD, J. J., SODHI, J. S., MCGUFFIN, L. J., BUXTON, B. F. & JONES, D. T. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol,* 337**,** 635-45.
WARNE, T., SERRANO-VEGA, M. J., BAKER, J. G., MOUKHAMETZIANOV, R., EDWARDS, P. C., HENDERSON, R., LESLIE, A. G. W., TATE, C. G. & SCHERTLER, G. F. X. 2008. Structure of a β(1)-adrenergic G protein-coupled receptor. *Nature,* 454**,** 486-491.
WASTERNACK, C. & HAUSE, B. 2013. Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Annals of Botany,* 111**,** 1021-1058.
WAUGH, D. S. 2005. Making the most of affinity tags. *Trends Biotechnol,* 23**,** 316-20.
WEATHERITT, R. J., DAVEY, N. E. & GIBSON, T. J. 2012. Linear motifs confer functional diversity onto splice variants. *Nucleic Acids Res,* 40**,** 7123-31.
WEATHERITT, R. J. & GIBSON, T. J. 2012. Linear motifs: lost in (pre)translation. *Trends Biochem Sci,* 37**,** 333-41.
WEIGEL, D. & GLAZEBROOK, J. 2002. *Arabidopsis: A Laboratory Manual* Cold Spring Harbor Laboratory Press. .
WEIGEL, D. & GLAZEBROOK, J. 2006. Transformation of Agrobacterium Using Electroporation. *Cold Spring Harbor Protocols,* 2006**,** pdb.prot4665.
WEIS, K. 2003. Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell,* 112**,** 441-51.
WHITE, S. H. & VON HEIJNE, G. 2008. How Translocons Select Transmembrane Helices. *Annual Review of Biophysics,* 37**,** 23-42.
WILKINS, K. A., BOSCH, M., HAQUE, T., TENG, N., POULTER, N. S. & FRANKLIN-TONG, V. E. 2015. Self-Incompatibility-induced Programmed Cell Death in Papaver pollen involves dramatic acidification of the incompatible pollen tube cytosol. *Plant Physiology*.
WILLIAMS, B. & DICKMAN, M. 2008. Plant programmed cell death: can't live with it; can't live without it. *Molecular Plant Pathology,* 9**,** 531-544.

WINGLER, A., VON SCHAEWEN, A., LEEGOOD, R. C., LEA, P. J. & PAUL QUICK, W. 1998. Regulation of Leaf Senescence by Cytokinin, Sugars, and Light : Effects on NADH-Dependent Hydroxypyruvate Reductase. *Plant Physiology,* 116**,** 329-335.

WISE, A. A., LIU, Z. & BINNS, A. N. 2006. Three Methods for the Introduction of Foreign DNA into Agrobacterium. *In:* WANG, K. (ed.) *Agrobacterium Protocols.* Totowa, NJ: Humana Press.

WOMBACHER, R. & CORNISH, V. W. 2011. Chemical tags: applications in live cell fluorescence imaging. *J Biophotonics,* 4**,** 391-402.

WOOLFSON, D. N. 2005. The design of coiled-coil structures and assemblies. *Adv Protein Chem,* 70**,** 79-112.

WRIGHT, P. E. & DYSON, H. J. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol,* 293**,** 321-31.

WROBLEWSKI, T., TOMCZAK, A. & MICHELMORE, R. 2005. Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. *Plant Biotechnol J,* 3**,** 259-73.

WU, A., ALLU, A. D., GARAPATI, P., SIDDIQUI, H., DORTAY, H., ZANOR, M.-I., ASENSI-FABADO, M. A., MUNNÉ-BOSCH, S., ANTONIO, C., TOHGE, T., FERNIE, A. R., KAUFMANN, K., XUE, G.-P., MUELLER-ROEBER, B. & BALAZADEH, S. 2012. JUNGBRUNNEN1, a Reactive Oxygen Species–Responsive NAC Transcription Factor, Regulates Longevity in Arabidopsis. *The Plant Cell,* 24**,** 482-506.

XIA, Y., YU, K., NAVARRE, D., SEEBOLD, K., KACHROO, A. & KACHROO, P. 2010. The glabra1 Mutation Affects Cuticle Formation and Plant Responses to Microbes. *Plant Physiology,* 154**,** 833-846.

