The Role of Cdc48 Adapter Proteins Npl4 & Ufd1 in Chloroplast Associated Protein Degradation

Simon Mark Thomson

Keble College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy in

Interdisciplinary Biosciences

Trinity Term 2022

The Role of Cdc48 Adapter Proteins Npl4 & Ufd1 in Chloroplast Associated Protein Degradation

Simon Mark Thomson

Keble College

DPhil in Interdisciplinary Biosciences

Trinity 2022

Abstract

The vast bulk of the chloroplast proteome is imported post-translationally through protein translocation complexes called TOCs (translocon at the outer envelope of chloroplasts). The regulation of such an import system offers a key mechanism to impact chloroplast metabolism and plastid development, both through the modulation of the proteome. In response to biotic, abiotic, and developmental cues, the TOC proteins are targeted by the ubiquitin-proteasome system for degradation. Membrane-embedded TOC proteins must first be extracted from the outer membrane before degradation by the cytosolic UPS. This is accomplished by retrotranslocation through a protein channel, with the motive force provided by an ATPase called CDC48 — this pathway has been termed CHLORAD (chloroplast associated protein degradation). CDC48 functions in a variety of cellular processes at different subcellular locations. The localisation and activity of CDC48 is determined by a wide range of adapter proteins. In UPS processes, CDC48 is typically recruited by a heterodimer of NPL4 and UFD1. There are two NPL4, and four UFD1 isoforms in the model higher plant *Arabidopsis thaliana*. To examine their potential role in CHLORAD, I performed bioinformatic analysis on protein sequences and predicted structures; interaction and localisation experiments, using microscopy and immunoprecipitation; and finally conducted a mutant analysis of knockout and overexpressing lines, looking at developmental and stress responses. Three UFD1 and two NPL4 proteins can participate within the CHLORAD system by recruitment of CDC48. While the proteins are nucleocytosolic, interactions with TOC proteins were sustained at the

chloroplast outer envelope membrane. There appears to be a degree of functional redundancy between sets of the isoforms; this is supported by phylogenetic reconstruction showing isolated clades in the *Arabidopsis* family. Few phenotypes were observed in the analysis of single mutants. One significant one on TOC protein abundance may be the product of a dominant negative mutation generated through a translational fusion with an epitope tag.

The functional significance of NPL4 and UFD1 in CHLORAD requires much additional consideration in the future, with particular emphasis on the analysis of additional genetic lines. Phenotypes may be masked by functional redundancy between NPL4 and UFD1 isoforms, so the investigation of appropriate multiple mutants would be highly informative. Additionally, analysis of crosses to plastid protein import mutants would determine if there are suppressive genetic interactions, which would support a function in CHLORAD. Additional focus should be placed on the generation of domain deletion mutants to validate bioinformatic predictions of function — Ufd1C should be the prime candidate in this regard, to assess how interactions are sustained with a remarkably unique predicted structure. Finally, consideration should also be given to the potential role in biotic stress tolerance, as both CHLORAD and CDC48-NPL4-UFD1 have independently understood contributions to the immune system.

Contents

Acknowledgements

I must thank the Oxford Interdisciplinary Bioscience Doctoral Training Partnership (DTP) for providing training and guidance throughout my studies. In particular, I would like to thank Professor Gail Preston from the DTP for her kindness and support. In turn, I must also thank the Biotechnology and Biological Sciences Research Council (BBSRC) for their financial support of the DTP programme and my work (grant number BB/M011224/1). I would also like to thank Keble College for their welfare support.

I would also like to offer my sincerest thanks to my supervisor Professor Paul Jarvis for his years of advice, guidance, and support. I am grateful to all members of the Jarvis group for ensuring that our laboratory ran functionally. I would like to offer particular special thanks to Na Li who has acted as my mentor in the laboratory since the commencement of this project; she has always been readily available to offer technical guidance and support, and I am hugely indebted to her for all she has taught me. I would like to express my gratitude for having known the following Department members, and for all their kindness and advice: Nick Kruger, Qihua Ling, Jun Fang, Pedro Bota, Beatriz Moreno-Garcia, Sreehar Nellaepalli, Sabri Ali, Robert Sowden, Jonathon Bester, Pablo Pulido, Gem Toes-Crichton, Sam Watson, Nattaphorn Buayam (Aom), and Anne Lau.

This work would not have been possible without the unwavering support and encouragement from my family. I also offer special thanks to my partner Lee-Anne, for all her encouragement and emotional support over the years.

Statement of COVID-19 Impact

This project was heavily influenced by the SARS-CoV-2 (COVID-19) pandemic. Following the outbreak of COVID-19 globally, the UK government imposed its first lockdown in March of 2020. The Department of Plant Sciences (abbreviated henceforth to 'the Department') was subsequently closed for a period of months. During this time, access was only granted to individuals to continue the propagation of existing genetic material. This was determined by a booking rota for people to enter the Department several times a week, to allow people to water and manage all users' plants. It later emerged that during this period my own lines had been mistakenly discarded by another individual, after I discovered them with other trays of discarded plants. This resulted in the permanent loss of many T1 and T2 transgenic lines, as well several crossed F0 plants.

Following the initial lockdown, the Department then limited building access for a period of approximately one year. This period included subsequent UK-wide lockdowns. Access was managed by a first come, first serve booking rota. The rota implemented a maximum user capacity, well below the number of active users, to allow for social distancing and account for which individuals were in close contact. This limited the possible number of working hours in the laboratory, as well as when certain types of experiments could be performed. The pandemic also affected scientific industry, creating delays in logistics and manufacturing. Chemicals, reagents, and molecular biology kits were all affected, resulting in slower delivery times and limited availability up until the end of the project.

While an extension of funding was granted covering a period of 6 months, the net result was still a considerable reduction in the time available to complete this project. It also reduced the potential for collaboration (e.g. with the proteomics facility), and limited opportunities to learn or improve on techniques by demonstration. The pandemic also took a personal toll on my health, resulting in the suspension of one term. To mitigate this impact, the plan of the project was adjusted, to reduce some aspects of experimental work and increase the content of computational work. However, this shift was not entirely compensatory. The completed project is thus limited in its the ability to address the original aims, which fundamentally relied on experimentation.

Statement of Authorship

All of the experiments, and all of the materials relating to Npl4 and Ufd1, were prepared and performed by the sole author. Much of Chapter I is modified from my own publication (Thomson *et al.*, 2020).

Materials not produced by the author were antibodies, plasmid constructs and genetic lines corresponding to genes other than Npl4 and Ufd1. These materials were generated previously by members of the Jarvis group, donated to the group by collaborators, or purchased from commercial suppliers.

Abbreviations

26SP, 26S proteasome

AAA+, ATPases associated with various cellular activities

AGE, agarose gel electrophoresis

APS, ammonium persulphate

At, *Arabidopsis thaliana*

ATP, adenosine triphosphate

AtpB, ATP synthase subunit beta

BiFC, bimolecular fluorescence complementation

BLAST, basic local alignment search tool

BME, β-mercaptoethanol

BSA, bovine serum albumin

CAB3, Chlorophyll a/b binding protein 3

Cdc48, cell division cycle 48

cDNA, complementary DNA

CDS, coding sequences

CHIP, C-terminus of Hsc70-interacting protein

CHLORAD, chloroplast associated protein degradation

Clp, caseinolytic protease

Co-IP, co-immunoprecipitation

Col-0, Columbia-0 ecotype

CRISPR, clustered regularly interspaced palindromic repeats

CRL, cullin-RING ligase

Cryo-EM, cyrogenic electron microscopy

CTAB, hexadecyltrimethylammonium bromide

Cue1, coupling of ubiquitin conjugation to ER degradation 1

Cue8, CAB3 under expressed 8

Deg, degradation of periplasmic proteins

DEK5, defective kernel5

DMSO, dimethyl sulfoxide

DNA, Deoxyribonucleic acid

dNTPs, deoxyribonucleoside triphosphates

DTT, dithiothreitol

E1, ubiquitin-activating enzyme

E2, ubiquitin-conjugating enzyme

E3, ubiquitin ligase

EDTA (ethylenedinitrilo)tetraacetic acid

ER, endoplasmic reticulum

ERAD, ER-associated protein degradation

fc2, ferrochelatase 2

FtsH, filamentation temperature-sensitive H

$F_{x,}$ filial generation x

G, g-force

GFP, Green fluorescent protein

Glk, Golden2-like

GO, gene ontology

gRNA, guide RNA

GTP, guanosine triphosphate

GUN1, genomes uncoupled1

HECT, homologous to the E6-AP carboxyl terminus

HIP, Hsp70 interacting protein

Hop, Hsp70/Hsp90-organising protein

Hrd1, HMG-CoA reductase degradation 1

IEM, the inner envelope membrane

IMS, intermembrane space

kDa, kilodalton

KO, knock-out

KOC1, regulatory kinase at the outer chloroplast membrane 1

LB, lysogeny broth

MAD, mitochondria-associated protein degradation

MAFFT, multiple alignment using fast Fourier transform

MES, 2-(*N*-morpholino) ethanesulfonic acid

MPN, Mpr1/Pad1 N-terminal

MS, Murashige and Skoog

MUB, membrane anchored ubiquitin fold

MV, methyl viologen

NASC, Nottingham Arabidopsis Stock Centre

NBM, Npl4 binding motif

NEB, New England Biolabs

Npl4, nuclear protein localisation 4

NU, Npl4-Ufd1 heterodimer

NVL2, nuclear VCP-Like 2

OD, optical density

OEM, the outer envelope membrane

Oep7, outer envelope protein 7

Omp, Outer-membrane protein

Otu1, ovarian tumour 1

OX, overexpression

p35s, CamV35s promoter

PAR, photosynthetically active radiation

PCR, polymerase chain reaction

PEG, polyethylene glycol

PI4Ks, phosphoinositide4-kinase

POTRA, polypeptide transport associated

Ppi, plastid protein import

ppo1-1, protoporphyrinogen IX oxidase 1

Pre-proteins, precursor proteins

PSI-BLAST, position-specific iterative basic local alignment search tool

PTMs, post-translational modifications

PUB, PNGase/UBA or UBX containing proteins

PUL, PLAP, Ufd3p, Lub1p

Pux, plant UBX

RbcL, Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit

RHA1A, RING-H2 FINGER A1A

RING, really interesting new gene

RMSD, root mean square deviation of atomic position

RNA, Ribonucleic acid

ROS, reactive oxygen species

RPN10, regulatory particle non-ATPase 10

RT, Reverse Transcription

SAP5, stress-associated protein 5

SDS, sodium dodecyl sulphate

SDS-PAGE, sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SHP box, binding site 1

Soh1, suppressor of *tic100* 1

SP, suppressor of *ppi1* locus

SPL, SP1-Like

sqRT-PCR, Semi-quantitative RT-PCR

Tam, translocation and assembly module

TAP, tandem affinity purification

TBST-T, Tris-buffered-saline-Tween

TEMED, tetramethylethylenediamine

TGS, Tris-Glycine-SDS running buffer

TIC, translocon at the inner envelope of chloroplasts

$T_{n,}$ Generation of transgenic plants

TOC, translocon at the outer envelope of chloroplasts

TP, transit peptide

TPR, tetratricopeptide repeat

TRAFD1, tumour necrosis factor receptor associated factor domain containing 1

Tris-HCl, Tris(hydroxymethyl)aminomethane-HCl

TUBE2, tandem ubiquitin-binding entity

UBL, ubiquitin-like

UBX, ubiquitin regulatory X

UBXL, UBX-like

Ufd1, ubiquitin fusion degradation 1

UPS, ubiquitin-proteasome system

v/v, $\frac{Volume\ (mL)}{Volume\ (mL)}$

VBM, VCP-binding motif

VCP, valosin containing protein

VIM, VCP interacting motif

Vms1, VCP mitochondrial stress-responsive 1

w/v, $\frac{Weight\ (g)}{Volume\ (mL)}$

YFP, yellow fluorescent protein

Chapter I

General Introduction

1.1 Plastid Diversity and Evolutionary Origin

1.1.1 Plastids are Functionally Diverse

Plastids are a family of organelles found within plants and algae which house a range of essential metabolic and biosynthetic functions (Jarvis & López-Juez, 2013; Rolland *et al.*, 2012). In higher plants, plastids are inherited as undifferentiated proplastids, which also exist in meristematic tissues (Jarvis & López-Juez, 2013). Plastid differentiation is dependent on specific developmental and environmental cues (van Wijk & Baginsky, 2011), ensuring their optimal function in the cell; the various forms may also interconvert throughout development.

The prototypical member of the group is the chlorophyll-pigmented chloroplast, which functions as the site of photosynthesis. Chloroplasts may develop from etioplasts, especially in early development, and they form in dark-grown tissues which contain precursor metabolites to enable a rapid chloroplast transition (Jarvis & López-Juez, 2013). Chromoplasts are another plastid form which contain large quantities of carotenoids (Sadali *et al.*, 2019). Plastids may also serve as energy storage organelles, such as amyloplasts, which are packed with starch granules. Senescing tissues will contain gerontoplasts which are formed from chloroplasts through disassembly involving autophagy (Jarvis & López-Juez, 2013). Plastids additionally have substantial metabolic and biosynthetic roles, including: nitrogen and sulfur assimilation; and the synthesis of amino acids, fatty acids, lipids, nucleotides, isoprenoids and alkaloids (Rolland *et al.*, 2012).

Plastids exist as integrated components of the cell. They participate in bidirectional signalling with the nucleus to coordinate gene expression for organelle biogenesis (Kakizaki *et al.*, 2009; Woodson *et al.*, 2013) or the regulation of plastid metabolism (Bräutigam *et al.*, 2009). Plastids

must exchange metabolites with other cellular compartments to facilitate processes such as photorespiration; they thus contain numerous membrane embedded metabolite transporters (Rolland *et al.*, 2012).

1.1.2 Evolution of Plastids

Plastids are descended from ancestral cyanobacteria, the organelle seemingly arising from a single endosymbiosis event (Sibbald & Archibald, 2020). This event is predicted to have occurred approximately 900 million years ago, giving rise to plastids shared by modern lineages of land plants, green algae, red algae, and glaucophytes (Shih & Matzke, 2013). Hallmarks of the endosymbiotic origin are the retention of a functional genome (the plastome), which is interpreted by eubacterial-type transcription and translation machineries, and a double-membraned envelope (McFadden, 2014). Over the course of plant evolution, roughly 98% of the endosymbiont's protein-coding sequences were either lost or relocated to the nuclear genome by the process of endosymbiotic gene transfer (Timmis *et al.*, 2004). More than 100 genes are retained in the chloroplast, and these encode vital factors such as core genetic components (Leister, 2003). Of the theories posited to explain why this subset of genes has been retained (Barbrook *et al.*, 2006), perhaps the most substantiated is that of colocation for redox regulation of gene expression (Allen, 2015): rapid compensatory effects are needed to respond to alterations in the redox state of components of the photosynthetic electron transport chain, preserving their function. The remaining 2000-3000 plastid proteins are encoded in the nucleus and imported post-translationally from the cytosol (Nakai, 2018; Paila *et al.*, 2015; Sjuts *et al.*, 2017).

1.2 Plastid Protein Import and its Regulation

The protein import apparatus and its regulation is an elaborate topic, and will be the focus of the majority of the remaining sections in this chapter. With respect to the components of the protein import system, the proteins found within higher plants (typically deriving from studies

in *Arabidopsis thaliana*) will be focused on — due to their relevance to this study. However, it should be noted that two breakthrough papers were recently published that provided cryo-electron microscopy (cryo-EM) structural data on the TOC-TIC supercomplex in *Chlamydomonas reinhardtii* (Jin *et al.*, 2022; Liu *et al.*, 2023). These remarkable studies provide valuable information on the components and organisation of the protein import complex in lower plants. There exist several differences between the protein import system in higher plants and lower plants, which are of note, including: the rigidity of the lower plant complex — as no TOC-TIC supercomplex has been successfully isolated from higher plants, it is speculated that it exists in a more transient state in higher plants between the two envelope membranes — and of the proteins in the complex, some of which are not conserved across taxa.

Nuclear encoded chloroplast proteins are typically synthesised as precursor proteins (pre-proteins). Pre-proteins possess a cleavable N-terminal targeting signal called a transit peptide (TP). Certain pre-proteins may be recognised by cytosolic chaperones, which interact with the transit peptide, to facilitate delivery to the chloroplast surface. Delivery into the chloroplast is achieved via translocon complexes in the outer (TOC) and inner (TIC) envelope membranes. These complexes specifically recognise transit peptide sequences and mediate their translocation into the plastid stroma. Once pre-proteins have entered the stroma, the transit peptide is removed by a the stromal processing peptidase (Trösch & Jarvis, 2011). Maturation of the protein may then complete by final folding and assembly in the stroma, or proteins may be routed in a partially unfolded state to the thylakoids (Celedon & Cline, 2013; Schünemann, 2007) or the inner envelope membrane (IEM) (Li *et al.*, 2017; Li *et al.*, 2015; Viana *et al.*, 2010).

There is now a clear consensus and large body of evidence to support the notion that flux of proteins imported through TOC and TIC complexes can vary greatly, according to the developmental stage, environmental cues, or stress conditions (Chu & Li, 2018; Suzuki *et al.*,

2015; Thomson *et al.*, 2020; Wang *et al.*, 2016; Watson *et al.*, 2018). To allow for this variation in protein flux, the import process needs to be adequately regulated; this is important in order to maintain an optimally functioning chloroplast proteome (Yang *et al.*, 2019). The cytosolic ubiquitin-proteasome system (UPS) has been shown to act as an important regulator of chloroplast protein import, by the negative regulation of TOC protein abundance (Ling & Jarvis, 2015a). The UPS is a eukaryotic protein degrading mechanism which is involved in numerous protein homeostasis processes in the cell (Vierstra, 2009).

As a bacterially-derived organelle, the chloroplast contains no identified UPS machinery internal to its envelope membranes, where E3 ligases have been identified. Indeed, the plastid possesses its own proteolytic machinery internally, in the form of approximately 20 proteases of prokaryotic origin, formed of three classes: FtsH (filamentation temperature-sensitive H), Deg (degradation of periplasmic proteins) or Clp (caseinolytic protease) (Fu *et al.*, 2022). Thee Clp and FtsH enzymes act to maintain internal protein homeostasis (Watson *et al.*, 2018). In addition, FtsH and Deg proteases are important regulatory components of photosystem II, aiding in the turnover of the frequently photodamaged D1 reaction centre protein (Kato & Sakamoto, 2009). Alternatively, under conditions of stress or senescence, the entirety, or part, of the organelle may be degraded by autophagy (Izumi *et al.*, 2017; Nishimura *et al.*, 2017; Rochaix & Ramundo, 2018; Woodson *et al.*, 2015).

1.3 Protein Targeting to the Chloroplast Surface

Chloroplast TPs are highly varied in their amino acid composition. It has been hypothesised that this variation is functional, as it may reflect different interacting domains for cytosolic, chloroplast membrane, and stromal chaperones and sorting factors (Lee & Hwang, 2018), or with different plastid forms (Li & Teng, 2013). Pre-proteins may be navigated to the plastid by cytosolic chaperones, which through their interaction inhibit the formation of secondary structures and maintain a largely unfolded conformation sufficient for import through the TOC

and TIC pores (Flores-Pérez & Jarvis, 2013). There are two cytosolic chaperone systems which have been reported to guide pre-proteins, though both of these systems are poorly characterised, and the mechanisms and physiological significance require elucidation. Hsp90 has been proposed to operate with Hsp70/Hsp90-organising protein (Hop) and immunophilin FKBP73 to deliver pre-proteins to the outer envelope membrane (OEM) (Fig. 1.1), where it may dock at Toc64 (Fellerer *et al.*, 2011; Panigrahi *et al.*, 2013). However, such delivery to Toc64 is seemingly not essential for protein import (Aronsson *et al.*, 2007; Hofmann & Theg, 2005). Alternatively, Hsp70 (Fig. 1.1) has been shown to interact with an undefined 14-3-3 protein to recognise phosphorylated TPs and deliver them to translocation complexes (May & Soll, 2000). Yet again the necessity of this process is unclear, as mutation of the implicated phosphorylation sites did not impede targeting to the chloroplast (Nakrieko *et al.*, 2004).

The TP mediated route through TOC-TIC is clearly the canonical model for protein delivery to plastids. However, there do exist several additional targeting pathways that deliver plastid proteins. These are even less well studied, and it is believed that their clients may number in only the hundreds of proteins (Jarvis & López-Juez, 2013). OEM proteins are not typically synthesised with an N-terminal TP, with targeting information instead residing within a transmembrane domain itself (Kim *et al.*, 2019; Lee *et al.*, 2017). An interesting and important exception is the β-barrel protein Toc75. Delivery and insertion of this protein of the TOC complex is mediated by a bipartite targeting peptide, the first part of which is a canonical TP (Day *et al.*, 2014; Inoue & Keegstra, 2003). Subsequently there is evidence to suggest that other β-barrel proteins may also be inserted through a pathway involving TPs (Day *et al.*, 2019), but those which seemingly are not are targeted to the chloroplast by their penultimate β-barrel strand (Klinger *et al.*, 2019) with the aid of TOC machinery (Day *et al.*, 2019). Lastly, there exist at least two non-canonical pathways of chloroplast protein targeting, which deliver internal proteins lacking N-terminal signals in a TOC-independent fashion (Jarvis, 2004), or involve passage through the endomembrane system (Baslam *et al.*, 2016).

1.4 The TOC Complex and Protein Import in the Outer Envelope Membrane

The TOC complex is present in the outer envelope membrane of chloroplasts and mediates the translocation of preproteins across this first membrane. The principal components of the TOC complex are three proteins with distinct roles, called Toc33, Toc159 and Toc75 (Jarvis, 2008). The associated nomenclature of the numbering is based on their observed molecular weights in kDa by protein gel migration. Toc33 and Toc159 are receptors that possess cytosol-projecting GTPase domains that bind to the TPs of pre-proteins (Jarvis & López-Juez, 2013). In higher plants, both receptors belong to protein families and so exist in multiple functional isoforms (Jackson-Constan & Keegstra, 2001).

1.4.1 Toc159

The Toc159 family (Toc159, Toc132, Toc120, Toc90) is the larger and more complex group of receptors. They possess three domains: a GTPase domain (Leipe *et al.*, 2002), located centrally (known as the 'G' domain); a large C-terminal membrane-anchoring domain (Paila *et al.*, 2015) (known as the 'M' domain); and finally an N-terminal intrinsically disordered acidic domain, which may act to convey recognition specificity (Agne *et al.*, 2010; Inoue *et al.*, 2010). Both the A and G domains extend into the cytosol to interact with preproteins. The A domain is frequently cleaved by an unknown protease, as such Toc159 was initially believed to be a far smaller protein (Perry & Keegstra, 1994). Although the A-domain is seemingly nonessential (Schnell *et al.*, 1994), it has also been observed that full-length Toc159 was more efficient at facilitating the import of preproteins than the variant with a cleaved A domain (Bolter *et al.*, 1998). The A domain may also be highly phosphorylated, and it has been proposed that this post translational modification may function in preprotein client specificity (Agne *et al.*, 2010). This view may be supported by the discovery of the membrane embedded kinase Koc1 (kinase at the outer chloroplast membrane 1), as *koc1* plants showed reduced import efficiency (Zufferey *et al.*, 2017). The M domain is strikingly atypical, containing no apparent hydrophobic regions or transmembrane domains (Paila *et al.*, 2015). It is seemingly a

necessary structural element for the formation of the TOC complex (Lee *et al.*, 2003), and has also been observed to interact with preproteins and β strands of Toc75 (Kouranov & Schnell, 1997), suggesting that Toc159 may imbed or position highly proximally to the Toc75 pore (Wallas *et al.*, 2003). Recent structural prediction using Alphafold2 would suggest that the Toc159 M-domain forms a β-barrel structure. Indeed, the lower plant homologues of Toc159, Toc90/Toc120, were recently revealed to form a hybrid channel with Toc75 in *Chlamydomonas* via interaction of their respective β-barrel strands (Jin *et al.*, 2022; Liu *et al.*, 2023).

The G domain is a related GTPase domain to that found in Toc33 (Leipe *et al.*, 2002). It is part of a family of GTPase domains that are characterized by low rates of GTP hydrolysis (Reddick *et al.*, 2007) and require dimerization to function (Weibel *et al.*, 2003). Functional GTPase domains seem to be essential for preprotein import, although protein import can tolerate either single mutants of Toc33 or Toc159 G domains (Agne *et al.*, 2009; Aronsson *et al.*, 2010), double mutants are seemingly lethal.

1.4.2 Toc33

Members of the Toc33 family (Toc33 and Toc34) have a relatively simple architecture, with a C-terminal membrane-embedded domain (Li & Chen, 1997) and an N-terminal GTPase domain (Paila *et al.*, 2015). Toc34 is expressed weakly in most plant tissues, though it is much elevated in roots (Gutensohn *et al.*, 2000). Plants without it also lack the typical features of plastid protein import mutants, with chloroplast ultrastructure and protein import seemingly unaffected in *ppi3* plants (Sjuts *et al.*, 2017). Toc33 is expressed to a higher degree, particularly in earlier stages of development in photosynthetic tissue. The *ppi1* mutant of Toc33 was discovered during a forward genetic screen (Jarvis *et al.*, 1998), and demonstrates clear defects in the import of photosynthetic preproteins (Kubis *et al.*, 2003).

1.4.3 Distinct TOC Complexes

The families of Toc159 and Toc33 proteins have led to the theory that there are distinct TOC import complexes with different pre-protein preferences and abundances during development (Bischof *et al.*, 2011; Gutensohn *et al.*, 2000; Inoue *et al.*, 2010; Kubis *et al.*, 2004). Toc132 and Toc120 form distinct receptor complexes and are not able to rescue the knock-out phenotype of Toc159 (*ppi2*), and *vice versa* (Inoue *et al.*, 2010; Kubis *et al.*, 2004). Multiple studies have suggested that Toc159 has preferential interaction with photosynthetic pre-proteins and that Toc132 and Toc120 are essential for basal functionality in all plastid forms (Bischof *et al.*, 2011; Kubis *et al.*, 2004). Toc90, on the other hand, may be more functionally similar to Toc159, as it is able to rescue the *ppi2* phenotype (Infanger *et al.*, 2011). The specificity of Toc33-type receptors is also not as pronounced; for instance, overexpression of Toc34 can rescue the Toc33 knock-out's (*ppi1*) phenotype (Paila *et al.*, 2015).

1.4.4 Toc75 and TOC Pore

Pre-proteins are translocated through the OEM via a channel made by the β-barrel protein Toc75 (Fig. 1.1). Toc75 belongs to the bacterially-descended Omp85 protein superfamily (Schleiff & Becker, 2011). A feature of this family is a soluble N-terminal polypeptide transport associated (POTRA) domain (Koenig *et al.*, 2010), which in the case of Toc75 extends into the intermembrane space (IMS) and performs a proposed chaperone-like activity (O'Neil *et al.*, 2017). The Toc75 channel has been found to reach a maximum diameter of roughly 30 Å (Ganesan *et al.*, 2018), which has led to speculation that it may be capable of incorporating some folded proteins, or even partially folded larger proteins. The molecular weight and stoichiometry of the TOC Toc33:Toc75:Toc159 core complex is still not precisely defined. It has been determined to be in the range of 4:4:1 (Schleiff, Soll*, et al.*, 2003) to 6:6:2 (Kikuchi *et al.*, 2006), and so the channel may be formed by multiple copies of the Toc75 protein and/or in combination with Toc159 family proteins (Ganesan & Theg, 2019) – which may account for the proposed maximum pore diameter, if multiple β-barrel proteins are capable of contributing to the pore via interaction of β-barrel strands. Accordingly, the mass of the complex also has

substantial variation, with the same studies finding it may exist either as approximately 500 kDa (Schleiff, Eichacker*, et al.*, 2003) or 800-1000 kDa (Kikuchi *et al.*, 2006).

1.4.5 TP Interaction with the TOC Complex

Initial TP interaction with the GTPase domains of Toc159 and Toc33 is transient and energy independent (Ma *et al.*, 1996), potentially allowing for rapid and sequential interaction with pre-proteins (Richardson *et al.*, 2018). Later, the TP interacts with the POTRA domain of Toc75 and the soluble IMS protein Tic22, before GTP hydrolysis (Richardson *et al.*, 2018). While GTP hydrolysis at both receptors is not necessary for protein import *in vivo* (Agne *et al.*, 2009; Aronsson *et al.*, 2010), it is required for successful protein translocation *in vitro* (Richardson *et al.*, 2018), reinforcing the notion that the GTPase receptors are the first points of contact (Andrès *et al.*, 2010; Paila *et al.*, 2015; Sjuts *et al.*, 2017). The TP may bind simultaneously to Toc33 and Toc159, as each preferentially bind to a distinct region of the peptide (Wiesemann *et al.*, 2019). Bound TP may then encourage heterodimer formation between Toc33 and Toc159, as well as GTP hydrolysis (Lumme *et al.*, 2014), leading to an activated translocon conformation which the pre-protein can pass through (Paila *et al.*, 2015).

1.5 The Intermembrane Space

In the IMS, Tic22 is a well-evidenced component of the import system. There are two functionally redundant isoforms in *Arabidopsis* (Kasmati *et al.*, 2013). These have been shown to act as chaperones (Glaser *et al.*, 2012; Tripp *et al.*, 2012) and facilitate pre-protein delivery to the TIC complex, through passive interactions with proteins in the OEM and IEM (Kouranov *et al.*, 1998) (Fig. 1.1).

An important recent study identified the IEM protein Tic236 as an element of a 1.25 MDa TOC-TIC supercomplex (Chen *et al.*, 2018). Tic236 has been proposed to provide a physical link between the TOC and TIC complexes. The protein is anchored in the IEM, where it associates

with Tic20, and mediates a physical link to the OEM through direct interaction with the POTRA domain of Toc75 (Chen *et al.*, 2018) (Fig. 1.1). However, alternative functions were found in a study of its maize orthologue, defective kernel5 (DEK5) (J. Zhang *et al.*, 2019). DEK5 was suggested to function in envelope biogenesis, through mediating the insertion of β-barrel proteins involved in protein import (such as Toc75), as well as assisting in the insertion of proteins involved in metabolite transport (J. Zhang *et al.*, 2019). From their summation, DEK5 sustains a conserved function to its bacterial homologue, TamB. Translocation and assembly module (Tam) proteins are found in gram negative bacteria and assist in the integration of β-barrel membrane proteins. Defects in TamB in bacteria result in altered membrane morphology and reduced metabolite export (Iqbal *et al.*, 2016). TOC protein abundance was unchanged in the *tic236* mutant, which lead the authors to propose that Tic236 had a distinct function (Chen *et al.*, 2018). In contrast, *dek5* mutants did in fact demonstrate reduced TOC protein abundance, and thus the impact on protein import could well be a secondary effect (J. Zhang *et al.*, 2019). Ultrastructure analysis revealed a reduction in the proportion of envelope relative to other chloroplast compartments in *dek5* (J. Zhang *et al.*, 2019), but a similar analysis was not done for *tic236* plants (Thomson *et al.*, 2020).

1.6 The TIC Complex and Protein Import at the Inner Envelope Membrane

Unlike the TOC machinery at the OEM, the identity of the TIC components is still not entirely resolved. In a landmark study, the 1 MDa TIC complex (Kikuchi *et al.*, 2013) was identified which has led to a re-evaluation of the formerly proposed models for pre-protein delivery through the TIC apparatus (Nakai, 2018). The 1 MDa complex consists of Tic214 (encoded by the chloroplast gene *Ycf1*), Tic100, Tic56, Tic20 and Tic21 (Kikuchi *et al.*, 2013) (Fig. 1.1).

1.6.1 The TIC Pore

The identity of the pore does not have a complete consensus, but it is likely that the translocon channel is formed by Tic20, which has four membrane-spanning α-helical domains (Kovács-

Bogdan *et al.*, 2011; Thomson *et al.*, 2020). Based on complex mass, three copies of Tic20 could theoretically exist within the 1 MDa complex (Kikuchi *et al.*, 2013), so a pore size of up to 30 Å has been predicted (Ganesan & Theg, 2019) – comparable to the current predictions for the TOC complex. Tic21 has a similar structure to Tic20, though the proteins are not closely related. Based on mutant analysis, it may have a role in protein import also and function in a complementary way to Tic20 (Teng *et al.*, 2006); though another study identified its function within iron homeostasis (Duy *et al.*, 2007).

1.6.2 Classical Components of the TIC Complex

The originally identified components of the TIC import machinery include the proteins Tic110, Tic40, Tic20 and Tic21 (Fig.1.1) (Thomson *et al.*, 2020). The functional relationship between Tic110 and Tic40 and the 1 MDa complex is unclear, but it may be Tic110 and Tic40 act in the later stages of import to mediate interactions with chaperones (Nakai, 2018) or have client specific preferences based on the composition of the pre-protein TP (Lee & Hwang, 2019).

The structure of Tic40 has only a single N-terminal transmembrane domain, and a large stromal extension. The stromal region contains numerous tetratricopeptide repeat (TPR) domains, which are known to promote protein-protein interactions (Paila *et al.*, 2015). In addition, the stromal region also contains a region with conserved sequence with HIP (Hsp70 interacting protein) and HOP (Hsp70/Hsp90 organizing protein) (Frydman & Hohfeld, 1997). It is on the basis of these domains, that is was suggested that Tic40 is involved int the recruitment of chaperones during the import process (Bedard *et al.*, 2007). Tic110 instead has two membrane spanning α helices, but a large stromal extension in a superficially similar way to Tic40. The stromal domain has been shown to bind to TPs (Inaba *et al.*, 2003), as well as the stromal chaperone Hsp93/ClpC (Flores-Pérez *et al.*, 2016). It was at one point proposed to form part of the TIC pore, but examination of the crystal structure found it implausible that it could form a channel (Tsai *et al.*, 2013).

Stromal chaperones are therefore believed to cooperate with Tic110 and Tic40 to facilitate pre-protein import and subsequent folding (Chou *et al.*, 2003; Inaba *et al.*, 2003). Translocation into the stroma is an energy-dependent process (Shi & Theg, 2013). Several chaperones, including cpHsp70, Hsp90C and Hsp93, have been implicated in the provision of this motive force (Paila *et al.*, 2015) (Fig. 1.1). It is clear however that the exact role of these chaperones requires clarification to build a mechanistic model of final delivery through the TIC complex. While cpHsp70 has been strongly linked to the role of the main protein import motor (Liu *et al.*, 2014; Shi & Theg, 2010; Su & Li, 2010), Hsp90C was also found to be essential for protein translocation (Inoue *et al.*, 2013). A motor function has also been proposed for Hsp93 (ClpC) (Akita *et al.*, 1997; Nielsen *et al.*, 1997), but recent data support the long-known interaction with the ClpP protease (Halperin *et al.*, 2001) showing that it works in a protein quality control process (Flores-Pérez *et al.*, 2016; Sjögren *et al.*, 2014).

1.6.3 Reconciling the Newly Identified Components of the TIC Complex

The TIC complex may be even more complex than previously understood, as yet another large complex, a 2 MDa Ycf2-FtsHi protein complex, was identified (Kikuchi *et al.*, 2018); this was proposed to act as a general import motor associated with the 1 MDa TIC complex (Kikuchi *et al.*, 2018) (Fig. 1.1). In view of the evidence supporting 30 Å pore diameters for the import channels (Ganesan & Theg, 2019), it was suggested that such a powerful ATPase might be specifically recruited to handle the import of recalcitrant or tightly-folded proteins (Herrmann, 2018). It does however seem plausible that cpHsp70 is the general import motor given the existing evidence, which must still be disproven (Liu *et al.*, 2014; Shi & Theg, 2010; Su & Li, 2010). The energetically more expensive Ycf2-FtsHi complex may only be utilised with select clients or under specific conditions (Thomson *et al.*, 2020).

The newly identified TIC complexes of the last decade remain enigmatic discoveries in need of further research. It is the first plastid import complex to contain a chloroplast-encoded protein (Tic214/Ycf1). The complex has been proposed to act as a general TIC translocon at

the IEM, evidenced by its co-purification with TOC proteins (Kikuchi *et al.*, 2013), an interaction that could in principle be mediated by Tic236 (Thomson *et al.*, 2020). Additionally, the 1 MDa complex as the general TIC translocoon is strongly supported by recent structural determination of the TOC-TIC supercomplex, identifying components of the 1 MDa complex in *Chlamydomonas* (Jin *et al.*, 2022; Liu *et al.*, 2023).

Mutant analysis of *ycf1* was not feasible, as knock-out plants demonstrated an embryo lethal phenotype – this is indicative of an essential function in plant development, potentially in chloroplast protein import. However, no defects in protein import were also observed when *Ycf1* translation was repressed by the specific plastid ribosomal inhibitor spectinomycin (Bölter & Soll, 2017). As Tic214 has also been found to have additional functions in the biogenesis of photosynthetic complexes in thylakoid membranes (Yang *et al.*, 2016), these other roles may contribute to the severity of the knock-out mutant's phenotype.

A viable knock-out mutant of *tic56* was attained, though this was only true for a truncated protein variant in the allele *tic56-2* (Kikuchi *et al.*, 2013). *tic56* plants display chlorotic phenotypes (Kikuchi *et al.*, 2013), which is typical of impaired chloroplast protein import. Analysis of *tic56* mutants have attributed the observed chlorotic phenotypes to defects in ribosome assembly (Agne *et al.*, 2017; Köhler *et al.*, 2016; Köhler *et al.*, 2015; Schäfer *et al.*, 2019). However, *tic56* has received conflicting evidence pertaining to its role in protein import, either having been observed to have reduced import efficiency (Kikuchi *et al.*, 2013) or no impact to it (Köhler *et al.*, 2015). This may have been related to different growth conditions or preparations of the isolated chloroplasts for the *in vitro* import assays (Loudya *et al.*, 2022).

Like *tic56*, complete knock-out *tic100* plants are also embryo lethal. However, two alleles of *Tic100* were recently identified which allowed for new investigations into the role of Tic100 in protein import. First identified in a mutant screen for *Chlorophyll a/b Binding Protein 3* (*CAB3*), the *CAB3 under expressed* 8 (*cue8*) mutant was generated (Lopez-Juez *et al.*, 1998). Through

whole genome sequencing, *cue8* was discovered to be a *tic100* mutant, renamed to *tic100^{cue8}* (Loudya *et al.*, 2020). The initial mutation present in *tic100^{cue8}* contains the amino acid substitution G366R. Strikingly, the phenotype of *tic100^{cue8}* is supressed by the allele *tic100^{soh1}* (*suppressor of tic100 1)*, in the *tic100^{cue8} tic100^{soh1}* double mutant (Loudya *et al.*, 2022). *tic100^{soh1}* conversely contains the mutation R345Q. It is believed that the introduction — and subsequent removal in the suppressor — of an additional positively charged arginine residue interferes with protein stability or interactions (Loudya *et al.*, 2022) – both mutations are present in a predicted β-sheet domain, expected to interface with lipids based on an Interpro annotation. These mutations in Tic100 offered an elegant system to investigate the protein's function. In summary, *tic100^{cue8}* plants showed substantially reduced protein import efficiency, which was independent of ribosome processes or retrograde signalling mediated transcriptional regulation (Loudya *et al.*, 2022) – these defects in protein import, plant and chloroplast development, were restored to near wild type levels in the *tic100^{cue8} tic100^{soh1}* mutant. In addition, *tic100^{cue8}* plants (like *tic56*) showed a decrease in the protein abundance of components of the 1 MDa complex, despite elevated gene expression (Loudya *et al.*, 2022). These data suggest that Tic100 is an important component of the 1 MDa complex and support the role of the 1 MDa complex in protein import, though the exact function of Tic100 in the complex is not well appreciated.

One the strongest arguments against the 1 MDa complex as a general import motor in plants, is the lack of conservation in the monocot family of grasses, Poaceae. The Poacea family is of particular agronomic interest as major crop species such as rice (*Oryza sativa*) and wheat (*Triticium aestivum*) are members. There are no identified orthologues of Tic214, Tic100 or Tic56 in the Poaceae, while all other components of the 1 MDa TIC complex remain conserved in higher plants (Jarvis & López-Juez, 2013). It is worth noting that while there are some additional proteins present in the Chlorophyta (Toc10, Toc39, Toc52, and Tic13) (Jin *et al.*, 2022) — which were presumably lost over evolutionary time in the Embryophyta — the 1 MDa complex components (Tic214, Tic100, Tic56, Tic20) are all still present (Jin *et al.*, 2022) and

presumably represent the ancestral state. There therefore exists several intriguing questions regarding the mechanism of protein import at the IEM in the Poaceae, and why proteins of the 1 MDa complex were lost during evolution. There do exist some clear differences in chloroplast biogenesis between the Poaceae and other flowering plants which may be the result of, or contribute to, differences in the TIC architecture – these include slower greening times of the tissue as well as the lack of a dual targeted RNA polymerase to the chloroplast and mitochondrion (as only single targeted enzymes are present for both organelles) (Borner *et al.*, 2015). There is an additional difference in the identity of the channel protein Tic20. In *Arabidopsis thaliana*, the dominant isoform of Tic20 in the 1 MDa complex is Tic20-I, whereas Tic20-IV is more abundant in root tissue (Kikuchi *et al.*, 2013). Intriguingly, the Poaceae orthologue of Tic20 present in photosynthetic tissue has greater sequence similarity to Tic20-IV (Nakai, 2018), suggesting further differences in the IEM translocon compared to other flowering plants.

To reconcile the evidence, it is likely that the 1 MDa TIC components are the ancestral state of the IEM protein import machinery. This is based on their phylogenetic presence, structural and functional significance observed in the Chlorophyta (Jin *et al.*, 2022; Liu *et al.*, 2023; Ramundo *et al.*, 2020), as well as in higher plants excluding the Poaceae (Kikuchi *et al.*, 2018; Kikuchi *et al.*, 2013; Loudya *et al.*, 2022; Loudya *et al.*, 2020; Nakai, 2018). There is likely a unique mode of import that has been acquired in the Poacea, which requires much additional focus in future studies. It is also plausible that the 1 MDa complex may act in multiple processes alongside protein import under specific conditions, with additional machinery operating either in series or in parallel. The extent of involvement of the 1 MDa complex in import may depend on the nature of the client proteins, or on specific developmental or environmental conditions (Thomson *et al.*, 2020). The classical components of import, including Tic40 and Tic110, may function in parallel or have a stronger role in post-import processing of pre-proteins in the stroma.