XIAO, Y.-P., WU, T.-X., HONG, Q.-H., SUN, J.-M., CHEN, A.-G., YANG, C.-M. & LI, X.-Y. 2012. Response to weaning and dietary L-glutamine supplementation: metabolomic analysis in piglets by gas chromatography/mass spectrometry. *Journal of Zhejiang University. Science. B,* 13**,** 567-578.

XIAO, Z., LIU, X., HENIS, Y. I. & LODISH, H. F. 2000. A distinct nuclear localization signal in the N terminus of Smad 3 determines its ligand-induced nuclear translocation. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 7853-7858.

XIE, H., VUCETIC, S., IAKOUCHEVA, L. M., OLDFIELD, C. J., DUNKER, A. K., OBRADOVIC, Z. & UVERSKY, V. N. 2007. Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res,* 6**,** 1917-32.

XIE, Q., ARNOLD, G. E., ROMERO, P., OBRADOVIC, Z., GARNER, E. & DUNKER, A. K. 1998. The Sequence Attribute Method for Determining Relationships Between Sequence and Protein Disorder. *Genome Inform Ser Workshop Genome Inform,* 9**,** 193-200.

YAMADA, T., ICHIMURA, K. & VAN DOORN, W. G. 2007. Relationship between petal abscission and programmed cell death in Prunus yedoensis and Delphinium belladonna. *Planta,* 226**,** 1195-1205.

YANG, C.-J., ZHANG, C., LU, Y.-N., JIN, J.-Q. & WANG, X.-L. 2011. The Mechanisms of Brassinosteroids' Action: From Signal Transduction to Plant Development. *Molecular Plant,* 4**,** 588-600.

YANG, D.-L., YANG, Y. & HE, Z. 2013. Roles of Plant Hormones and Their Interplay in Rice Immunity. *Molecular Plant,* 6**,** 675-685.

YANG, D.-L., YAO, J., MEI, C.-S., TONG, X.-H., ZENG, L.-J., LI, Q., XIAO, L.-T., SUN, T.-P., LI, J., DENG, X.-W., LEE, C. M., THOMASHOW, M. F., YANG, Y., HE, Z. & HE, S. Y. 2012. Plant hormone jasmonate prioritizes defense over growth by interfering with gibberellin signaling cascade. *Proceedings of the National Academy of Sciences,* 109**,** E1192–E1200.

YEGAMBARAM, K., BULLOCH, E. M. M. & KINGSTON, R. L. 2013. Protein domain definition should allow for conditional disorder. *Protein Science,* 22**,** 1502-1518.

YEUNG, E. C. & MEINKE, D. W. 1993. Embryogenesis in Angiosperms: Development of the Suspensor. *The Plant Cell,* 5**,** 1371-1381.

YOSHIDA, S., ITO, M., NISHIDA, I. & WATANABE, A. 2001. Isolation and RNA gel blot analysis of genes that could serve as potential molecular markers for leaf senescence in Arabidopsis thaliana. *Plant Cell Physiol,* 42**,** 170-8.

YOSHIDA, S., ITO, M., NISHIDA, I. & WATANABE, A. 2002. Identification of a novel gene HYS1/CPR5 that has a repressive role in the induction of leaf senescence and pathogen-defence responses in Arabidopsis thaliana. *Plant J,* 29**,** 427-37.

YOUNG, C. L., BRITTON, Z. T. & ROBINSON, A. S. 2012. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnology Journal,* 7**,** 620-634.

YU, X., LI, L., ZOLA, J., ALURU, M., YE, H., FOUDREE, A., GUO, H., ANDERSON, S., ALURU, S., LIU, P., RODERMEL, S. & YIN, Y. 2011. A brassinosteroid transcriptional network revealed by genome-wide identification of BESI target genes in Arabidopsis thaliana. *The Plant Journal,* 65**,** 634-646.

YUAN, Z. C., EDLIND, M. P., LIU, P., SAENKHAM, P., BANTA, L. M., WISE, A. A., RONZONE, E., BINNS, A. N., KERR, K. & NESTER, E. W. 2007. The plant signal salicylic acid shuts down expression of the vir regulon and activates quormone-quenching genes in Agrobacterium. *Proc Natl Acad Sci U S A,* 104**,** 11790-5.

ZASLOFF, M. 1983. tRNA transport from the nucleus in a eukaryotic cell: carrier-mediated translocation process. *Proc Natl Acad Sci U S A,* 80**,** 6436-40.