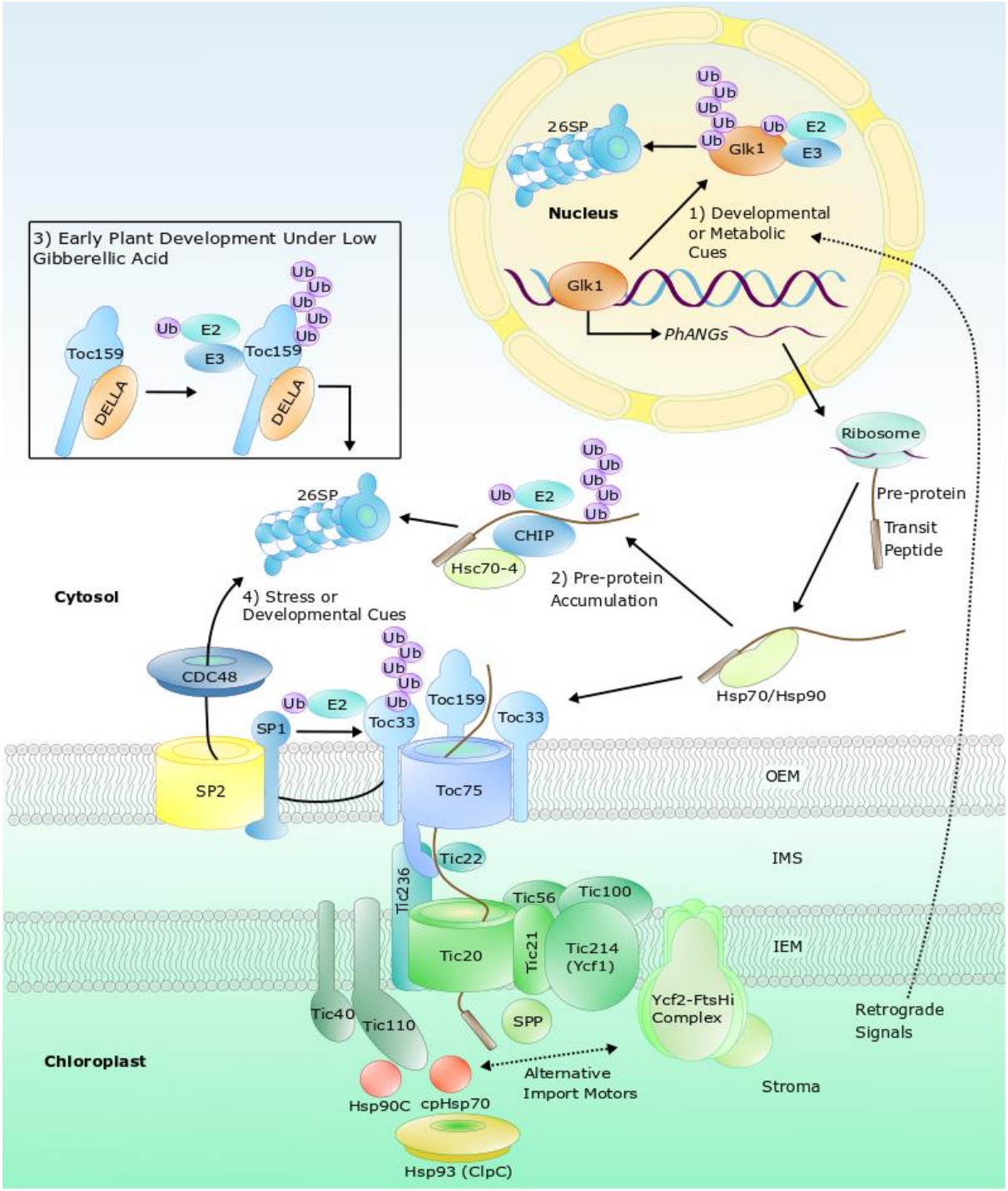


Figure 1.1 Adapted from Thomson *et al*. (2020): an overview of protein import into the chloroplasts and its regulation by the UPS, Nucleus-encoded chloroplast proteins are synthesised in the cytosol as pre-proteins and post-translationally imported into the chloroplast. The pre-protein carries an N-terminal transit peptide which holds guidance information and initially allows interaction with cytosolic chaperones (e.g., Hsp70, Hsp90). The pre-protein is then recognised by receptor GTPases in the OEM, Toc33/34 and Toc159/132/120/90, which heterodimerize to allow the pre-protein to pass through the Toc75

pore into the IMS. Passage through the IEM is mediated by Tic20, which is reported to be part of the 1 MDa TIC complex containing Tic21, Tic56, Tic100 and Tic214 (Ycf1). Completion of translocation into the stroma is powered by an ATP-dependent import motor, which may be composed of stromal molecular chaperones (e.g., cpHsp70, Hsp90C) or a 2 MDa Ycf2-FtsHi complex. The Tic40 and Tic110 proteins are also involved in the import process, and may operate downstream in conjunction with stromal chaperones. Hsp93 (ClpC) has been proposed to perform a quality-control function at the point of import, or to act in the import motor. Upon arrival in the stroma, the transit peptide is cleaved from the pre-protein by the SPP. The UPS regulates protein import in a variety of ways: 1) The transcription factor Glk1, which regulates the expression of pre-protein-encoding PhaNGs, may be degraded by the UPS in response to unknown retrograde signals (from the chloroplast to the nucleus) in response to developmental or metabolic cues. 2) Accumulation of pre-proteins in the cytosol may trigger their UPS degradation to prevent formation of cytotoxic aggregates, and this is mediated by the chaperone Hsc70-4 and the E3 ligase CHIP.3) Before germination, DELLA factors repress chloroplast biogenesis under low gibberellic acid conditions by binding to Toc159 and triggering its UPS degradation. 4) During stress or particular phases of development, the CHLORAD system directly targets the TOC apparatus for proteolysis, with ubiquitination being mediated by the E3 ligase SP1. The targeted TOC proteins are retrotranslocated from the membrane via the channel protein SP2, using motive force provided by the cytosolic AAA+ ATPase CDC48. The Toc33 and Toc159 GTPase isoforms are displayed due to the identified role of CHLORAD to supress the import of photosynthetic proteins in response to abiotic stress (Ling & Jarvis, 2015b), as photosynthetic proteins are the primary clients of these isoforms (Jarvis & López-Juez, 2013). All the other TOC GTPase isoforms have also been identified as substrates of SP1, as *sp1* plants displayed imbalances in all TOC proteins during de-etiolation (where chloroplasts develop from dark grown etioplasts) (Ling et al., 2012). Dashed lines indicate uncertainty.

1.7 The Ubiquitin Proteasome System

The UPS is a major regulatory system involved in the targeting of misfolded or unnecessary proteins for degradation. It functions within many biological processes, for example hormone signalling (Christians *et al.*, 2009). Ubiquitination (or ubiquitylation) is a post-translational modification involving the addition of one or more copies of the 8.5 kDa ubiquitin protein to lysine residues of target proteins (Vierstra, 2009). The addition of polyubiquitin chains signals the protein for degradation by the nucleocytosolic 26S proteasome (26SP) (Vierstra, 2009). The 26SP is an ATP-dependent proteolytic complex formed from a cylindrical 20S core particle and a 19S regulatory particle; ubiquitinated proteins are recognised by the regulatory particle which guides them to the core where they are degraded (Sako *et al.*, 2014). Importantly, this cytosolic machinery also targets proteins from organelles, most famously the endoplasmic reticulum (ER) in ER-associated protein degradation (ERAD) (Liu & Li, 2014), but also the endosymbiotically-derived mitochondria and chloroplasts (Hua & Vierstra, 2016). Indeed, proteomic studies of plant ubiquitinomes have identified a large number of internal chloroplast proteins which are likely to be UPS targets (Kim *et al.*, 2013; Svozil *et al.*, 2014; Xie *et al.*, 2019).

Ubiquitination requires an enzyme cascade involving the activation and targeted conjugation of ubiquitin. An E1 ubiquitin activase first forms a thioester bond with ubiquitin in an ATP-dependent reaction (Vierstra, 2009). The ubiquitin moiety is then transferred to an E2 ubiquitin conjugase. Finally, an E3 ubiquitin ligase conveys substrate specificity through a selective interaction with its targets, catalysing the transfer of ubiquitin to the target from the E2 enzyme (Vierstra, 2009). Reiterative rounds of conjugation onto ubiquitin lysine residues result in the formation of a polyubiquitin chain degradation signal. Upon degradation, the ubiquitin moieties are recycled through the action of deubiquitinating enzymes. The E3 ligases are necessarily numerous and diverse, given their role in specificity, with roughly 1,400 E3 proteins in *Arabidopsis*, far outnumbering the ~40 E2 and 2 E1 enzymes (Ling & Jarvis, 2015a). In plants,

there are four classes of E3 ligase: homologous to the E6-AP carboxyl terminus (HECT), really interesting new gene (RING), U-box, and cullin-RING ligase (CRL) (Ling & Jarvis, 2015a). Each class has a different mechanism of action and subunit composition, with particular diversity in the substrate-interacting domain or component (Vierstra, 2009).

1.7.1 Degradation of Precursor Proteins

Initial evidence for the UPS regulation of chloroplast proteins came from observations of cytosolic E3 activity targeting pre-proteins (Lee *et al.*, 2009). The C-terminus of Hsc70-interacting protein (CHIP) E3 ligase was initially shown to direct the degradation of Clp and FtsH precursors under high-light conditions (Shen, Adam*, et al.*, 2007; Shen, Yan*, et al.*, 2007). A subsequent study revealed that in the *Arabidopsis* Toc159 mutant *plastid protein import2* (*ppi2*), pre-proteins accumulated in the cytosol and the expression of the cytosolic chaperone Hsc70-4 (an Hsp70 isoform) was elevated (Lee *et al.*, 2009). The Hsc70-4 was shown to interact with the TP of pre-proteins and recruit them to CHIP for ubiquitination and degradation by the 26SP (Lee *et al.*, 2009) (Fig. 1.1). This system is suggested to function as a quality-control process to degrade mis-targeted proteins and/or prevent the accumulation of pre-proteins in the cytosol, as unfolded proteins may accumulate into cytotoxic aggregates (Lee *et al.*, 2009). A recent study in wheat identified another cytosolic E3 ligase, stress-associated protein 5 (SAP5), which interacts with the pre-protein of Hsp90C to trigger its degradation (N. Zhang *et al.*, 2019).

1.7.2 The UPS and Chloroplast Biogenesis

A further influence of the UPS on cytosolic events controlling chloroplast biogenesis was reported to occur during early plant development, pre-germination. DELLA proteins inhibit seed germination in processes regulated by UPS-mediated degradation (Li *et al.*, 2016). Germination depends on the accumulation of the hormone gibberellic acid, which downregulates DELLA factor accumulation to enable germination and, in turn, chloroplast

biogenesis (Shanmugabalaji *et al.*, 2018). It was reported that all DELLA factors can interact with cytosolic Toc159, prior to its assembly into the TOC complex, and initiate its degradation by the 26SP (Shanmugabalaji *et al.*, 2018) (Fig. 1.1). Low gibberellic acid conditions also resulted in the UPS-dependent downregulation of pre-proteins (Shanmugabalaji *et al.*, 2018).

The UPS also exerts indirect effects on chloroplast development through nuclear activities (Hirosawa *et al.*, 2017; Hua & Vierstra, 2016; Ling & Jarvis, 2015a; Rochaix & Ramundo, 2018). In *Arabidopsis*, the two Golden2-like (Glk) transcription factors function redundantly to promote the expression of photosynthetic proteins, thereby promoting chloroplast biogenesis (Tokumaru *et al.*, 2017). Glk1 itself is regulated at the transcriptional level through plastid-to-nucleus signals mediated by genomes uncoupled1 (GUN1) in response to the developmental state of the organelle (Kakizaki *et al.*, 2009). Interestingly, Glk1 has also been found to be regulated at the posttranscriptional level by the UPS in response to an as yet unknown, GUN1-independent plastid signal (Tokumaru *et al.*, 2017) (Fig. 1.1). This signal may derive from an environmental or developmental source to control chloroplast biogenesis (Tokumaru *et al.*, 2017).

1.8 Chloroplast-Associated Protein Degradation

1.8.1 Sp1 and Sp2

The first evidence for direct interaction between the UPS and chloroplast proteins *in situ* came from the discovery of the ubiquitin-dependent degradation of TOC complexes. A forward-genetic screen identified the RING type E3 ligase, suppressor of *ppi1* locus 1 (Sp1) (Ling *et al.*, 2012). Located in the OEM (Ling *et al.*, 2017), Sp1 possesses two transmembrane domains separated by an intermembrane space domain, which acts in target recognition, and a C-terminal cytosolic RING domain (Ling *et al.*, 2012). Sp1 directly interacts with all TOC proteins, mediating their ubiquitination and degradation (Fig. 1.1); thus, there is an increase in TOC protein abundance when Sp1 is lost. Accordingly, in the double mutant *sp1 ppi1* an enhanced greening (or suppression of the *ppi1* single mutant's chlorotic phenotype) is

produced — a phenotype mirrored by UPS inhibition (Ling *et al.*, 2012). The observed greening phenotype in the *sp1 ppi1* double mutant is attributed to the relative increase in the protein abundance of other TOC proteins, allowing for increased protein import despite the lack of Toc33. Further experiments revealed that such degradation of TOC complexes provides for important regulation of the import machinery, and may act to alter the proteome, functions and developmental fate of the organelle (Ling & Jarvis, 2015b). This regulation can also help to promote the plant's tolerance of abiotic stress, by depleting the import of photosynthetic proteins and thus suppressing photosynthesis and the tendency to overproduce reactive oxygen species during stress (Ling & Jarvis, 2015b).

To be degraded by the cytosolic 26SP, polyubiquitinated TOC proteins first need to overcome the physical and energetic barriers to their removal from the OEM. Degradation of ER membrane and mitochondrial outer membrane proteins involves retrotranslocation across the membrane before degradation by the 26SP (Liu & Li, 2014; Wu *et al.*, 2016), and it was thought that protein degradation at the chloroplast OEM may involve an analogous process. Strikingly, an additional product of the suppressor screen that identified Sp1 was the Omp85 protein Sp2, and this OEM protein was shown to assist TOC retrotranslocation (Ling *et al.*, 2019). As an Omp85 family member, Sp2 shares homology with Toc75 and is capable of forming a channel, although it lacks a POTRA domain. Like Sp1, Sp2 physically interacts with TOC proteins, and it is hypothesised to form the retrotranslocon channel (Ling *et al.*, 2019) (Fig. 1.1). The entry of substrates into the channel may occur by lateral gating in the membrane, as with other Omp85 proteins (Ganesan & Theg, 2019).

1.8.2 Cdc48 in CHLORAD

The motive force for extraction through Sp2 is provided by a cytosolic factor: cell division cycle 48 (Cdc48) is a homohexameric ATPase and a member of the ATPases associated with various cellular activities (AAA+) family of proteins (Ye *et al.*, 2017).

Much of the information available regarding the mechanism of action and structure of Cdc48 and associated UPS proteins is derived from studies in non-plant organisms; the organism of study will therefore be referred to in each instance.

In yeast, conformational changes induced by ATP hydrolysis create a motive force, allowing for the extraction and denaturation of bound substrates through the central pore (N. O. Bodnar & T. A. Rapoport, 2017; Twomey *et al.*, 2019) (Fig. 1.1). Cdc48 is known to function in a variety of cellular activities beyond protein homeostasis in eukaryotes, such as cell cycle regulation and autophagy, but it is especially well-known as the core motor component in ERAD (Bègue *et al.*, 2017). Of the five reported homologues of Cdc48 in *Arabidopsis*, CDC48A was found to associate with Toc33 by mass spectrometry (Ling *et al.*, 2019), and has recently been identified by proteomic analysis of the chloroplast envelope (Bouchnak *et al.*, 2019). Reconstitution experiments demonstrated that Cdc48 operates as a cytosolic motor to retrotranslocate ubiquitinated TOC proteins prior to their degradation, a process in which Sp2 was also shown to be critical (Ling *et al.*, 2019) (Fig. 1.1). This pathway of TOC degradation by the UPS involving Sp1, Sp2 and Cdc48 has been named chloroplast-associated protein degradation (CHLORAD) (Ling *et al.*, 2019).

1.9 Cdc48 and Adapter Proteins

1.9.1 Cdc48 Structure and Functions

Cdc48 is a molecular chaperone of numerous essential functions and has thus been the subject of much research across the life sciences. Cdc48 was originally discovered in yeast (*Saccharomyces cerevisiae*) in a genetic screen for regulators of the cell cycle (Moir *et al.*, 1982). The mammalian homologue has the alternative name of VCP (valosin containing protein), and the molecular chaperone has also been given the name P97 – but for duration of this thesis, it shall be referred to as Cdc48. As a note, the literature relating to Cdc48 often refers to a protein which modifies its activity as a “cofactor” (Buchberger *et al.*, 2015); to avoid

confusion with the typical biochemical definition of a cofactor, the term ‘adapter protein’ is used in its place here.

As aforementioned, it is a member of the of the AAA+ family of proteins and contains conserved ATPase domains across eukaryotes, with Walker A and Walker B motifs involved in ATP binding and hydrolysis (Bègue *et al.*, 2017). The remainder of the paragraph details structural and functional information that has been determined in studies of yeast and mammalian proteins, where there is notable conservation. As an assembled complex, Cdc48 has been determined to form a homohexamer formed by two concentric ATPase rings called D1 and D2 in (Wolf & Stolz, 2012) (Fig.1.2); these are stacked on top of each other forming a central pore. Above the ATPase rings lie N-terminal domains, involved in substrate and adapter protein binding (Davies *et al.*, 2008). The position of the N domains is dependent on the nucleotide state of the adjacent D1 ring: the N domain is coplanar in the ADP state, and rises up above the D1 ring when it’s in the ATP state (Davies *et al.*, 2008) (Fig 1.2). Additional movement induced by ATP hydrolysis involves the relative rotations of the D1 and D2 ring (Sato *et al.*, 2019). Through these movements, modelling based on mammalian structural information predicted that the pore size and distance between the D1 and D2 rings may change (Na & Song, 2016). The exact molecular mechanism of substrate processing and unfolding is not clearly understood, but it is thought that the movement of the ATPase rings and N domain account for substrate processing through the central pore. In addition, there exists a 76 amino acid long C-terminal extension, involved in regulating ATPase activity and adapter protein binding, which may also be modified by phosphorylation (Niwa *et al.*, 2012).

In eukaryotes, Cdc48 has been shown to participate in numerous cellular processes, including: protein homeostasis, membrane fusion, vesicular trafficking, DNA replication, chromatin degradation, apoptosis, autophagy, and immunity (Bègue *et al.*, 2017; Hänzelmann & Schindelin, 2017). In acting as a molecular chaperone, Cdc48 extracts or disassembles substrate proteins modified with ubiquitin, typically either from protein complexes,

membranes, or chromatin. Substrate proteins are typically then processed and degraded by the proteasome, but this isn't always the case: for example, in chromatin associated processes (e.g. DNA repair and transcriptional regulation), substrates may instead be recycled (Franz *et al.*, 2016).

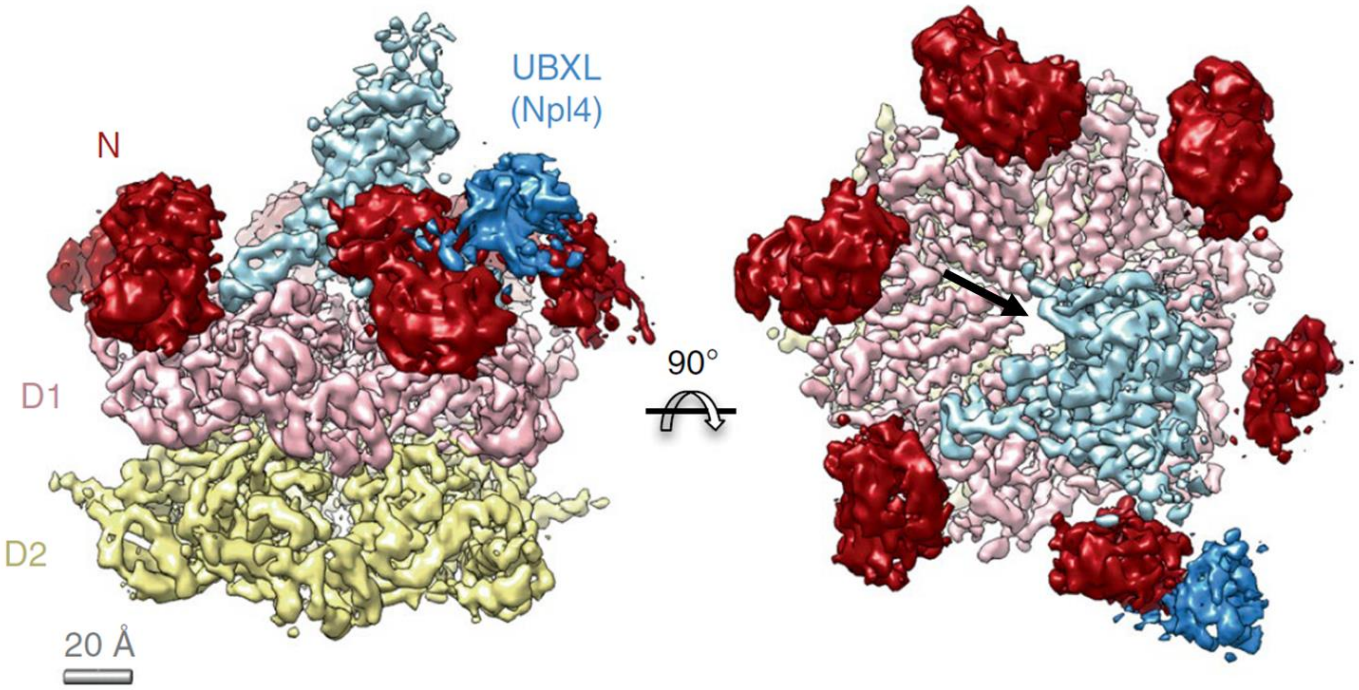


Figure 1.2 Adapted from Bodnar *et al*. 2018: cryo-electron microscopy density maps displaying structural information on the assembled Cdc48-Npl4-Ufd1 complex, using protein sequences from *Chaetomium thermophilum*. The left hand side image shows a side profile of the Cdc48 hexamer, with a 90° rotation in the right hand side image. The structure of the Cdc48 hexamer contains two concentric ATPase rings, with a lower D2 ring (green) and the D1 ring (grey) stacked on top of it. The N-terminal domains of Cdc48 (red) were captured in the 'up' state relative to the D1 ring, as the complex was assembled *in vitro* supplemented with ATP-γS. The central pore of Cdc48, through which substrates are unfolded and pulled through, is visible in the centre of the right hand image, indicated by a black arrow. The tower-like structure on top of Cdc48, was assigned as the zinc finger, MPN and C terminal domains of Npl4; also believed to be present there, is a C-terminal section of Ufd1

containing an Npl4 binding motif (as several Npl4 peptides in this region were protected in the presence of Ufd1, during hydrogen deuterium exchange mass spectrometry). The UT3 domain of Ufd1 was not visible in these density maps, presumably because of the flexibility of the domain, and it was assumed that a ubiquitinated substrate was required to fix it in place. The region bound to the N domain of Cdc48 (blue), was determined to be the UBXL domain of Npl4. This was positioned in a hydrophobic cleft of the Cdc48 N domain.

1.9.2 ERAD

The functions of Cdc48 have been extensively studied in relation to ERAD (Christianson & Ye, 2014), due to its relevance in human diseases – though the understanding of ERAD in plants is not as mature. ERAD is of particular importance here also, due to shared conceptual challenges with CHLORAD. In ERAD, terminally misfolded proteins are polyubiquitinated and retrotranslocated to the cytosol for degradation by the proteasome. The model organism for ERAD studies has been *Saccharomyces cerevisiae*, which will be the default source of information in the following descriptions – known information in plants will be highlighted where appropriate. There are three different ERAD pathways, dependent on whether the misfolded lesion is located in the ER lumen, membrane or cytosolic side of the ER, each of which involve a different E3 ligase (Wu & Rapoport, 2018). To focus on the membrane pathway, it is thought that substrates enter the 6 transmembrane spanning retrotranslocon Hrd1 (HMG-CoA reductase degradation 1) laterally through the membrane (Wu & Rapoport, 2018). Hrd1 has dual functionality, as it is also an E3 ligase, with two functionally redundant homologs in *Arabidopsis* (Liu & Li, 2014). Ubiquitination occurs on the cytosolic side of the protein, with the ER membrane anchored protein Cue1 (coupling of ubiquitin conjugation to ER degradation 1) (Hirsch *et al.*, 2009) activating a E2 cytosolic conjugase (e.g. Ubc7) and the E3 Hrd1 (Metzger *et al.*, 2013). Hrd1 has been shown to cooperate with membrane proteins Hrd3, and occasionally Usa1, in a complex (Horn *et al.*, 2009). Cdc48 is brought to the complex by the membrane recruitment protein Ubx2 (Neuber *et al.*, 2005). Cdc48 interaction with ubiquitinated substrates is mediated by a heterodimer of adapter proteins Npl4 (nuclear

protein localisation 4) and Ufd1 (ubiquitin fusion degradation 1) (Braun *et al.*, 2002). Following retrotranslocation and unfolding via motive force from Cdc48, the substrate is likely processed by a deubiquitinase, such as Otu1 (ovarian tumour 1) in both yeast and *Arabidopsis* (Stein *et al.*, 2014; Zang *et al.*, 2020) – deubiquitinase activity may be required to cleave the polyubiquitin chain following unfolding through Cdc48, otherwise continued engagement with the Cdc48 complex is possible (Ji *et al.*, 2022). Finally, the unfolded substrate is presumed to be further processed before degradation by the proteasome – though these later steps are not clearly understood, even in yeast.

1.9.3 Regulation of Cdc48 Activity

The functions of Cdc48 are regulated through two systems: adapter proteins binding to either the N or C terminal domains via distinct binding motifs, or post-translational modifications (PTMs) (Hänzelmann & Schindelin, 2017). Mammalian Cdc48 has been identified as having a range of PTMs, including SUMOylation, ubiquitination, acetylation, and phosphorylation, through proteomics (Fang *et al.*, 2016). Such PTMs are believed to play a role in mammalian Cdc48 complex assembly, stability, binding site competition and affinity (Venne *et al.*, 2014). While there is considerable data relating to the position and types of these PTMs, the understanding of the physiological significance resulting from various modifications is still quite poor.

Thus far, roughly 40 different adapter proteins are known to exist, from yeast and mammalian literature (Buchberger *et al.*, 2015). Accordingly, Cdc48 would appear to have broadly strong functional conservation across taxa to allow for consistent interactions with a wide range of partners, with similar crystal structures produced in studies of yeast and mammalian proteins (Bodnar *et al.*, 2018; Gao *et al.*, 2022). The diversity of these adapters allows Cdc48 to operate in various cellular locations and processes. It is likely that adapter proteins are tightly regulated to ensure Cdc48 functions adequately, but thus far this aspect is not well examined. These adapters essentially form three functional groups: substrate recruiting (e.g. Npl4-Ufd1 to the

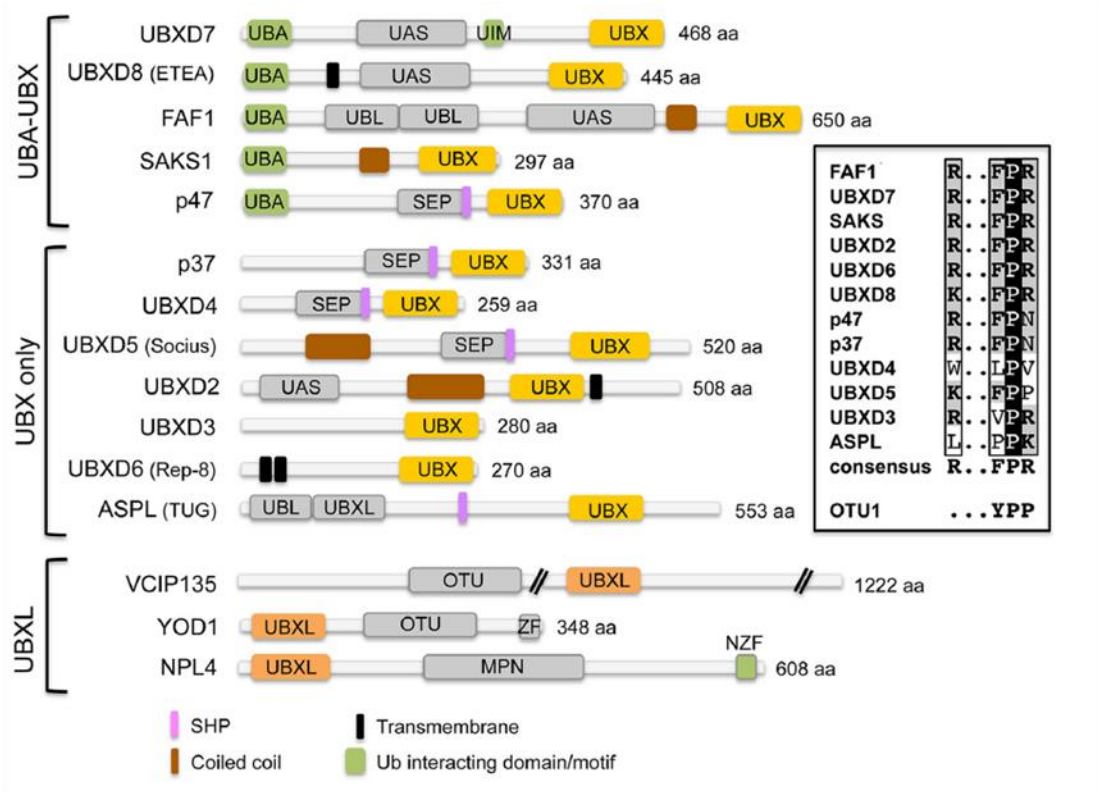
UPS), substrate processing (e.g. the Otu1 deubiquitinase), and Cdc48 regulators (e.g. SVIP to recycle Cdc48). Only a minority of these proteins are known to bind to Cdc48 via the unstructured C-terminus, and these include proteins possessing PUB (PNGase/UBA or UBX containing proteins) (Allen *et al.*, 2006) or PUL (PLAP, Ufd3p, Lub1p) (Zhao *et al.*, 2007) domains. The majority of adapters bind to the N domain of Cdc48 via domains of UBX (ubiquitin regulatory X), UBXL (UBX-like), or motifs of VBM (VCP-binding motif), VIM (VCP interacting motif) and the SHP box (binding site 1) (Hänzelmann & Schindelin, 2017) (Fig. 1.3). The UBX domain is a ubiquitin-like β-grasp fold and is found in the majority of Cdc48 adapter proteins; at least 13 UBX proteins have been demonstrated to interact with Cdc48 in mammals, binding to a cleft in the N domain (Kloppsteck *et al.*, 2012). The structurally similar UBXL domain (though there is little sequence similarity) is found in a few deubiquitinates, as well as Npl4 (Bruderer *et al.*, 2004) (Fig. 1.3). The VIM motif is formed by a short α-helix comprising of positively charged residues, separated by a series of alanine residues (Stapf *et al.*, 2011). It is found in a functionally diverse range of adapters, including the substrate binding protein in mitochondrial associated degradation Vms1 (VCP mitochondrial stress-responsive 1) (Heo *et al.*, 2010) (Fig. 1.3). The VBM motif consists of a short stretch of positively charged amino acids and is present in the E3 ligase Hrd1 (Buchberger *et al.*, 2015) (Fig. 1.3). The VBM motif has a far lower affinity than the VIM motif, and they compete for binding to the N domain cleft (Liu *et al.*, 2013). Finally, the SHP box motif is notably present in Ufd1 (Fig. 1.3). While mostly found in unstructured regions, the conserved motif typically adopts a slight kink; the SHP box has been shown to bind to a distinct area of the N domain compared to the UBX domain and VIM/VBM motifs (Le *et al.*, 2016).

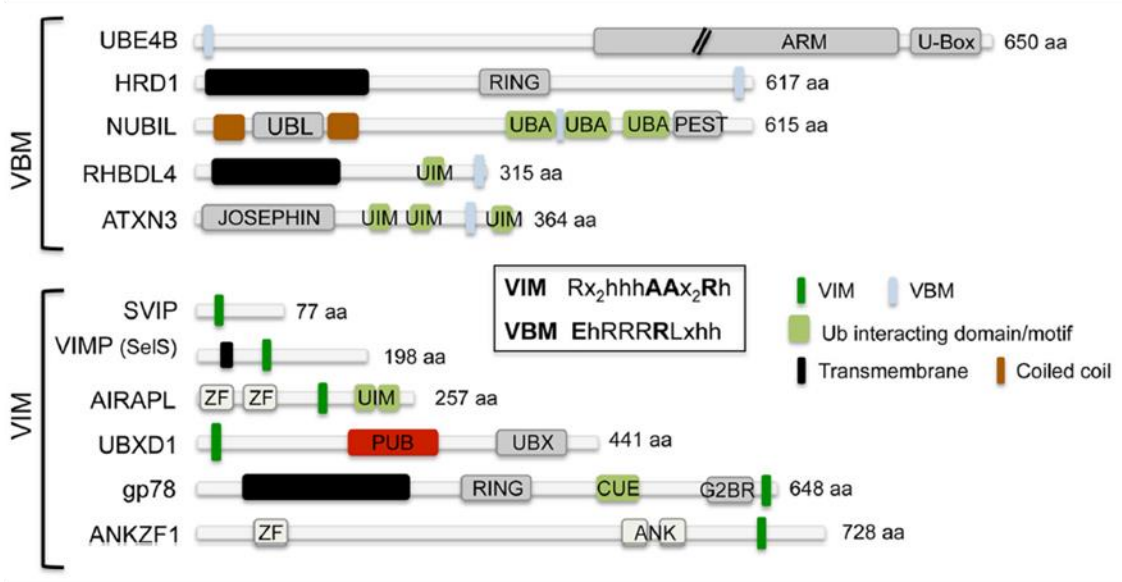
Due to the overlap in binding sites, there is often competition between adapter proteins, for instance between p47 and Npl4-Ufd1 (Bruderer *et al.*, 2004). However, most studies examining competition utilise purified proteins *in vitro*, and so it is feasible that there are unknown regulatory processes *in vivo* that act on binding site competition. Adapter proteins can adopt a range of stoichiometries in binding to Cdc48; for instance, the VIM protein SVIP

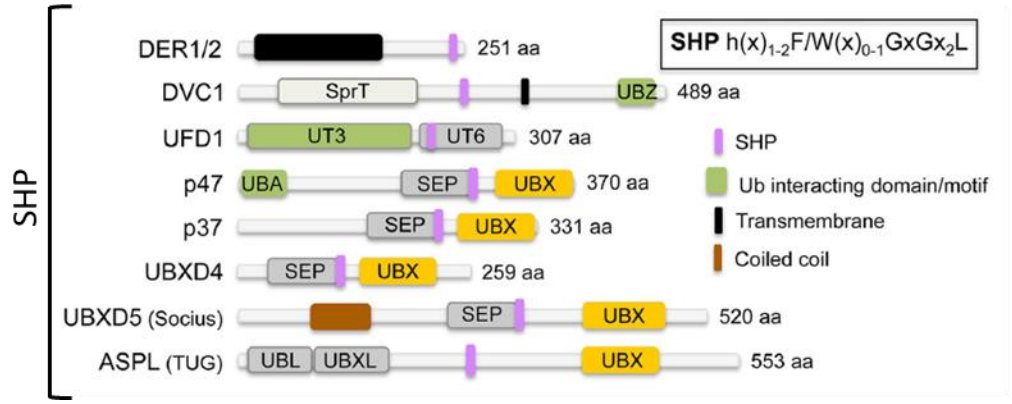
may bind to all six N domains (Hänzelmann & Schindelin, 2011). In the case of Npl4-Ufd1 (NU), with a bipartite binding ability, it was speculated whether it bound to the same or different N domains. Structural determination of the Cdc48-Npl4-Ufd1 complex by cryogenic electron microscopy, demonstrated that one copy of the NU heterodimer binds to opposite N domains of Cdc48 – in fact, Ufd1 may bind to two adjacent N domains simultaneously via two SHP boxes (Bodnar *et al.*, 2018).

The *Arabidopsis* genome encodes 15 UBX domain containing proteins, called plant UBX (Pux) proteins (Liu & Li, 2014). Kretzschmar *et al.* (Kretzschmar *et al.*, 2018) highlighted the role of Pux10 in protein extraction from lipid droplets, through tethering of Cdc48; notably they observe that Pux10 localises to the chloroplast OEM later in development. ERAD in yeast involves the UBX-containing protein Ubx2; this has been identified as a membrane-embedded tethering factor, as defects in Ubx2 caused impairments to ERAD (Schuberth & Buchberger, 2005). Pux10 may function in CHLORAD in an analogous manner.

Figure 1.3 (overleaf) Adapted from Hänzelmann & Schindelin 2017: An overview of the diversity of N-domain interacting Cdc48 adapter proteins in *Homo sapiens*. Domain architectures for each human protein sequence is shown, displaying functional domains known from experimental literature or from bioinformatic annotations by the authors. Sequences are grouped by their type Cdc48 interacting domain. Consensus sequence alignments for motifs in UBX, VIM, VBM, and SHP are displayed.


UBA-UBX
UBXD7
UBA
UAS
UIM
UBX
468 aa
UBXD8 (ETEA)
445 aa
FAF1
UBL
650 aa
SAKS1
297 aa
p47
SEP
370 aa
UBX only
p37
331 aa
UBXD4
259 aa
UBXD5 (Socius)
520 aa
UBXD2
508 aa
UBXD3
280 aa
UBXD6 (Rep-8)
270 aa
ASPL (TUG)
UBXL
553 aa
VCIP135
OTU
UBXL
1222 aa
YOD1
ZF
348 aa
NPL4
MPN
NZF
608 aa
FAF1 R..FPR
UBXD7 R..FPR
SAKS R..FPR
UBXD2 R..FPR
UBXD6 R..FPR
UBXD8 K..FPR
p47 R..FPN
p37 R..FPN
UBXD4 W..LPV
UBXD5 K..FPP
UBXD3 R..VPR
ASPL L..PPK
consensus R..FPR
OTU1 ...YPP
SHP
Transmembrane
Coiled coil
Ub interacting domain/motif


VBM
UBE4B
ARM
U-Box
650 aa
HRD1
RING
617 aa
NUBIL
UBL
UBA
PEST
615 aa
RHBDL4
UIM
315 aa
ATXN3
JOSEPHIN
364 aa
VIM
SVIP
77 aa
VIMP (SelS)
198 aa
AIRAPL
ZF
257 aa
UBXD1
PUB
441 aa
gp78
CUE
G2BR
648 aa
ANKZF1
ANK
728 aa
VIM Rx2hhhAAx2Rh
VBM EhRRRRLxhh
VIM
VBM
Ub interacting domain/motif
Transmembrane
Coiled coil


SHP
DER1/2
251 aa
DVC1
SprT
UBZ
489 aa
UFD1
UT3
UT6
307 aa
p47
370 aa
p37
331 aa
UBXD4
259 aa
UBXD5 (Socius)
520 aa
ASPL (TUG)
553 aa
SHP h(x)1-2F/W(x)0-1GxGx2L
SHP
Ub interacting domain/motif
Transmembrane
Coiled coil

1.10 Project Rationale: Npl4 and Ufd1 are Likely Participants in CHLORAD

Cdc48 is evidently an important component of CHLORAD (Li *et al.*, 2022; Ling *et al.*, 2019; Sun *et al.*, 2022). The question of how Cdc48 is recruited to function in CHLORAD has not been addressed previously. In UPS processes, Cdc48 interactions with ubiquitinated substrates are typically mediated by the heterodimer of Npl4 and Ufd1 (Bègue *et al.*, 2017). In these cases, it is the NU heterodimer that directly interacts with ubiquitin moieties to facilitate their unfolding from membranes or complexes by Cdc48 (Twomey *et al.*, 2019).

I proposed that Npl4 and Ufd1 are likely candidates to facilitate Cdc48's involvement in CHLORAD. This heterodimer is necessary in ERAD to allow Cdc48 to retrotranslocate proteins (N. O. Bodnar & T. A. Rapoport, 2017). Another likely candidate might have been the adapter proteins of mitochondria-associated protein degradation (MAD): Npl4 and Vms1 (Heo *et al.*, 2010). However, the exact role of Vms1 is unclear and it may not be necessary for protein degradation (Esaki & Ogura, 2012).

1.10.1 Model for Cdc48-Npl4-Ufd1 Substrate Extraction

Mechanistic understanding of the Cdc48-Npl4-Ufd1 complex has arisen from biochemical and structural analysis, leading to the development of a potential model of operation (Ji *et al.*, 2022; Twomey *et al.*, 2019) (Fig 1.4) — a simplified explanation is offered below.

Initially, the Cdc48 hexamer and Npl4-Ufd1 heterodimer form independently (Fig. 1.4.1). The NU heterodimer forms as a result of an interaction between the C-terminal NBM (Npl4 binding motif) domain of Ufd1 (Le *et al.*, 2016) and the central MPN (Mpr1/Pad1 N-terminal) domain of Npl4 prior to binding to Cdc48 (Bodnar *et al.*, 2018; Nguyen *et al.*, 2022; Sato *et al.*, 2019). Cdc48 and NU heterodimer assemble to form the initial interaction with the polyubiquitinated chain (bound to a substrate) without the need for ATP (Fig. 1.4.1). There are multiple binding sites present across the heterodimer for folded ubiquitin molecules; these have been observed

in multiple studies at the N terminal region of Ufd1 and the C terminal region of Npl4 (Le *et al.*, 2016; Sato *et al.*, 2019; Twomey *et al.*, 2019). Additionally, there are zinc finger domains in yeast which bound to the D1 ring of Cdc48 (Bodnar *et al.*, 2018) — these were found to be functional in the unfolding activity of Cdc48 in yeast, as reduced unfolding activity was observed when both CHCC type motifs were mutated in the domain (Bodnar *et al.*, 2018). The initiator ubiquitin in the chain which binds to the complex is random, but must contain several folded ubiquitin moieties upstream in the chain, which bind to the Npl4 tower and the UT3 domain of Ufd1 (Fig. 1.4.2). The initiator ubiquitin is spontaneously unfolded by thermal processes, a portion then binds to a groove in Npl4 and the N-terminus of the ubiquitin inserts into the D2 ring (which shifts the equilibrium to the unfolded state). Atomic force microscopy has demonstrated that unfolding of ubiquitin may be initiated by force applied to the N-terminal Lys^{48} residue (Carrion-Vazquez *et al.*, 2003), which binds to the Npl4 groove; proximity to the Cdc48 pore may then be sufficient to disrupt and unfold the remaining ubiquitin molecule's β sheets (Irback *et al.*, 2005).

The unfolded ubiquitin molecule was also observed to engage with loops located in the D2 ring, which may act as signal to engage ATP binding. ATP hydrolysis drives conformational movements of the ring (and thus the loops), pulling and unfolding the subsequent residues in the chain through interactions with aromatic groups in the central pore (Ji *et al.*, 2022) (Fig. 1.4.2-3). Throughout this process, there may been to be at least three folded ubiquitin molecules bound to NU to secure the substrate; the N domain is also expected remain in a conformation above the D1 ring. Translocation of the initiator ubiquitin into the D2 ring is expected to dislodge it from the Npl4 groove, which in turn releases the Ub1+ ubiquitins from Npl4, while Ub1- are expected to remain attached to the UT3 domain of Ufd1 (Fig. 1.4.3-4). Translocation through the Cdc48 pore continues by a remarkable action: the C-terminus of the initiator ubiquitin is translocated with the K48 branchpoint attached, Ub1+ ubiquitins were determined to remain outside of the Cdc48 pore, in an unfolded conformation (Fig. 1.4.5). This is expected to be accomplished by partial lateral opening of protomers of the Cdc48 hexamer,

which is theoretically possible as opening of the protomers is feasible by the observation of spiral structures using cryo-EM (Ji *et al.*, 2022; Pan *et al.*, 2021; Twomey *et al.*, 2019). Cdc48 then translocates the proximal K48 linked Ub1- ubiquitin. The branch point in the chain would presumably necessitate the simultaneous translocation of two polypeptide chains at this point (Fig. 1.4.6). It is hypothesised that this occurs, as translocation of single chains reiteratively is energetically wasteful, and because the related ATPase Vps4 can accommodate two strands (Han *et al.*, 2019). Following translocation, proximal Ub1- ubiquitins are refolded, while the substrate remains unfolded (Fig. 1.4.7-8). The presence of a deubiquitinase is seemingly necessary to cleave the folded distal Ub1+ ubiquitin molecules, otherwise reiterative engagement with the Cdc48 complex is possible (not shown in model) (N. O. Bodnar & T. A. Rapoport, 2017; Ji *et al.*, 2022). The unfolded substrate may engage with shuttling factors or be directly degraded by the 26SP (Fig. 1.4.9).

While this model is detailed and seemingly in agreement with existing literature, it still requires much additional evidence. It is also limited by structural resolution of the D1 ring, hence there is a poor understanding of its involvement. Yet, at least one conclusion can be drawn from multiple lines of evidence: the Npl4-Ufd1 heterodimer is a vital component of this process.