ZHANG, C., ZHAO, X. & ZHANG, M. Q. 2007. Functional in Silico Analysis of Gene Regulatory Polymorphism. *Bioinformatics for Geneticists.* John Wiley & Sons, Ltd.

ZHANG, J., BILLINGSLEY, M. L., KINCAID, R. L. & SIRAGANIAN, R. P. 2000. Phosphorylation of Syk activation loop tyrosines is essential for Syk function. An in vivo study using a specific anti-Syk activation loop phosphotyrosine antibody. *J Biol Chem,* 275**,** 35442-7.

ZHANG, K., HALITSCHKE, R., YIN, C., LIU, C.-J. & GAN, S.-S. 2013. Salicylic acid 3-hydroxylase regulates Arabidopsis leaf longevity by mediating salicylic acid catabolism. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 14807-14812.

ZHANG, W., REN, G., WU, B., LIU, X., WANG, G., SONG, Y., XU, H., WEN, H. & WANG, Z. 2008. Amino acid substitutions in leucine zipper motif in the F-specific domain of human parainfluenza virus 3 HN protein play important roles in the protein function. *Intervirology,* 51**,** 311-21.

ZHANG, X., HENRIQUES, R., LIN, S. S., NIU, Q. W. & CHUA, N. H. 2006. Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nat Protoc,* 1**,** 641-6.

ZHAO, B., LI, L., TUMANENG, K., WANG, C. Y. & GUAN, K. L. 2010a. A coordinated phosphorylation by Lats and CK1 regulates YAP stability through SCF(beta-TRCP). *Genes Dev,* 24**,** 72-85.

ZHAO, J., PENG, P., SCHMITZ, R. J., DECKER, A. D., TAX, F. E. & LI, J. 2002. Two Putative BIN2 Substrates Are Nuclear Components of Brassinosteroid Signaling. *Plant Physiology,* 130**,** 1221-1229.

ZHAO, R., GISH, K., MURPHY, M., YIN, Y., NOTTERMAN, D., HOFFMAN, W. H., TOM, E., MACK, D. H. & LEVINE, A. J. 2000. Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. *Genes Dev,* 14**,** 981-93.

ZHAO, X., LI, G. & LIANG, S. 2013. Several affinity tags commonly used in chromatographic purification. *J Anal Methods Chem,* 2013**,** 581093.

ZHAO, Z., ANDERSEN, S. U., LJUNG, K., DOLEZAL, K., MIOTK, A., SCHULTHEISS, S. J. & LOHMANN, J. U. 2010b. Hormonal control of the shoot stem-cell niche. *Nature,* 465**,** 1089-1092.

ZHOU, N. E., KAY, C. M. & HODGES, R. S. 1992. Synthetic model proteins. Positional effects of interchain hydrophobic interactions on stability of two-stranded alpha-helical coiled-coils. *J Biol Chem,* 267**,** 2664-70.

ZHOU, X.-Y., SONG, L. & XUE, H.-W. 2013. Brassinosteroids Regulate the Differential Growth of Arabidopsis Hypocotyls through Auxin Signaling Components IAA19 and ARF7. *Molecular Plant,* 6**,** 887-904.

ZHU, S., JEONG, R.-D., VENUGOPAL, S. C., LAPCHYK, L., NAVARRE, D., KACHROO, A. & KACHROO, P. 2011. SAG101 Forms a Ternary Complex with EDS1 and PAD4 and Is Required for Resistance Signaling against Turnip Crinkle Virus. *PLoS Pathogens,* 7**,** e1002318.

ZHU, Y., NAM, J., CARPITA, N. C., MATTHYSSE, A. G. & GELVIN, S. B. 2003. Agrobacterium-mediated root transformation is inhibited by mutation of an Arabidopsis cellulose synthase-like gene. *Plant Physiol,* 133**,** 1000-10.

ZIMMERMANN, P., HEINLEIN, C., ORENDI, G. & ZENTGRAF, U. 2006. Senescence-specific regulation of catalases in Arabidopsis thaliana (L.) Heynh. *Plant, Cell & Environment,* 29**,** 1049-1060.

ZIPFEL, C., KUNZE, G., CHINCHILLA, D., CANIARD, A., JONES, J. D., BOLLER, T. & FELIX, G. 2006. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell,* 125**,** 749-60.