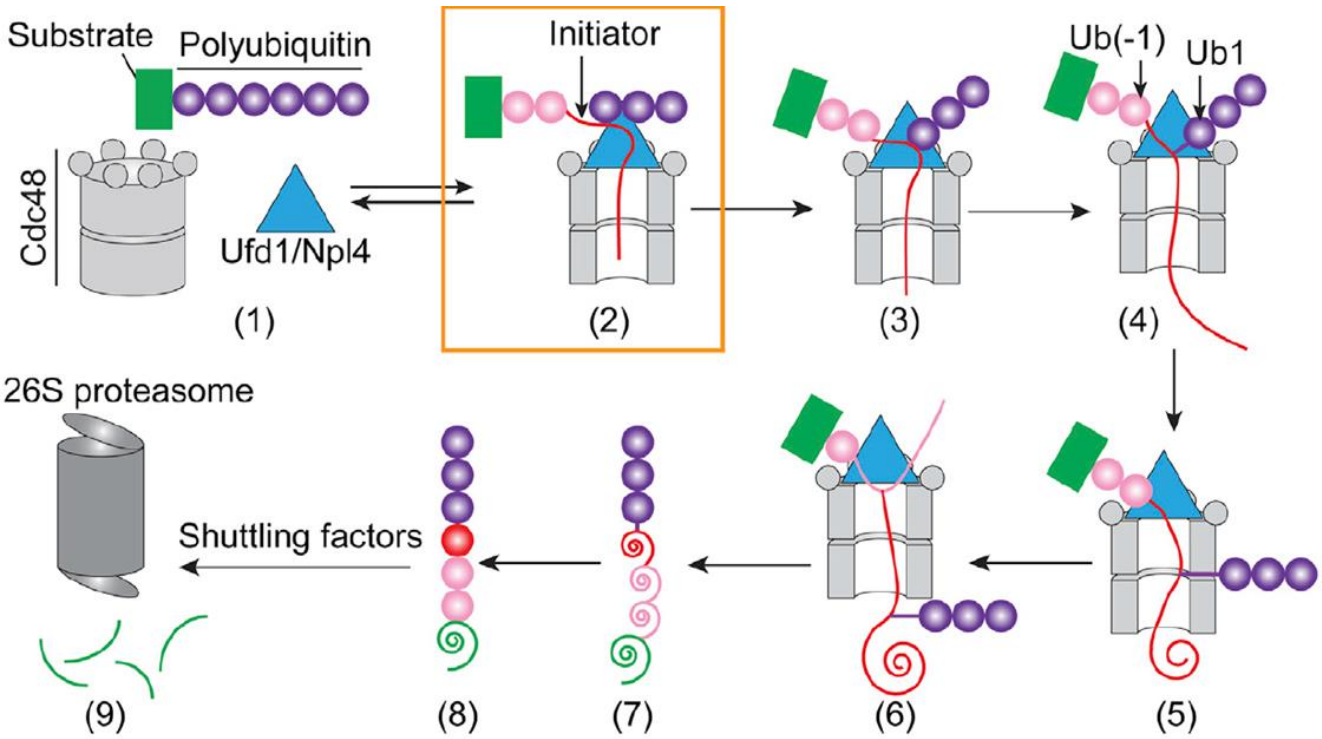


Figure 1.4 Adapted from Ji *et al.* 2022: a mechanistic model for substrate processing based on *in vitro* experiments using purified Cdc48, Npl4, Ufd1, and ubiquitinated substrate proteins from *Saccharomyces cerevisiae*. The boxed initiation stage is most evidenced by structural methods. The initiating ubiquitin moiety is coloured red, while proximal and distal ubiquitins are shown as pink and purple circles respectively. Translocated and unfolded amino acid sequences are shown as spirals.

1.10.2 The Current Understanding of Npl4 and Ufd1 in *Arabidopsis*

In *Arabidopsis thaliana*, there have been reported to be two isoforms of Npl4 and four isoforms of Ufd1. These were identified by BLAST searching of mammalian ERAD protein sequences (Liu & Li, 2014). Thus far, publications exist pertaining to both Npl4 genes, and all bar one variant of Ufd1 (At4g15420). Studies of these proteins in plants are limited, with their inclusion in only three publications to date. In the existing literature, Npl4 and Ufd1 isoforms have been given the following names (Li *et al.*, 2022; Mérai *et al.*, 2014): AT2g47970, NPL4A; AT3g63000, NPL4B; AT2g29070, UFD1A; AT2g21270, UFD1B; AT4g38930, UFD1D; and At4g15420 is not named.

Galvão *et al.* (Galvão *et al.*, 2008) investigated the functions of kinases AtPI4Kγ4 and AtPI4Kγ7, predicted to encode type II PI4Ks (phosphoinositide4-kinase), in *Arabidopsis*. These proteins undergo autophosphorylation and phosphorylate serine and threonine residues of proteins. Both proteins however contain N-terminal ubiquitin-like (UBL) domains. UBL domain containing proteins often interact with components of the UPS. Through co-immunoprecipitations followed by mass spectrometry, the authors discovered that the kinase AtPI4Kγ4 interacts with the UPS proteins UFD1A, and RPN10 (regulatory particle non-ATPase 10). In addition, AtPI4Kγ4 phosphorylates the C-terminal of Ufd1, which may regulate its binding affinity and activity in the UPS.

Mérai *et al*. (Mérai *et al.*, 2014) describe Cdc48's function in degrading sumoylated centromeres in the nucleus. They found that NPL4A and NPL4B function in this with a degree of redundancy, with only the double knock-out mutant phenocopying a *cdc48A* mutant's effect on centromere condensation. The authors of the latter study additionally observed interaction of CDC48A with NPL4A, NPL4B, UFD1A and UFD1B by mass spectrometry, following co-immunoprecipitation of CDC48A isolated from pollen.

A highly relevant publication was released earlier this year, unveiling new experimental evidence of the functions of Npl4 and Ufd1 in chloroplast biology. Internal chloroplast proteins RbcL and AtpB are reportedly ubiquitinated and interact with a complex of Cdc48-Npl4-Ufd1 (Li *et al.*, 2022). The headline result of this publication may be that of the ubiquitination of internal chloroplast proteins; this result was further evidenced by a publication from the Jarvis group later in the year (Sun *et al.*, 2022).

Isolated chloroplast lysates were probed with anti-UBQ11 antibody, a marker for ubiquitination, which revealed patterns of ubiquitination in membrane envelope, stromal, and thylakoid membrane fractions. Inhibition of the proteasome resulted in increased levels of ubiquitination, suggesting that chloroplast proteins were degraded by the UPS. Single and double knock-out mutants of isoforms were examined to see the impact of reduced *NPL4A*, *NPL4B*, *UFD1B* and *UFD1C* expression on ubiquitination. There was no observed impact on chloroplast ubiquitination in the single mutants, however there were elevated levels in the double mutants of *npl4a npl4b* and *ufd1b ufd1c* – this was also observed for an RNAi line of *CDC48A*. These data suggest redundancy between Npl4 and Ufd1 isoforms, and the involvement of Cdc48 to downregulate the quantity of ubiquitinated proteins from the chloroplast – likely via their retrotranslocation and degradation by the UPS, as inhibition of the UPS also resulted in increased quantities of ubiquitinated chloroplast proteins.

To examine interactions with internal chloroplast proteins, the authors chose to focus on RbcL (Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit) and AtpB (adenosine triphosphate (ATP) synthase subunit beta) (Li *et al.*, 2022). The justification for this was that these two proteins, as well as AtpA, were the most abundant chloroplast genome encoded proteins by mass spectrometry, following co-immunoprecipitation (Co-IP) of overexpressed NPL4A-GFP. The high abundance of RbcL is a concerning factor (Mergner *et al.*, 2020), as the chance of type I error is increased, as evidenced by its detection in their WT samples in the supplementary data (Li *et al.*, 2022). Nonetheless, interactions between NPL4A and

UFD1B with AtpB and RbcL were sustained by three independent measures (Li *et al.*, 2022); further interactions were not observed with AtpA.

Ubiquitination of RbcL and AtpB are enhanced by *Npl4* and *Ufd1,* as determined by a tandem ubiquitin-binding entity (TUBE2) assay using double knock-out mutants (Hjerpe *et al.*, 2009). To examine an aspect of the functional significance of Npl4 and Ufd1 in chloroplast biology, the authors used methyl viologen (MV) as a generator of superoxide to create ROS (reactive oxygen species) stress. As before, single mutants were indistinguishable from WT, but double mutants demonstrated reduced chlorophyll content and photosynthetic performance. They also describe increased transcription and expression of Npl4 and Ufd1 in response to ROS stress from MV (approximately 2-fold increase relative to WT). They then examined transcription and expression in two mutant lines which generate endogenous ROS stress: *ppo1-1* (protoporphyrinogen IX oxidase 1) and *fc2* (ferrochelatase 2) (Li *et al.*, 2019; Woodson *et al.*, 2015). Here they only observed upregulation in *ppo1-1,* though they do not provide statistical analysis of these data nor quantification of the blots again (Li et al. 2022, Fig. 4).

The publication from Li *et al.* (2022) has substantial overlap with the scope of this project. The timing of its publication means that it is a useful resource for the discussion of results, but that it had no bearing on the planning or execution of this project. Therefore, the results in the subsequent chapters are presented regardless of these new findings, but are discussed in their context.

It is clear from existing literature that the functions of Npl4 and Ufd1 are not holistically understood, especially in how they operate in plants. There was also no existing evidence for their participation in CHLORAD. Based on the importance of CHLORAD and of Cdc48-Npl4-Ufd1 in UPS processes, I believed this to be viable and worthy area of research. I therefore sought to establish whether Npl4 and Ufd1 proteins had a role in CHLORAD in *Arabidopsis*.

1.11 General Aims of Research

The overarching purpose of my project was to determine whether any of the Npl4 and Ufd1 isoforms functioned in CHLORAD, and to what extent they were necessary or adaptive components. The depth of research into these proteins in other eukaryotic organisms is highly beneficial, allowing me to make assumptions about many molecular characteristics. However, the evolutionary distance between these and the plant proteins means that distinct properties and functions may have been acquired.

Bioinformatic analysis thus formed the initial part of the investigation. I aimed to understand the proteins' domain architectures and predicted structures: to predict whether there may be functional or operational differences based on structural homology. I also sought to complete phylogenetic analysis to examine sequence relationships within the Viridiplantae. An intriguing question lies in the number of Npl4 and Ufd1 copies in *Arabidopsis*. Examination across plant species may reveal to what extent multiple copies are adaptive or simply the product of non-functional processes.

The first experimental evidence aimed to establish that Npl4 and Ufd1 have the capacity to, and do, physically interact with CHLORAD proteins and substrates. This was addressed by analysing their subcellular location of Npl4 and Ufd1 proteins in cells and *in planta*; I also aimed to determine whether localisation could be convincingly determined at the chloroplast OEM. Additionally, the localisation of those interactions were examined with fluorescence microscopy. To validate observations of protein-protein interactions, I planned to perform tests of physical interaction using the techniques of co-immunoprecipitation, and tandem affinity purification (TAP) followed by mass spectrometry.

Finally, I aimed to examine the functional significance of Npl4 and Ufd1 in CHLORAD using mutant analysis. Genetic crosses with knock-out and overexpressing mutants were planned

to TOC and CHLORAD mutants, to determine whether phenotypic suppression was visible. These may also reveal the extent of redundancy between isoforms, and in turn if any set of proteins has unique functions in CHLORAD. Experiments were planned to investigate chloroplast physiology, particularly in the contexts of environmental stress and development. These would allow me to determine the adaptive significance of Npl4 and Ufd1 genes in CHLORAD

Chapter II

Results I: Bioinformatics and Phylogenetics

2.1 Abstract

The UPS is an important regulatory system in plants, comprised of thousands of proteins in *Arabidopsis*. As with the CHLORAD UPS protein Cdc48, Npl4 and Ufd1 share multiple isoforms in *Arabidopsis*. I endeavoured to begin by analysis be examining the information present in their sequences through bioinformatic and phylogenetic techniques. Through this, I first determined that Npl4 and Ufd1 proteins are likely all soluble, with the strongest predictions for their localisation being in the nucleus and cytoplasm. I then examined potential interacting proteins through curated databases, identifying many UPS proteins as predicted functional partners. All Npl4 and Ufd1 genes show medium to high gene expression throughout development, and there are no major perturbations which affect expression. Next, phylogenetic analysis revealed that Npl4 proteins are likely the result of a recent genome duplication. Ufd1 proteins are more evolutionary distant, forming three paralogous groups. This may suggest increased functional diversity within Ufd1 proteins. Through a series of bioinformatic tools, I generated functional domain annotations; comparative analysis with yeast proteins shows many shared features. These annotations were broadly supported through predictions of protein structure via Alphafold. Through these analyses, it became clear that one Ufd1 variant (called Ufd1C) showed substantial deviation in structure when compared to the other Ufd1 proteins, and it may be questionable whether it is a true Ufd1 homologue. Unique to Ufd1C is a zinc finger domain. Here I demonstrate that it is a TRAF type zinc finger with a C3HC4 motif. It likely has a function in protein binding, but it is unclear what significance exists beyond this. In conclusion, sequence analysis of Npl4 and Ufd1 proteins demonstrate that they meet many requirements for participation in CHLORAD, through their predicted

structures and functions. However, no co-expression or predicted interactions show any involvement specially in chloroplast related processes.

2.2 Introduction

The UPS has emerged as an important regulator of protein homeostasis in plants. Its importance and complexity is reflected in plant genomes, with UPS components comprising a startling 6% of the predicted *Arabidopsis thaliana* proteome (Vierstra, 2009). This is reflected in the retained number of isoforms for Cdc48, Npl4 and Ufd1. Cdc48 is a highly conserved protein found within all eukaryotes; there are also several predicted bacterial homologs, though they share low sequence identity (Barthelme & Sauer, 2016). In *Arabidopsis*, there are five sequences of Cdc48 (CDC48 A, B, C, D, and E) (Bègue *et al.*, 2017). These are represented across plant taxa, though they can be differentiated into two groups. The first contains CDC48A, B and C, and these share highly conserved features and sequences with yeast and mammalian Cdc48 (such as N terminal domains). The second consists of CDC48D and E, which only share ATPase domains, and have considerably more structural diversity at their N- and C- termini. These may also share homologs with other eukaryotic proteins, such as mammalian NVL2 (nuclear VCP-Like 2) which is involved in ribosome biogenesis in the nucleus (Nagahama *et al.*, 2004).

It was discovered through TAP purification of AtToc33 and mass spectrometry, that CDC48A (referred to as simply ‘Cdc48’ hereon) may be an interacting partner. It subsequently emerged through careful molecular and genetic analysis, that Cdc48 operates in CHLORAD to provide the motive force to retrotranslocate proteins (Ling *et al.*, 2019). Studies examining the known UPS adapter proteins of Cdc48, Npl4 and Ufd1, are limited in plants (Li *et al.*, 2022; Mérai *et al.*, 2014); whereas studies in yeast and mammals are well established (N. Bodnar & T. Rapoport, 2017; Hänzelmann & Schindelin, 2017). Indeed, there has been no sequence analysis performed on these genes in *Arabidopsis*.

Like Cdc48, there are multiple predicted homologs of Npl4 and Ufd1 (Liu & Li, 2014). Through analysis of Npl4 and Ufd1 sequences in *Arabidopsis,* I hoped to garner more information about homology to other eukaryotic sequences, potential interactions, and functions.

2.3 Results and Discussion

2.3.1 Nomenclature

An important point to begin with are the differences in nomenclature between the existing literature and this thesis. My naming of Npl4 and Ufd1 proteins at the start of this project was not done with the knowledge of the existing precedent for the *Arabidopsis* proteins (Mérai *et al.*, 2014). I initially conducted TBLASTN searching of conserved Npl4 and Ufd1 domains from yeast (UBXL and UT3 respectively), to identify potential homologs in *Arabidopsis* (Table 2.1); through this, I was able to reproduce the identification of *Npl4* and *Ufd1* genes by Liu & Li (2014). For internal consistency, the names I originally chose were retained in the generation of all materials and experiments throughout this project. In the discussion of the presented results hereon, my own nomenclature will be favoured (Table 2.1): Npl4 (At2g47970), Npl4L (At3g63000), Ufd1A (At2g21270), Ufd1B (At2g29070), Ufd1C (At4g15420), and Ufd1D (At4g38930).

TAIR Code	This Thesis	Mérai *et al.* 2014
At2g47970	Npl4	NPL4A
At3g63000	Npl4L	NPL4B
At2g21270	Ufd1A	UFD1B
At2g29070	Ufd1B	UFD1A
At4g15420	Ufd1C	Not defined
At4g38930	Ufd1D	UFD1C

Table 2.1 Nomenclature of *Arabidopsis* Npl4 and Ufd1 proteins in this thesis and their format in the existing literature; TAIR (the Arabidopsis information resource) codes show the shared identifying code.

2.3.2 Predicted Features and Gene Expression

Broadly, Cdc48 functions in the nucleus and cytoplasm – where it interacts with the surface of other organelles like the chloroplast (Bègue *et al.*, 2017; Hänzelmann & Schindelin, 2017; Ling *et al.*, 2019). Other eukaryotic homologs of Npl4 and Ufd1 are all soluble proteins. However, as there exist a range of membrane bound Cdc48 adapters, it is not entirely inconceivable that such domains may have been acquired during evolution. I therefore examined whether Npl4 and Ufd1 proteins possessed transmembrane domains. ARAMEMNON (Schwacke *et al.*, 2003) is a database containing over 20 membrane topology predictors. It may therefore be considered a reasonably reliable resource for topology prediction, through presenting the output of multiple predictive algorithms. I found that all Ufd1 and Npl4 isoforms contained no predicted transmembrane spanning regions in the consensus output. It is therefore likely that all Npl4 and Ufd1 isoforms are soluble proteins. There were some hydrophobic α-helical regions predicted by the HmmTop_V2 algorithm: these include an N terminal region of Ufd1C, and central region of Ufd1A, B and D. However, HmmTop predicts the highest estimates of transmembrane proteins in *Arabidopsis* (Schwacke *et al.*, 2003); it is therefore likely to overestimate transmembrane abundance, important in this case as it is the sole algorithm producing said prediction.

In aiming to predict the subcellular location of Npl4 and Ufd1, I examined the predictive database SUBA4 (Hooper et al. 2017). SUBA4 incorporates localisation data from existing literature: examining fluorescent proteins, mass spectrometry, and various protein-protein interaction experiments. It also incorporates over 22 subcellular localisation prediction algorithms to produce predictions of likely subcellular compartments. Experimental data from mass spectrometry was incorporated, but unexpectedly this did not include Mérai *et al.* (2014). Instead these data resulted from a range of proteomic studies. These found that Ufd1A could be detected in the nucleus (Goto *et al.*, 2019), the extracellular matrix (cell wall) (Kraner *et al.*, 2017), cytosol (Ito *et al.*, 2011), and plasma membrane (de Michele *et al.*,

2016). In addition, there were proteomic data to suggest cytosolic localisations for Npl4 and Npl4L (Ito *et al.*, 2011). The predictive scores for all isoforms are displayed in Table 2.2; this output provides as scores for different subcellular compartments, relative to a max, which may be interpreted as the most likely location. On this basis, the cytoplasm is the most likely subcellular location for Npl4, Npl4L and Ufd1A. SUBA4 finds that the nucleus is more likely for Ufd1B, Ufd1C and Ufd1D. The score for Ufd1A tallies accordingly with the incorporated proteomic data. It is not surprising that the scores are low for the plastid for any of these proteins, as they possess no targeting signals – examined using TargetP V2.0 (Emanuelsson *et al.*, 2007). What is perhaps more surprising are the low scores for the endoplasmic reticulum, and for the nucleus for both Npl4 and Npl4L. This is a concerning result and highlights the fact that predictive databases are never wholly reliable – given the existing publication demonstrating a nuclear localisation for Npl4 and Npl4L and its involvement in an exclusively nuclear process (Mérai *et al.*, 2014). There is also a higher score associated with the mitochondrion for Npl4, but it is unclear what the exact basis for this score is. There does not appear to be any targeting signal, though it may be related to co-expression; for instance, with substrates or functional partners in MAD. However, if this were the case, it would be surprising that components in ERAD were not also detected. Ultimately, experimentation is required to validate these observations.

Potential protein-protein interactions were then initially predicted using the BAR *Arabidopsis* interactions viewer database (Fucile *et al.*, 2011). This tool provided no information about Ufd1B and Ufd1C. The predictions given predominantly lacked annotation and experimental support; thus the list of proteins identified by BAR is not presented, but described briefly below. Npl4 was predicted to interact with 40 proteins, involved in a wide variety of processes by GO (gene ontology) annotation; these were predominantly nuclear located. It was thus difficult to draw any meaningful conclusions from this interactome. Fewer interactions were found for Npl4L (14 proteins), notably finding Ufd1A and Ufd1D, as well as several nuclear proteins (e.g. Rpa2). The Ufd1A interactome (17 proteins) contained

Cdc48A, Cdc48C, Npl4, Npl4L, SUMO system proteins (e.g. Sae2) and again many nuclear proteins (e.g. Mad2). Ufd1D possessed many similar interactions to Ufd1A (e.g. Sae2 and Mad2) but was only predicted to interact with isoforms Npl4L and Cdc48C. The only identified ERAD proteins by BAR were found to interact with Ufd1D (Otu2 and Hrd1). Overall, these predicted interactions offer limited utility due to the inherently low reliability of the predictions without experimental support.

To pursue predicted protein-protein interactions further, I then turned to the STRING database (Szklarczyk *et al.*, 2021). The STRING database is constructed on similar principles to SUBA4: it uses literature mining, databases on interacting proteins, co-expression data, and the transfer of known interactions from one organism to another. Overall, it is a more comprehensive and well-curated database than BAR. The database was searched under the highest stringency conditions, and the highest ranked proteins are displayed in Table 2.2. The effect of literature mining and transfer of known interactions from other organisms, might risk the generation of some tautologous results. That could certainly be argued here, where I see heterodimer partners feature as the most likely interacting partners for Npl4 and Ufd1 isoforms respectively – are these real in *Arabidopsis*, or simply carry over from what is known in yeast and mammals? I also see the relevant isoforms of Cdc48 (A, B, and C) appearing with all proteins bar Ufd1B – which is a curious and non-obvious omission. In addition to these expected interactions, there are additional UPS components (Table 2.2). These include the E3 ligases RHA1A (RING-H2 FINGER A1A) and proteasomal subunits RPN10 and RPN13. In addition, the proteins MUB4 (membrane anchored ubiquitin fold 4) and MUB5 are repeatedly present across the isoforms. These are E2 interacting enzymes, with a ubiquitin like β grasp, with identified homology to yeast and human UPS proteins; they are thought to aid in the recruit of UPS activity at the plasma membrane (Dowil *et al.*, 2011). The repeated occurrence of these proteins may therefore be indicative of additional localisation of Npl4 and Ufd1 at the plasma membrane as well. There are several additional uncharacterised proteins predicted as well, including: AT5G15400, a

putative E4 ubiquitin elongating factor; AT3G24530, a AAA+ ATPase; AT3G18860, a WD-40 repeat family protein; and AT1G14200, a potential E3 ligase. AT4G14250 is annotated as a pseudogene. Overall, it may be concluded from these data that there could well be functional redundancy between isoforms, due to shared interactions. Continued searching of the predicted interacting proteins list reveals more UPS components, and no substrates were found that are involved in CHLORAD.

Finally, the expression of Npl4 and Ufd1 was explored. This may provide information on developmental or condition specific functions, as well as revealing whether any of the isoforms are in fact pseudogenes. In addition, co-expression data is a useful resource for predicting whether proteins may function in the same process. Genes which are co-expressed under the same conditions may do so in response to the same signal, to regulate when proteins are synthesised to act in a particular process; co-expression analysis has become a powerful tool to identify genes which are involved in the same pathway (Movahedi *et al.*, 2012). GENEINVESTIGATOR (Hruz *et al.*, 2008) was used to examine gene expression data from microarray experiments: expression of genes throughout development (Fig 2.1) and perturbations (not shown). All six of the *Arabidopsis* adapter protein genes were expressed at medium or high relative levels in all tissues (Figure 2.1). CHLORAD is known to be of developmental importance during chloroplast biogenesis and senescence. There was no evident change of expression during these stages of development (Figure 2.1). There is, however, a marked increase in the expression of Npl4, Npl4L and Ufd1A during the seed stage. Of the adapter proteins, Ufd1B is expressed to the lowest degree at all stages – the other proteins are expressed to a more comparable level. There was no notable impact of perturbations on gene expression. Co-expression data showed no CHLORAD genes – identified genes primarily consisted of UPS associated proteins, including numerous E3 ligases. There appeared to be weak co-expression between the putative paralogs: for instance, Npl4L did not feature in the top 100 co-expressed genes of Npl4. This might be indicative of discrete functions. UPS processes occur at all stages of the

cell cycle, throughout development and during abiotic and biotic stress. It could therefore be argued that the lack of change in expression is a reflection of constant activity and preparedness of the cell to engage in UPS activity. Though even if true, it makes it highly challenging to find any link to CHLORAD in these results by co-expression analysis.

Protein	Subcellular Compartment								
	Nucleus	Cytosol	ER	Golgi	Plastid	Mitochondrion	Plasma Membrane	Extracellular	Cytoskeleton
NPL4 (At2g47970)	2	**16**	2	0	0	10	0	2	0
NPL4L (At3g63000)	4	**22**	0	0	2	4	0	0	0
UFD1A (At2g21270)	10	**18**	0	0	2	4	10	10	2
UFD1B (At2g29070)	**10**	4	6	6	0	2	2	8	0
UFD1C (At4g15420)	**12**	6	0	0	2	4	0	0	0
UFD1D (At4g38930)	**10**	6	0	0	4	4	0	0	0

Table 2.2 Predicted subcellular location of proteins produced by the SUBA4 toolkit. Compartment scores are relative to max. Values highlighted in bold are the strongest predicted subcellular location.

	Input Sequence					
	NPL4 (At2g47970)	NPL4L (At3g63000)	UFD1A (At2g21270)	UFD1B (At2g29070)	UFD1C (At4g15420)	UFD1D (At4g38930)
Predicted Functional Partners	UFD1D	UFD1A	NPL4L	NPL4	NPL4	NPL4
	UFD1B	CDC48A	CDC48A	NPL4L	NPL4L	NPL4L
	UFD1A	UFD1D	NPL4	AT4G02950	RHA1A	CDC48A
	CDC48A	UFD1B	RPN10	RPN10	MUB4	RPN10
	CDC48B	CDC48B	CDC48C	AT4G03360	RPN10	AT3G18860
	CDC48C	CDC48C	CDC48B	RHA1A	AT4G14250	AT3G24530
	UFD1C	UFD1C	AT5G15400	AT3G18860	CDC48B	CDC48B
	MUB4	MUB4	PI4KG4	MUB4	AT1G14200	RHA1A
	AT3G24530	AT3G24530	RPN13	MUB5	AT3G24530	AT4G04180
	AT4G04180	AT4G04180	AT3G18860	AT3G24530	CDC48A	CDC48C

Table 2.3 Predicted protein-protein interactions produced by the STRING protein network toolkit. The top ten predicted interacting proteins (high confidence, >0.900) are shown for each Npl4 and Ufd1 isoform.

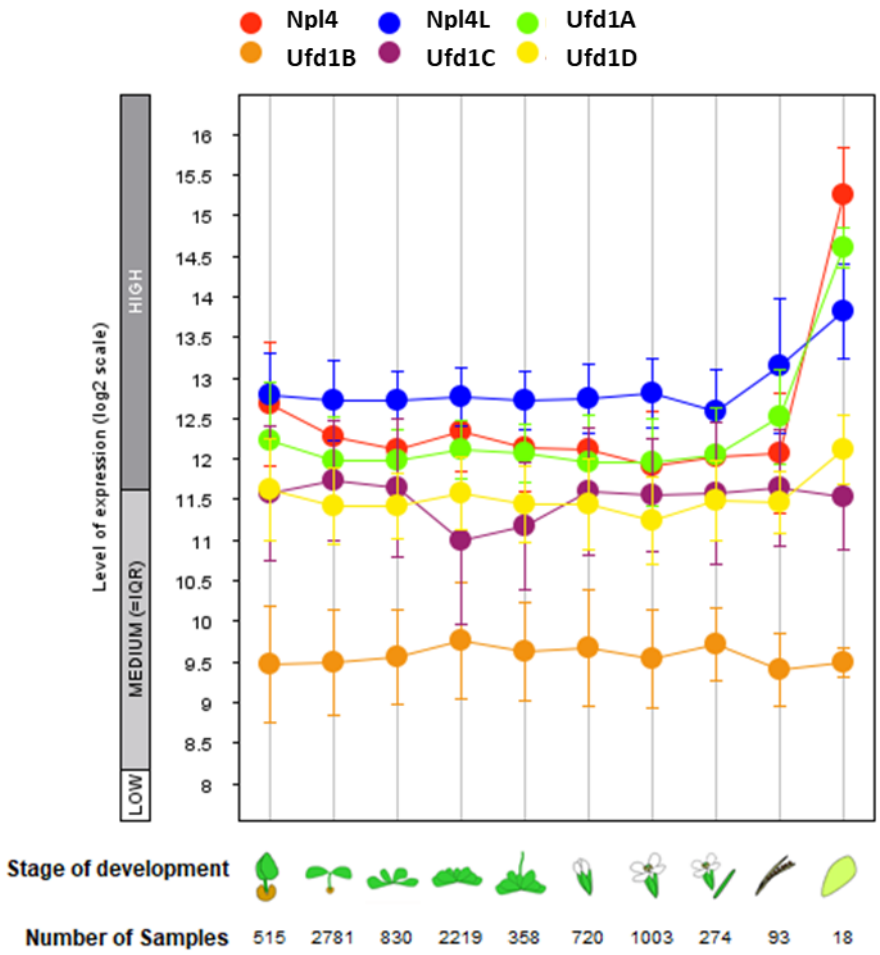


Figure 2.1 Multiple Line graph of gene expression levels of Npl4 and Ufd1 isoforms throughout plant development. The graph was produced using GENEVESTIGATOR, with data supplied by the Affymetrix genome array database for Arabidopsis.

2.2.3 Evolutionary Relationships of Npl4 and Ufd1

Phylogenetic tree reconstruction was then used to assess the evolutionary relationships within the NPL4 and UFD1 protein families in plants. I used BLASTP (basic local alignment search tool protein) searches to identify related sequences of AtUfd1 and AtNpl4 in the Viridaeplantae using Phytozome (Goodstein *et al.*, 2012). Phytozome is a useful database of curated genomic data, but it is limited in the number of species represented and biases towards higher plants of scientific and/or agronomic interest. Tree reconstruction was based on the principle of maximum likelihood.

Npl4 appears to be a conserved protein across plants, represented across taxa in basal lineages such as the chlorophytes. Both *Arabidopsis* Npl4 sequences are clustered in a node within the Brassicaceae (Fig. 2.2). There appeared to be orthologs of Npl4 and Npl4L shared with other Brassicaceae species. This is plausibly the result of a gene (or genome) duplication within the family, which is supported by the high sequence identity (83.5%). There are not two isoforms observed in basal lineages, such as Chlorophyta. It is possible that retaining multiple isoforms was an ancestral trait lost by most lineages, but it seems far more plausible that the ancestral state has only one protein — as it is known from studies on the evolutionarily divergent species *Homo sapiens* and *Saccharomyces cerevisiae,* that there is only a single isoform of Npl4 that is present. However, multiple isoforms of Npl4 were observed in several other eudicot families, such as the Fabaceae, Rosaceae and Crassulaceae (Fig. 2.2). This again may be the result of gene/genome duplication events or which are a common feature of plant evolution and responsible for higher copy numbers of many genes (Panchy *et al.*, 2016). Alternatively, retention of multiple copies could be sustained by selective pressure: this could relate to a functional consequence of multiple isoforms, such as acquiring distinct involvement in different pathways or redundancy to counter pathogenic selective antagonism.

The evolutionary reconstruction of Ufd1 sequences is far more complex. Initially, BLAST searching of plants with full length Ufd1A was sufficient to identify sequences from across plant taxa, as well as the AtUfd1B and AtUfd1D sequences. This led to the initial phylogenetic tree reconstruction of Ufd1 (Fig. 2.3). Like Npl4 and Npl4L, Ufd1A and Ufd1D appear to have diverged recently within the Brassicaceae, and they also share high sequence identity (78.5%). This again is likely the result of genome duplication event, and one might predict similar functionality between these proteins on this basis. Ufd1B, however, exists in a separate group as a putative paralog (Fig. 2.3). Intriguingly, there exists split branching in both the monocots and dicots. This would seemingly suggest that the existence of paralogous Ufd1 sequences is a conserved feature of higher plants, though they may

have separate evolutionary origins – as the monocot sequences are more similar to other monocot sequences and *vice versa* for dicots. This could suggest convergent evolution between these paralogous groups — though it is highly speculative to suggest that these necessarily sustain conserved divergent functions. Which may be the ancestral group is not clear, though Ufd1B does have more sequence similarity to yeast Ufd1, than Ufd1A or Ufd1D do. It is also interesting that many species appear to have multiple copies of Ufd1, in both subgroups. For instance, there are a total of eight Ufd1 sequence from *Glycine max* (soybean). Once again, gene/genome duplications may be the likely explanation – but this degree of sequence number is not observed in the Npl4 phylogeny (Fig. 2.2). The basal species all seem to have one copy of Ufd1, aside from the moss *Sphagnum fallax.* Thus, it seems more likely that the ancestral state is to have only one Ufd1 protein, as in yeast and mammals. It is also worth noting that the reconstruction of Ufd1 does not follow the same evolutionary sequence as expected. Notably, the basal angiosperm species *Amborella trichopoda* is grouped within a branch containing eudicot species, and the Ufd1A/Ufd1D type Ufd1 sequences. Additionally, the monocot Ufd1 sequence from *Zostera marina* is also placed at an adjacent branch to *Amborella*. The final tree topology may therefore not be optimal or resemble the true relationship between these sequences in nature.

In order to construct the phylogram for Ufd1C, BLAST searches had to be performed with that specific sequence. In these results, there is likewise no reciprocal identification of the other *Arabidopsis* Ufd1 sequences. This does raise the question of whether Ufd1C is a true Ufd1 homologue. However, it does not appear to be a pseudogene as there are related sequences identified across the Viridaeplantae (Fig.2.4). The architecture of this phylogeny is surprising, however, with moss species and the basal angiosperm *Amborella trichopoda* grouped on a node including the dicots. The monocot sequences are placed as a sister group to the outgroups (Fig. 2.4), suggesting that these surprisingly have greater sequence similarities to the yeast and human Ufd1 than older plant lineages. While his tree contains a smaller subset of sequences in general, it is disappointing that no sequences from

Chlorophyta species were identified — their absence may contribute to the peculiar topology between the monocots and eudicots.

In attempting to produce a completed Ufd1 phylogeny containing all identified *Arabidopsis* sequences, identified protein sequences from both Figure 2.3 and Figure 2.4 were placed in the same multiple sequence alignment. The resulting phylogenetic tree (Fig. 2.5) shows what is perhaps an unsurprising topology, given the seeming distinction of Ufd1C sequences. Intriguingly, the previously observed topological abnormality (Fig.2.3) is less pronounced here (Fig 2.5): *Zostera* is correctly placed in a branch with other monocot species; though *Amborella* is still not represented as the basal angiosperm and is placed on a branch with eudicot species. Ufd1C forms a distinct group, with retained topology (Fig. 2.4); likewise, the majority of the topology of the original Ufd1 tree (Fig. 2.3) is also retained, but as a sister group to Ufd1C. It therefore seems possible that there are three putatively paralogous groups of Ufd1 in plants: the Ufd1A/D type, Ufd1B type, and Ufd1C type (Fig 2.5).

Finally, phylograms were reconstructed to examine the relationships between Npl4 and Ufd1 sequences in model species. This was attempted to demonstrate a broader appreciation of the evolution of these proteins across eukaryotes. The following sections in the chapter detail intriguing differences that are apparent in the predicted domains and structures of these proteins, particularly in the Ufd1 isoform, Ufd1C; examination of Npl4 and Ufd1 in a broader eukaryotic context, may highlight where differences evolved. As model species are generally well characterized and have curated genomic and proteomic resources, these were chosen to represent more evolutionarily distant species. Included in the tree are diverse major phyla, including basal eukaryotes (like the ciliate *Tetrahymena thermophyla*) and higher eukaryotes (including a variety of chordates, such as *Homo sapiens*). Major clades in the Viridiplantae were also represented by model species, including chlorophytes (*Chlamydomonas reinhardtii*), bryophytes (*Physcomitrella patens*) monocots (*Oryza sativa*) and dicots (*Populus trichocarpa*).

Once again, a generally simpler portrait is painted for Npl4 sequences (Fig 2.6). The Viridiplantae are placed as sister to other eukaryotes, which is broadly in agreement with general phylogenetic reconstructions of these species using other data sets (Hedges, 2002). In the following sections, the Npl4 sequences in *Arabidopsis* are found to not have a zinc finger domain; this observation was sustained in the Viridiplantae by examination of the multiple sequence alignment produced using the MAFFT algorithm. Conserved residues in the CHCC zinc finger were identified in all other eukaryotic model organisms — this may be reflected in the branching observed (Fig 2.6). It is unclear whether the zinc finger is an ancestral feature. General consensus in the phylogeny of eukaryotes (Burki, 2014) would suggest that *Tetrahymena thermophyla* should have closer evolutionary relatedness to the Viridiplantae than the other eukaryotic species present. It may therefore be plausible that the zinc finger domain was lost during the evolution of the Viridiplantae. An additional observation from the multiple sequence alignment, was that the C-terminal zinc finger present in *Homo sapiens* Npl4 (Nguyen *et al.*, 2022), is solely present in other chordate species and may have evolved exclusively in this phylum (Fig 2.6).

The Ufd1 tree (Fig 2.7) reinforces the previous observations in the Viridiplantae (Fig 2.5) that Ufd1C is evolutionarily distant sequence. Intriguingly, searching for related sequences using BLASTP for Ufd1C identified several proteins annotated as TRAFD1 — this is explored in more detail in section 2.3.6. The group containing Ufd1C and TRAFD1 sequences is sister to all other Ufd1 proteins present (Fig 2.7); I therefore believe that it has a far older evolutionary origin. All Ufd1 proteins contain the same conserved UT3 domain, which defines them as Ufd1 proteins. However, the large difference between the Ufd1C related sequences and other Ufd1 proteins (especially the well-studied mammalian and yeast Ufd1 proteins), may further support the argument that it is not a true Ufd1 protein.

The remainder of the tree contains the Ufd1 proteins related to *Arabidopsis* Ud1A, Ufd1B and Ufd1D (Fig 2.7). The topology here is somewhat surprising, as the Viridiplantae, are positioned as most distantly related to the TRAFD1 type branch. The relationships between

the other eukaryotic species is close to what might be expected, when compared to a generally accepted phylogeny (Burki, 2014). It would also appear that the origin of divergent Ufd1 sequences is isolated within the Viridiplantae, and that this divergence occurred within the embryophyta. There are multiple sequences present for *Oryza sativa*, *Populus trichocarpa* and *Arabidopsis thaliana*. The origin of the split between the Ufd1B and Ufd1A/D type is uncertain; this is because one sequence from *Oryza sativa* is placed as more evolutionarily distant than the single Ufd1 isoform from the bryophyte *Physcomitrella patens.* Two potential origin points of this divergence are therefore marked (Fig. 2.7).

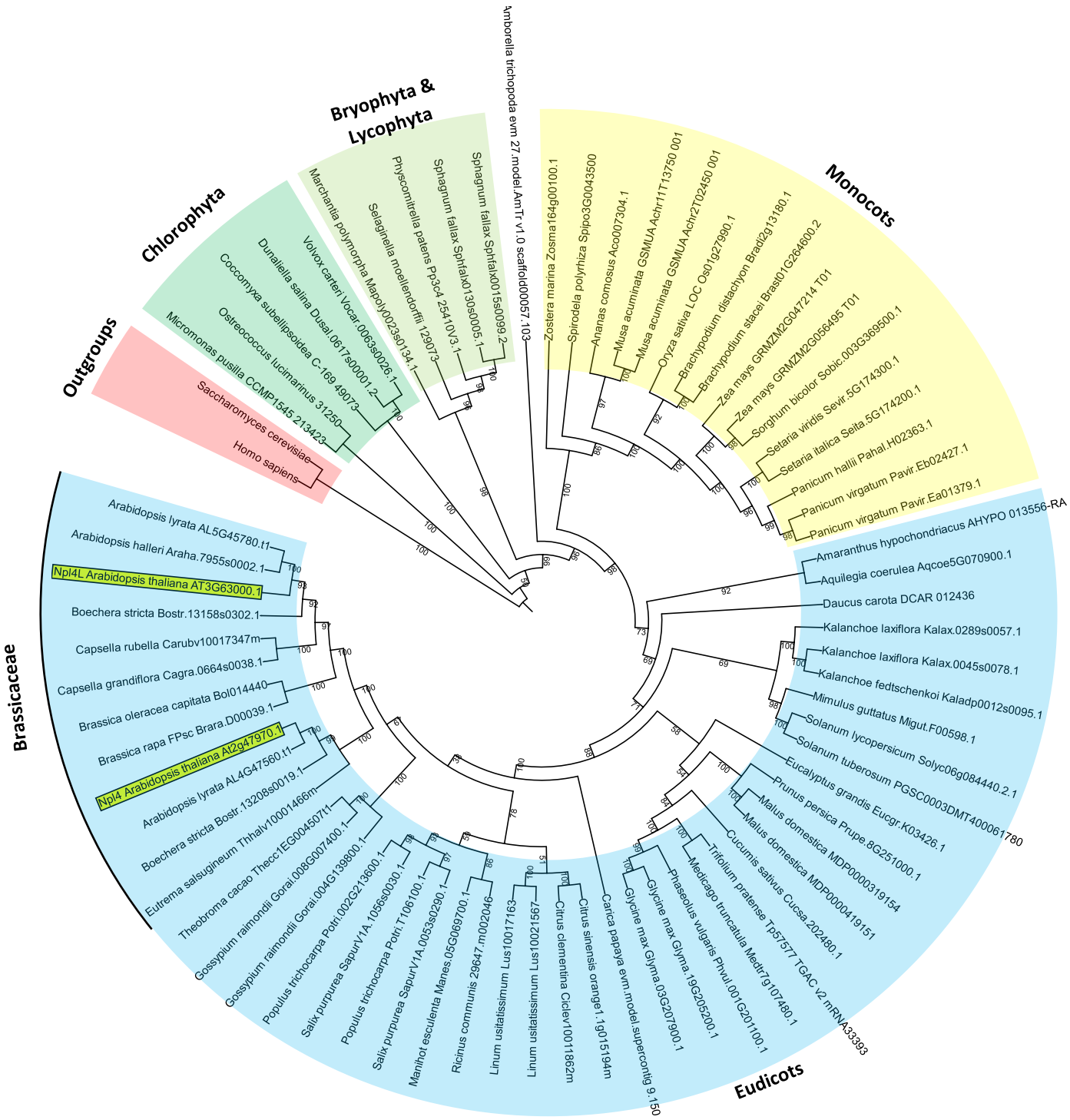


Figure 2.2 Maximum-likelihood phylogenetic tree reconstruction of Npl4 amino acid sequences across plants. Orthologue sequences were sourced from BLAST searches of human and yeast Npl4 sequences against the Phytozome database. Coloured blocks represent broad cladistic annotations of organisms with featured sequences. As a basal angiosperm, *Amborella trichopoda* is not grouped into any annotated clade. Angiosperms are broadly subdivided into monocotyledonous and eudicotyledonous species. The family Brassicaceae is shown with a black line, and *Arabidopsis* sequences are highlighted in

yellow. Human and Yeast sequences were used as the outgroups. Numbers on branches indicate bootstrap support from 1000 replicates.

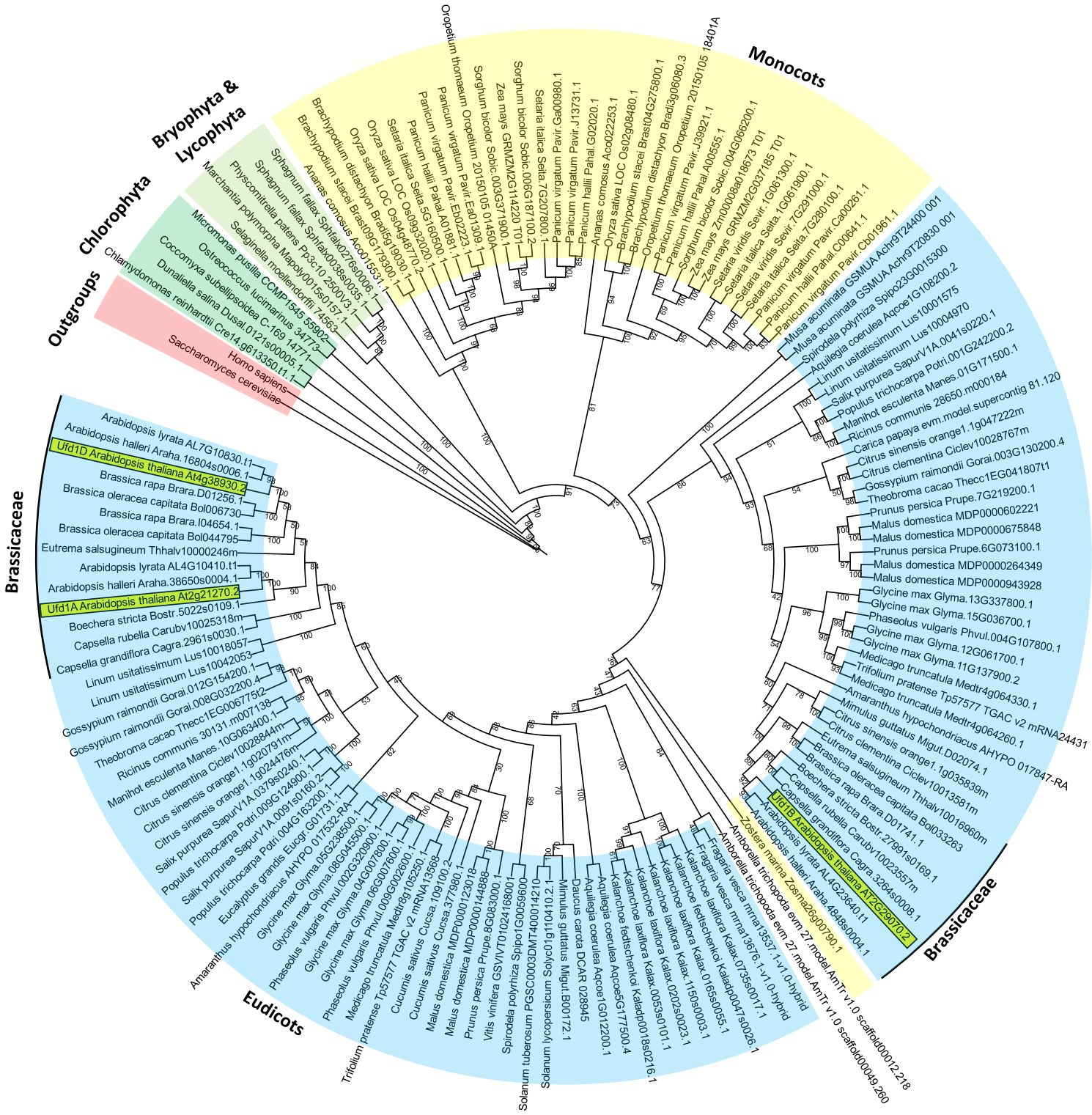


Figure 2.3 Maximum-likelihood phylogenetic tree reconstruction of Ufd1 amino acid sequences across plants, containing Ufd1A, Ufd1B, and Ufd1D. Orthologue sequences were sourced from BLAST searches of human and yeast Npl4 sequences against the Phytozome database. Coloured blocks represent broad cladistic annotations of organisms with featured sequences. As a basal angiosperm, *Amborella trichopoda* is not grouped into any annotated clade. Angiosperms are broadly subdivided into monocotyledonous and eudicotyledonous species. The family Brassicaceae is shown with a black line, and Arabidopsis sequences are highlighted in yellow. Human and Yeast sequences were used as the outgroups. Numbers on branches indicate bootstrap support from 1000 replicates.

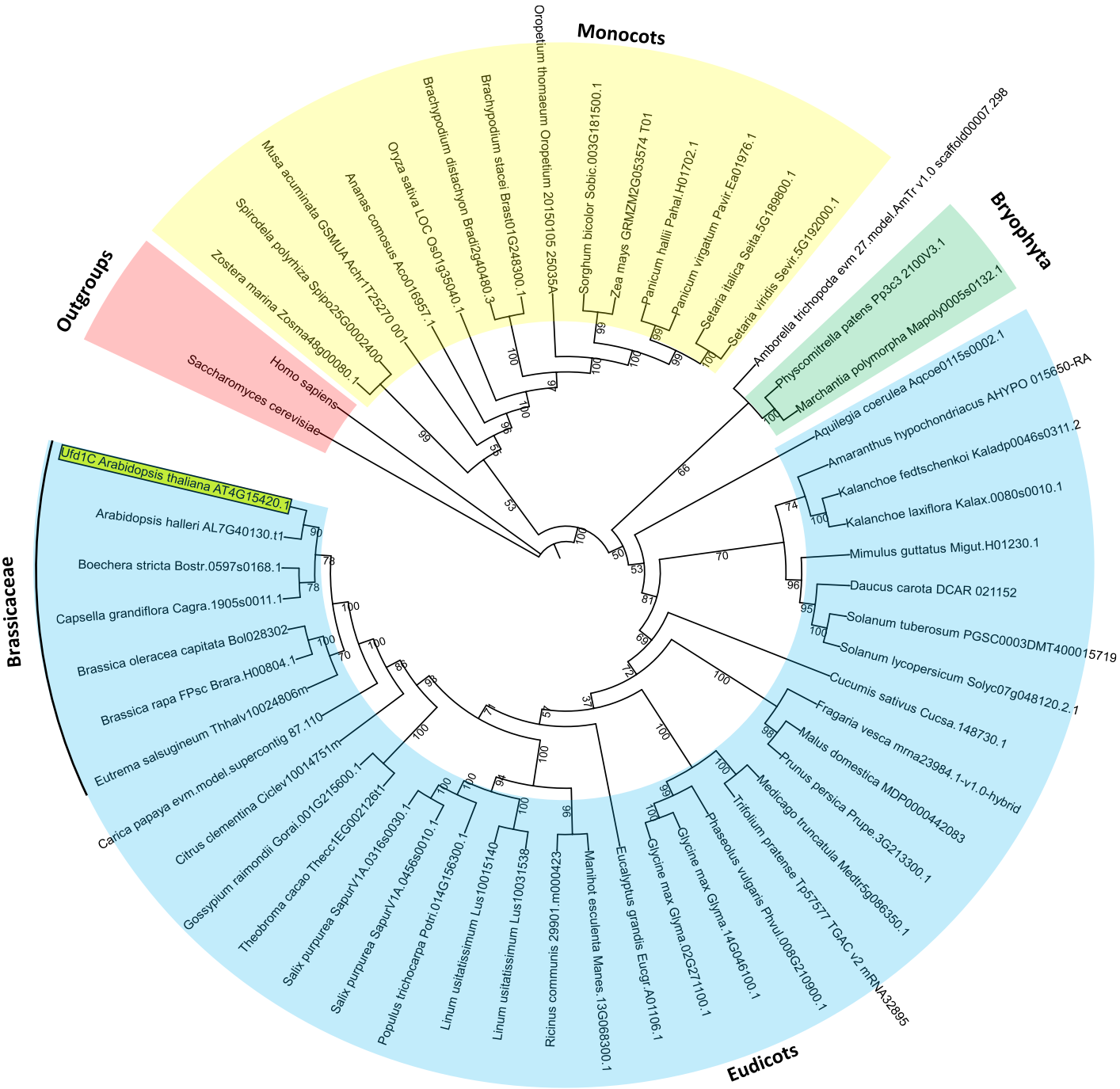


Figure 2.4 Maximum-likelihood phylogenetic tree reconstruction of Ufd1C amino acid sequences across plants. Orthologue sequences were sourced from BLAST searches of AtUfd1C against the Phytozome database. Coloured blocks represent broad cladistic annotations of organisms with featured sequences. As a basal angiosperm, *Amborella trichopoda* is not grouped into any annotated clade. Angiosperms are broadly subdivided into monocotyledonous and eudicotyledonous species. The family Brassicaceae is shown with a black line, and Arabidopsis sequences are highlighted in yellow. Human and Yeast sequences were used as the outgroups. Numbers on branches indicate bootstrap support from 1000 replicates.

Figure 2.5 Combined maximum-likelihood phylogenetic tree reconstruction of Ufd1 amino acid sequences across plants, containing all four Ufd1 isoforms in Arabidopsis. Coloured blocks represent broad cladistic annotations of organisms with featured sequences. As a basal angiosperm, *Amborella trichopoda* is not grouped into any annotated clade. Angiosperms are broadly subdivided into monocotyledonous and eudicotyledonous species. The family Brassicaceae is shown with a black line, and Arabidopsis sequences are highlighted in yellow. Human and Yeast sequences were used as the outgroups. Numbers on branches indicate bootstrap support from 1000 replicates.

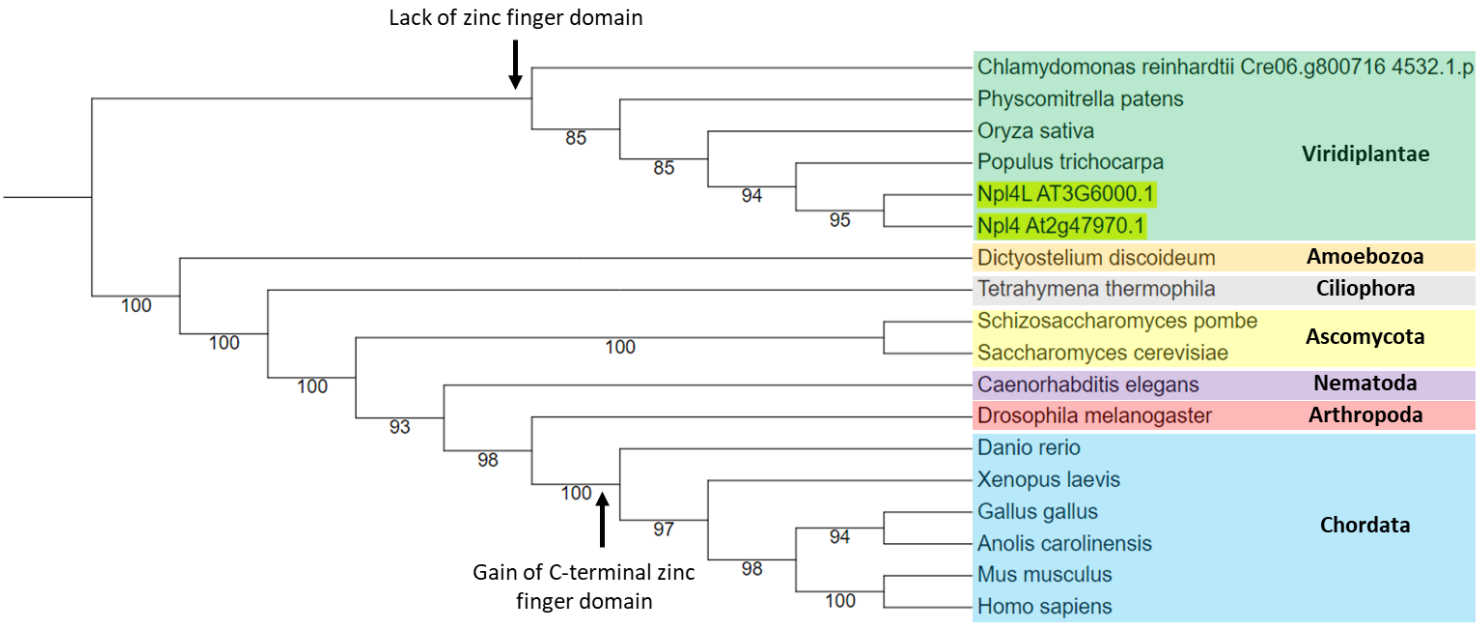


Figure 2.6 Maximum-likelihood phylogenetic tree reconstruction of Npl4 amino acid sequences in a selection of Eukaryotic model species. Coloured blocks denote clades by phylum – the Viridiplantae are the exception, and are broadly grouped together. *Arabidopsis* sequences are highlighted in yellow. Marked by black arrows are two key features in the evolution of Npl4 sequences: the lack of a zinc finger domain in the Viridiplantae and the gain of an additional zinc finger domain at the C-terminus in the Chordata. The Npl4 sequence from *Chlamydomonas* was not identified in the initial Npl4 phylogeny, and the gene identifier is thus displayed. Numbers on branches indicate bootstrap support from 1000 replicates.

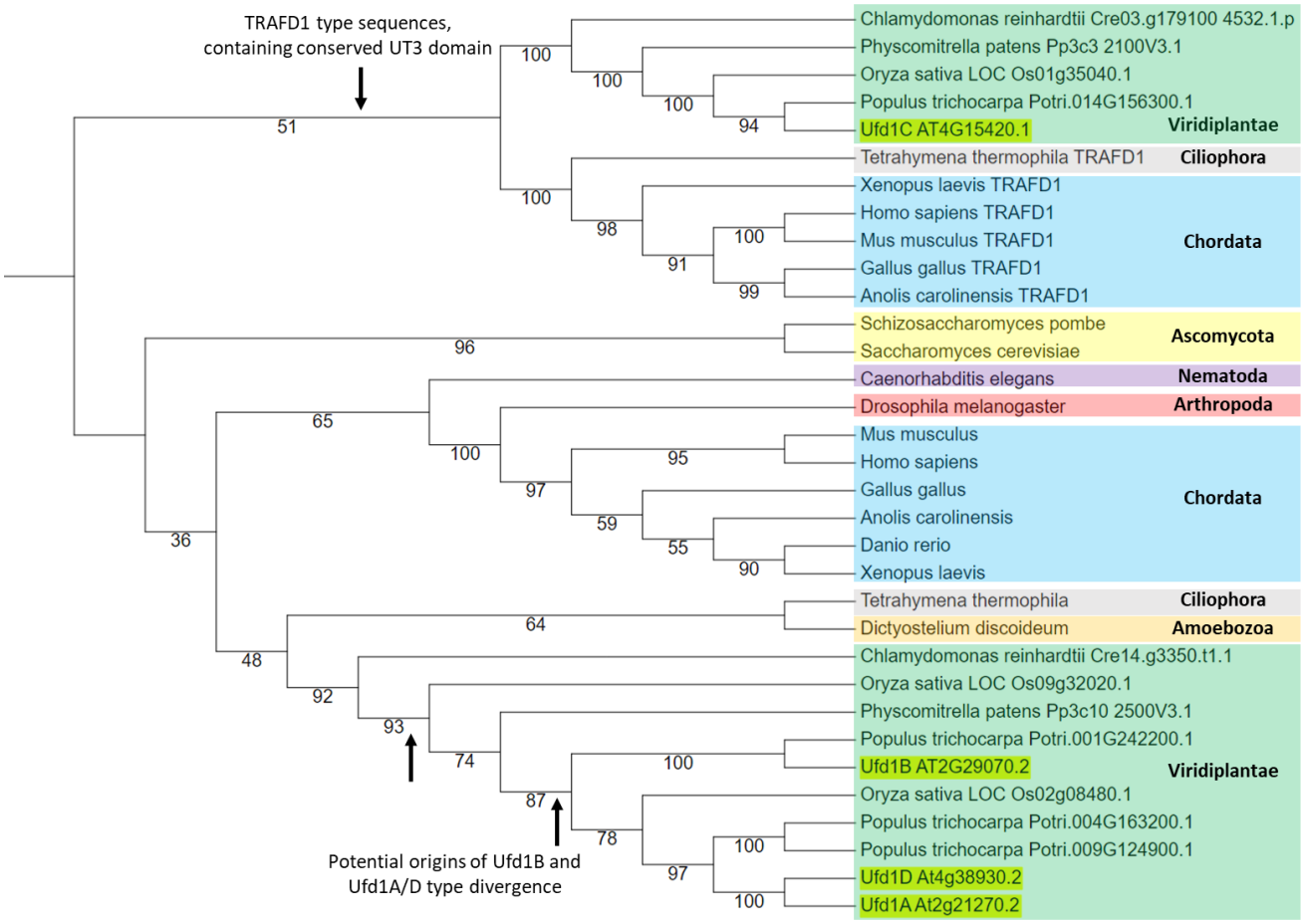


Figure 2.7 Maximum-likelihood phylogenetic tree reconstruction of Ufd1 amino acid sequences in a selection of Eukaryotic model species. Coloured blocks denote clades by phylum – the Viridiplantae are the exception, and are broadly grouped together. *Arabidopsis* sequences are highlighted in yellow. Marked by black arrows are key features in the evolution of Ufd1 sequences: the TRAFD1 type sequences appear as a sister to other Ufd1 sequences, and the potential origins of multiple isoforms of Ufd1 within the Viridiplantae. Gene identifiers are displayed for Viridiplantae sequences due to multiple being present for each species. Numbers on branches indicate bootstrap support from 1000 replicates.

2.3.4 Protein Domain Architecture

The Phyre2 (Kelley *et al.*, 2015) and Interpro (Mitchell *et al.*, 2019) servers were used to search for conserved structures and test domain homology. These data were integrated through annotating amino acid sequences of each isoform (Fig. 2.8). The strongest alignments produced by Phyre2 formed the principle basis of these annotations; this was due to its strength in identifying conserved structural, and therefore functional, sequences from PSI-BLAST (position-specific iterative basic local alignment search tool). Interpro domain annotations formed a secondary test to identify annotated sequences from bioinformatic databases. I additionally used multiple sequence alignment with the MAFFT (multiple alignment using fast Fourier transform) program (Katoh & Standley, 2013) to search for specific motifs shared with yeast proteins, which are annotated with high confidence.

Conflicts existed only in the form of omitted annotations from one or more of the methods, which may be expected given their varying strengths. The identification of conserved structures was aided by several published structures for Cdc48-Npl4-Ufd1 interfaces in other eukaryotes (Bodnar *et al.*, 2018). A ubiquitin-like β grasp structure is predicted at the N-terminus of Npl4 and Npl4L, consistent with the known structure of the UBXL domain and this was supported by an Interpro annotation. Structural homology predicted by Phyre 2, supported annotations of the enzymatically inactive MPN domain (Bodnar *et al.*, 2018) of Npl4, as well as the C-terminal domain (Fig 2.8). Surprisingly, no zinc finger domain was identified in either of the Npl4 proteins. Two zinc finger motifs are present in the zinc finger domain of Npl4 from yeast (Bodnar *et al.*, 2018; Twomey *et al.*, 2019), and these were found to be functional in enhancing the unfoldase activity of Cdc48. It is therefore intriguing that the *Arabidopsis* Cdc48-Npl4-Ufd1 complex may possess a difference in its mechanism of action, due to the lack of a zinc finger domain.

The Ufd1 isoforms showed greater variability, and this was especially the case for Ufd1C: the latter protein has an additional ~230 residues, and a unique C-terminal zinc finger

domain (Fig. 2.8). The conserved structure predicted in all *Arabidopsis* Ufd1 isoforms contains a double-psi β barrel motif, which is predicted to facilitate an interaction with ubiquitin (Bodnar *et al.*, 2018). I expect this structural annotation to correspond to the UT3 domain of Ufd1 proteins (2.6), as the N-terminal UT3 domain of human Ufd1 interacts with ubiquitin (Le *et al.*, 2016); this is reinforced by a Ufd1 signature match from Interpro for this region. An MPN domain was found in both Npl4 and Npl4L (Fig. 2.8) solely by Interpro.

Ufd1's interaction with Cdc48 is mediated by an SHP motif, which is conserved amongst nine human co-factors, and has the consensus sequence: h-x-x-F-p-G-x-G-x-x-h-G (Hänzelmann & Schindelin, 2017). Through multiple sequence alignment, I identified the SHP motif in the *Arabidopsis* Ufd1 proteins, aside from Ufd1C. I also identified an NBM domain in allUfd1 isoforms aside from Ufd1C. The difference in architecture between Ufd1C and the other Ufd1 proteins is striking. It raises further concerns that it may not be a true homologue to the Ufd1 proteins found in yeast and mammals, despite sharing the defining domain. How it might interact within the heterodimer and with Cdc48 requires further investigation.

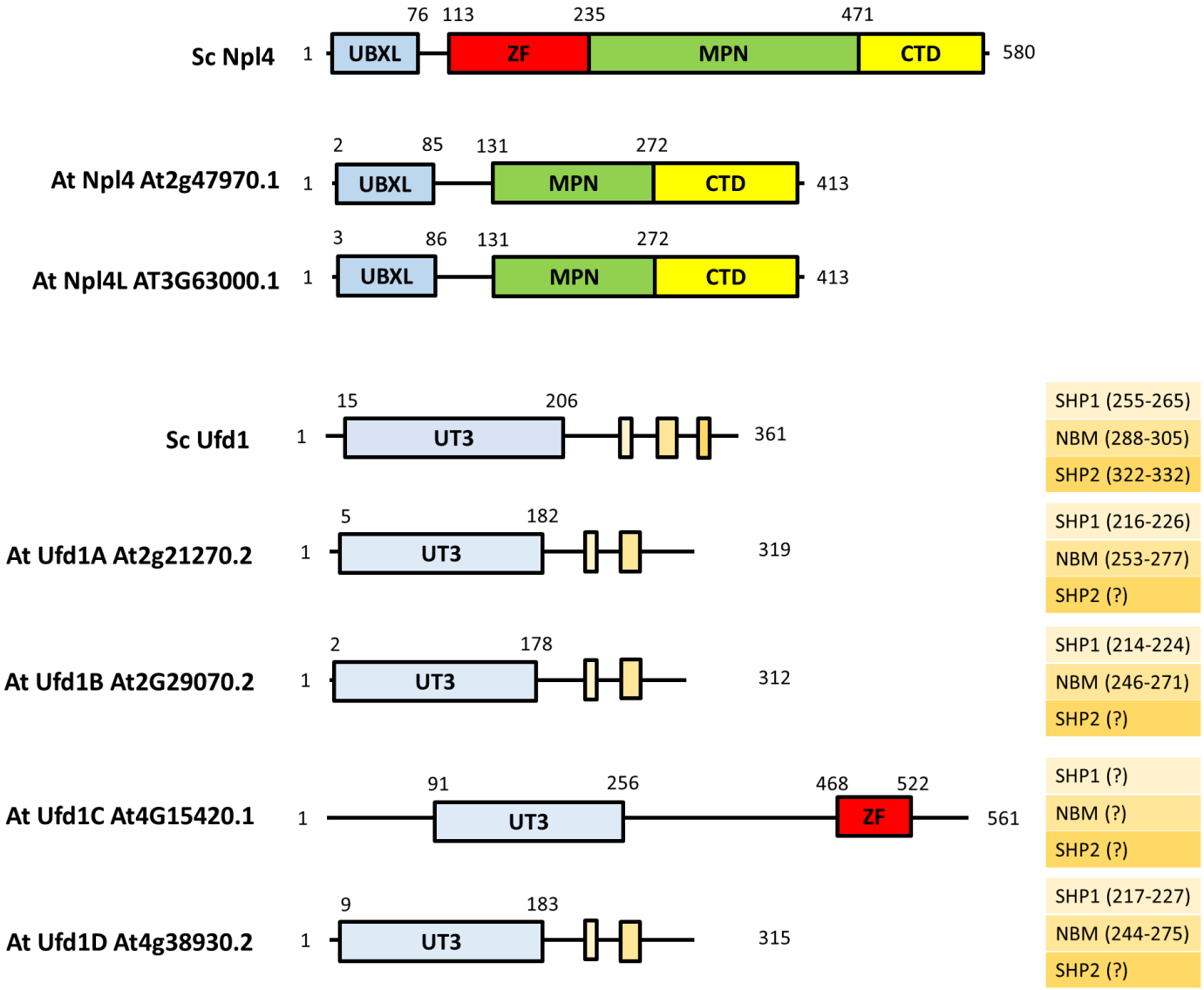


Figure 2.8 Schematic diagram illustrating the known and predicted domains of the yeast (Sc) and Arabidopsis (At) Npl4 and Ufd1 protein sequences. Sc Npl4 and Sc Ufd1 are reference sequences with domain annotations sourced from existing literature. Numbers represent amino acid positions in the protein sequence starting from the N terminus. Coloured blocks show domain types and positions. Ufd1 SHP and NBM positions are detailed adjacent to the sequence diagrams for clarity. Abbreviations: Npl4, Nuclear Protein Localisation 4; Ufd1, Ubiquitin Fusion Degradation 1; MPN, Mpr1/Pad1 N-terminal; ZF, zinc finger; NBM, NPL4 Binding Motif; SHP, binding site 1; UT3 is not defined; UBXL, Ubiquitin Binding region X Like; Ub, ubiquitin; aa, amino acids.

2.3.5 Predicted Protein Structures

Understanding protein structure can reveal mechanistic insights into protein function. A dramatic leap in the ability of scientists to predict protein structure has emerged with the development of Alphafold. Alphafold offers highly accurate structural prediction through advanced neural network layers (Jumper *et al.*, 2021). It offers predicted structures that far exceed competing methods and are close to experimentally determined structures in accuracy (Pereira *et al.*, 2021).

To begin with, predicted structures were downloaded from the Alphafold2 database (Fig. 2.9). For comparison, I included the structures predicted for the yeast Npl4 and Ufd1 proteins. These may act as a positive control given the existing experimentally determined structures, which improve the accuracy of the models. Many domain annotations (Fig. 2.8) are reinforced by these models, as exemplified by the N-terminal UBXL domains present in Npl4 sequences, which show comparable folded structures (Fig. 2.9). Clear differences are also readily apparent between the yeast and *Arabidopsis* Npl4 structures. The most obvious is the lack of a zinc finger domain, which is coloured initially in yellow on yeast Npl4 (Fig. 2.9). A potential consequence of this presence is the closer proximity of the UBXL domain with zinc finger and MPN domains in the yeast protein. Npl4 and Npl4L have a seemingly much more flexible linking sequence between these domains (Fig. 2.9). This could be a product of the rendered protein in isolation however, as interaction in a complex may alter the electrostatic interactions and the position of the UBXL domain. It was observed by Twomey *et al.* (2019) that while the yeast UBXL domain positions on a different N domain to that of Ufd1, the other Npl4 domains are in close proximity to the Ufd1 bound N domain – a flexible sequence may thus be functional here.

Through rendering electrostatic structures in the software ChimaeraX (Pettersen *et al.*, 2021), I attempted to determine whether I could observe the groove in the Npl4 backbone which housed unfolded ubiquitin (Ji *et al.*, 2022; Pan *et al.*, 2021; Twomey *et al.*, 2019). It did appear that one was visible, with sections of the zinc finger and MPN domain contributing to

a groove which matched the previously observed shape (Twomey *et al.*, 2019) (Fig 2.10). To attempt to validate this interaction, I utilised the data presented by Twomey *et al*. 2019 on the residues that contributed to the formation of the groove in *Saccharomyces cerevisiae* Npl4 (Fig 2.10). Through multiple sequence alignment using MAFFT, I plotted the residues that aligned with those present in *Arabidopsis* Npl4 sequences. The observed groove is approximately traced (Fig 2.10) but shows the same kink shape observed in the yeast structure. This could indicate that *Arabidopsis* Npl4 and Npl4L retain the structure capable of housing an unfolded ubiquitin moiety.

There is less overall definition in the yeast Ufd1 structure, outside of the N terminal UT3 domain, which has a characteristic double-psi β barrel fold followed by a mixed α/β roll structure (Park *et al.*, 2005) (Fig. 2.9). The two α helices, which are not conserved in *Arabidopsis* Ufd1 proteins, contribute to no known domain. The two SHP boxes and NBM region of yeast Ufd1 are found in three distinct kinks (in the region coloured in yellow-orange). These kinks may also be observed in Ufd1A, B and D – even if the second SHP box could not be located by multiple sequence alignment, there may appear to be conserved structure.

I must now discuss Ufd1C. The structure of this protein is highly enigmatic. Firstly, there exists an enormous unstructured region at the N-terminus, which is followed by ~100 amino acid long α helix (Fig. 2.9). It is unclear what both of these sections may contribute to the protein's function. We then reach the conserved UT3 domain, which is followed by an uncharacterised structure and domain, before finally reaching the zinc finger at the C-terminus. The structure following the UT3 domain, superficially resembles the double-psi β barrel fold of the UT3 domain – it is however structurally distinct, and not a detected domain by Interpro. The zinc finger will be examined in more detail in section 2.3.6. It is unclear what function Ufd1C may have based on this structure. The UT3 domains alone suggest a role in ubiquitin binding, but the substantial differences to the other Ufd1 proteins raise questions about whether it can form a heterodimer with Npl4 and interact with Cdc48.

To examine protein-protein interactions within the Cdc48-Npl4-Ufd1 complex, I used protein multimer prediction produced by Alphafold. The main limitation with this approach was computational. Unlike Alphafold2 (which is a database) running multimer prediction relies on real-time computation and thus the hardware to perform this. I utilised a Google Colab Fold server (Mirdita *et al.*, 2022), which contains the Alphafold multimer prediction pipeline running in a Jupyter notebook environment. Here, each session was limited to 12 Giga bytes of video random access memory on the graphical processing unit – exceeding this limit caused the runtime to fail. Consequently, only pairs of proteins could be computed in any single instance. The runtime also utilized a modified version of the Alphafold pipeline. This replaces the multiple sequence alignment algorithm of HHMer with MMseqs2, which is roughly 50x faster (and thus less computationally demanding) (Mirdita *et al.*, 2022).

First, I input yeast sequences of Npl4, Ufd1, and Cdc48 to examine heterodimer formation (Fig. 2.11). As before, this acted as a positive control for the method. Initially, it might be argued that my earlier hypothesis was supported in relation to the positioning of Npl4 and Cdc48: there is a relative change in the structure of the linking sequence between the UBXL domain and the zinc finger. The UBXL domain of Npl4 is shown to bind to Cdc48 within the cleft of the N domain, as experimentally demonstrated (Bodnar *et al.*, 2018). The Ufd1 interaction with Cdc48 is located at the kink produced by the initial SHP box (Fig. 2.11). Here the second SHP box does not present the same kink as observed in the monomer (Fig. 2.9). Ufd1 has been observed to bind to two N domains simultaneously via both SHP boxes; though as only one copy of Cdc48 could be computed, I could not examine this. The interaction between yeast Npl4 and Ufd1 was also broadly similar to experimentally determined crystal structures (Fig.1.2): the UFD1 NBM domain is in close proximity to the Npl4 MPN domain, with both the UBXL domain of Npl4 and UT3 domain of Ufd1 free to engage with different parts of the complex.

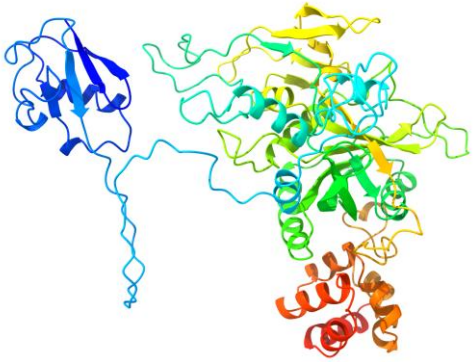
Next, I input the *Arabidopsis* sequences of Npl4 and Ufd1 to examine interactions within the heterodimer (Fig. 2.12). It should be noted that the gene model I have used throughout for

Ufd1C is At4g15420.1, whereas the Alphafold2 database contained At4g15420.2 – this is why the unstructured N-terminus is not present in future models (Fig.2.9; Fig. 2.12). These results are surprisingly different. In each case of Npl4/Npl4L with Ufd1A, B, or D, interactions between the proteins occur with the UT3 domain of Ufd1 nestled in between the UBXL and MPN domain. There is a slight difference with Ufd1C, which is shown to position proximal to the backbone of Npl4 and Npl4L, further away from the UBXL and MPN domains (though still involving the UT3 domain). It is notable that Npl4 and Npl4L share very similar structures, and this is consistent with their interactions with Ufd1 proteins. It seems unlikely that the conserved UT3 domain, which is a known ubiquitin interacting structure, would bind to this region – this would sequester it from grasping ubiquitin, raising further uncertainty about the function and operation of Ufd1 proteins in *Arabidopsis*. Multimer prediction is considerably less accurate than monomer prediction (Mirdita *et al.*, 2022), so without reference structures (as for the yeast proteins), caution in their interpretation may be wiser. The C-terminal region of Ufd1 is generally less structured, and with the reduced accuracy of this model prediction it lacks the previously observed kinks present in the NBM and SHP box domains – this may offer an explanation as to the observed differences in interaction.

Predictions of binding to Cdc48, on the other hand, are far closer to what is predicted from yeast (Fig. 2.13). Both Npl4 proteins demonstrate binding to the Cdc48 N domain cleft. The binding site of Ufd1A and B is located in a different position on the N domain the UBXL (as expected), but it does not involve the same kink structure or region where the SHP box is predicted – instead interaction occurs at the C-terminus. It is Ufd1D that shows the closest similarity to the yeast interaction (Fig. 2.11; Fig. 2.13). This difference may be related to accuracy limitations in assembling the model, as before (Fig. 2.12). For Ufd1C, the interaction takes places in a unique way: the N domain cleft of Cdc48 is engaged by the C terminal portion of the Ufd1C zinc finger. This is a reasonably plausible interaction in principle, as zinc finger domains are often involved in protein binding. However, binding to the N domain cleft may then create competition within the heterodimer of Npl4 as well.

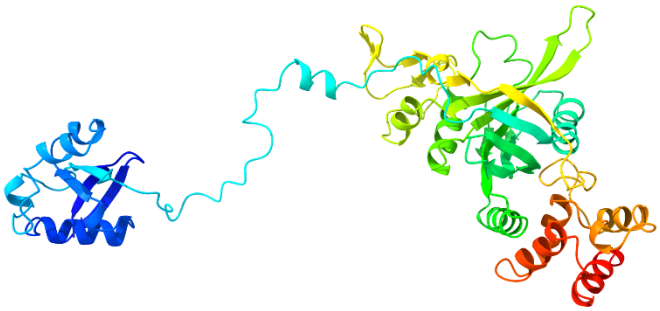
Nevertheless, this may not be an issue if the heterodimer is preassembled and binds to different N domains in the homohexamer.

Figure 2.9 (overleaf) Alphafold2 database protein structure predictions for Npl4 and Ufd1 protein structures from Yeast (Sc) and *Arabidopsis* (At). Yeast models are shown for reference, given existing experimental structural data. 3D models were rendered in ChimaeraX and coloured by a rainbow gradient according to amino acid position, from the N terminus (blue) to the C terminus (red). Models were rotated and positioned to an approximately similar orientation for image capture.

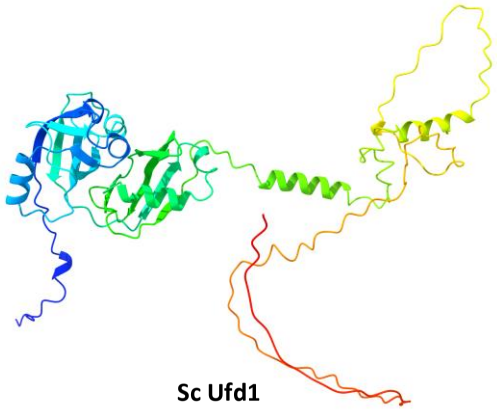


Sc Npl4

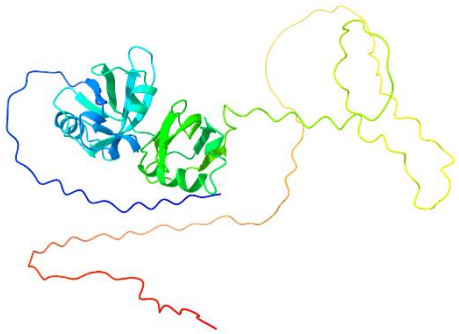
At Npl4



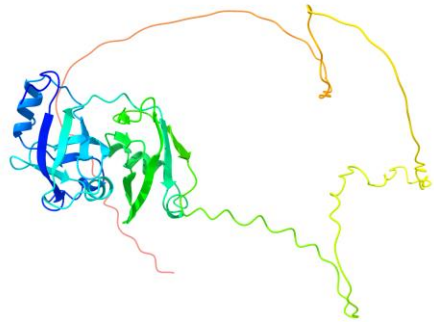
At Npl4L



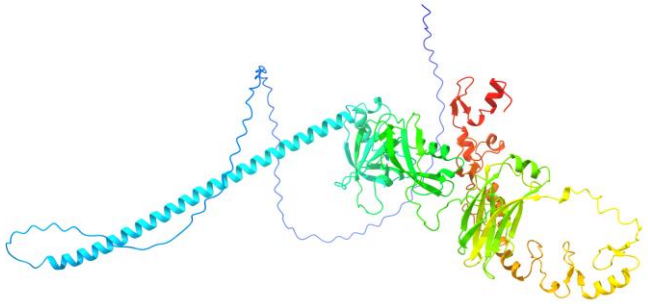
Sc Ufd1



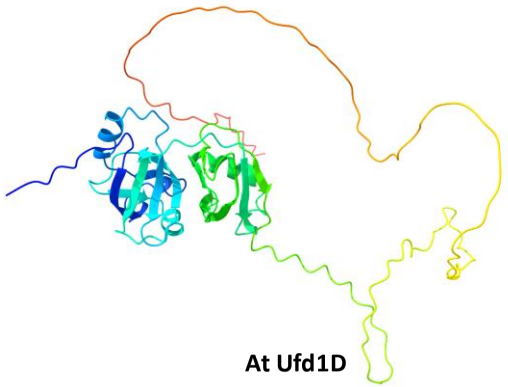
At Ufd1A



At Ufd1B



At Ufd1C



At Ufd1D

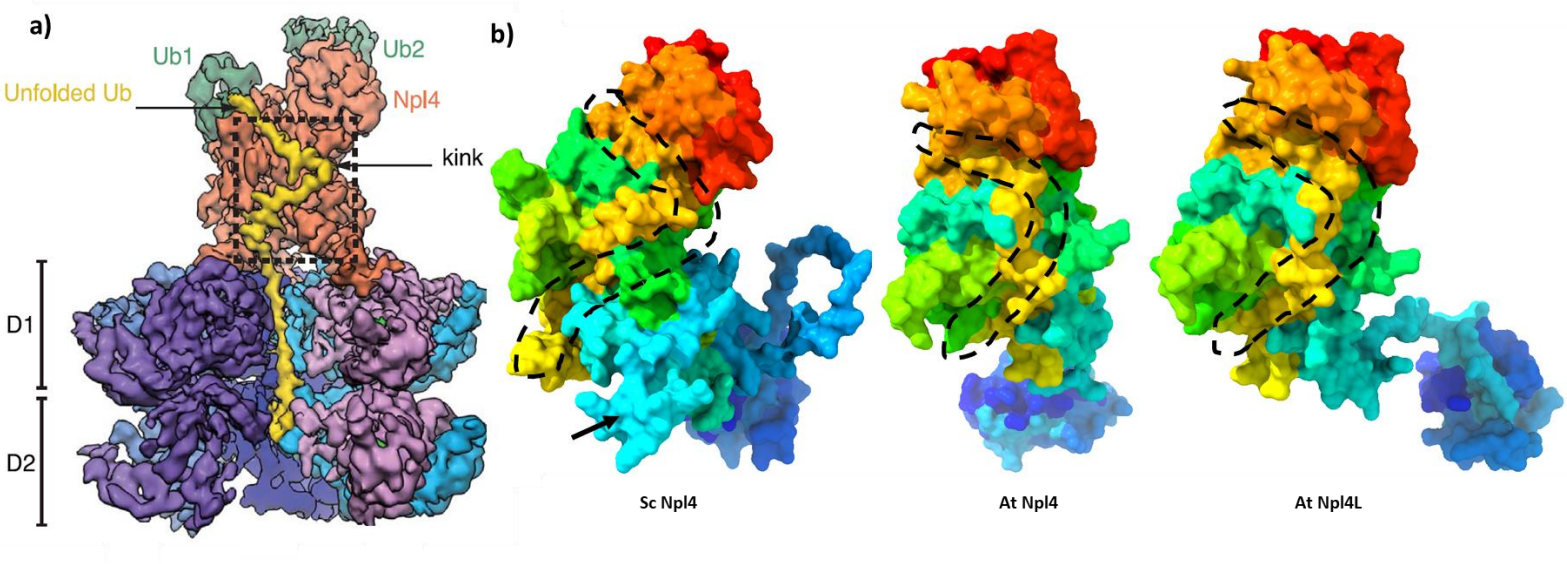


Figure 2.10 A structural groove which was observed to hold an unfolded ubiquitin moiety may be conserved in *Arabidopsis thaliana* Npl4 proteins. **a)** Adapted from Twomey *et al*. 2019: cryogenic electron microscopy density map of the Cdc48-Npl4-Ufd1 complex, in the presence of a polyubiquitin chain. Cdc48 is displayed as a cross section in the lower half of the image, with the central pore visible in between the D1 and D2 rings (N domains were omitted). Assembled on top of Cdc48 is the Npl4 tower (red). The crystallised structure contained an unfolded ubiquitin chain (yellow), bound to a groove in Npl4, exhibiting a kink shaped conformation (inside the dashed box); the N terminal portion of the ubiquitin chain extended into the Cdc48 pore to the D2 ring. Two additional ubiquitin moieties were observed to be attached to the Npl4 tower, labelled Ub1 and Ub2, distal to the unfolded ubiquitin. **b)** Alphafold2 database protein structure predictions for Npl4 structures from Yeast (Sc) and *Arabidopsis* (At). 3D models were rendered with their electrostatic potential in ChimaeraX and coloured by a rainbow gradient according to amino acid position, from the N terminus (blue) to the C terminus (red). Models were rotated and positioned to an approximately similar orientation for image capture. The yeast model is displayed as a positive control. The observed groove in Npl4 is indicated in each predicted structure by a dashed outline, showing the kink conformation. Experimental observation (Twomey *et al.* 2019) finds that the groove in Sc Npl4 may be traced by amino acids R253, I249, F470,

H468, R232, K165, H166, Q227, Y461, D460, F349, and W252: approximate positioning of the kink by multiple sequence alignment of residues of the groove is marked on predicted *Arabidopsis* Npl4 structures. The structure of the Sc Npl4 zinc finger domain, absent in At sequences, is marked by a black arrow.

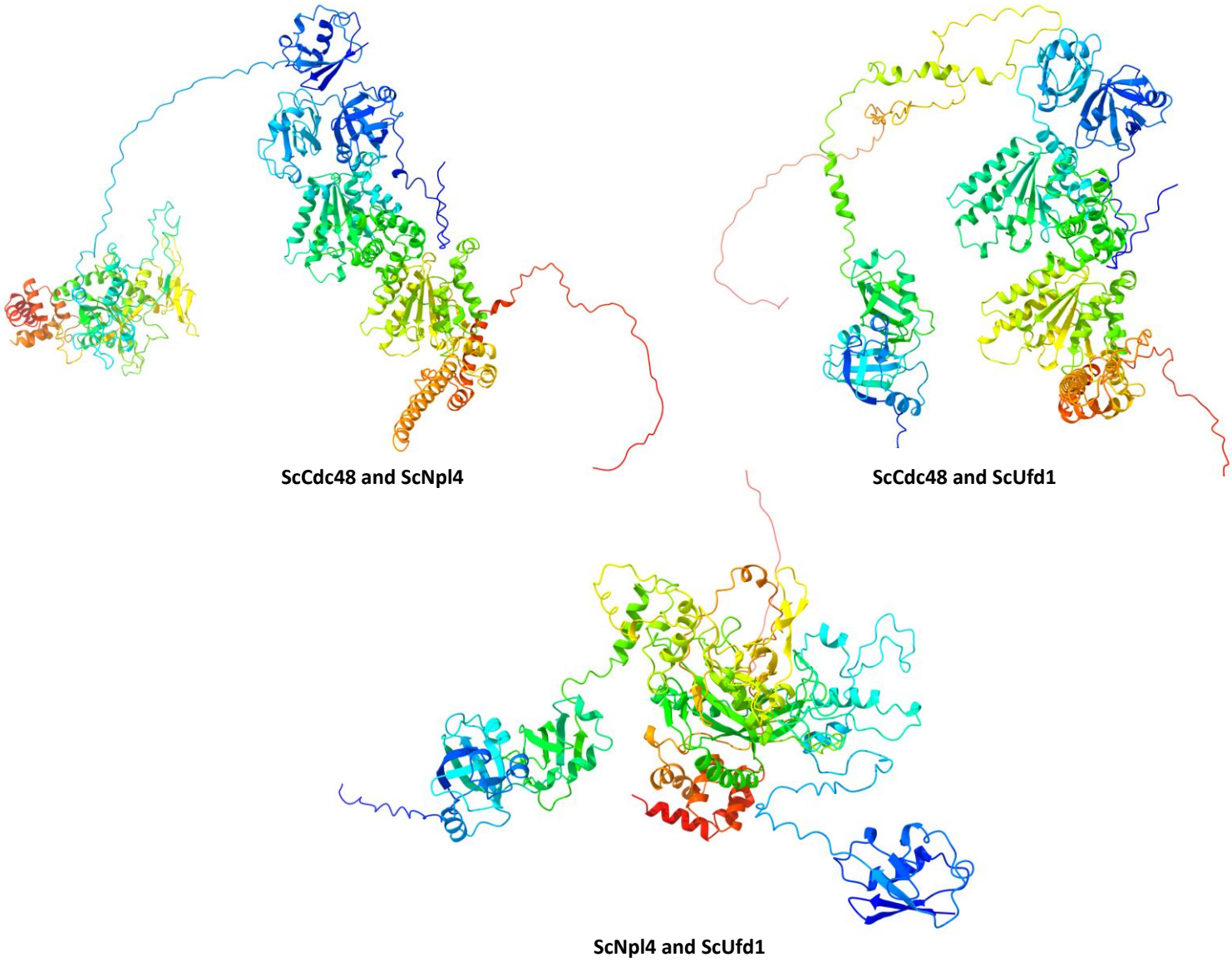
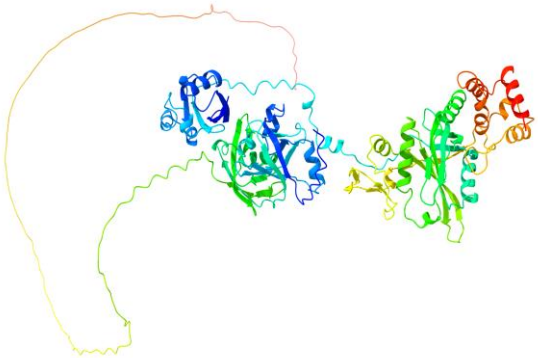


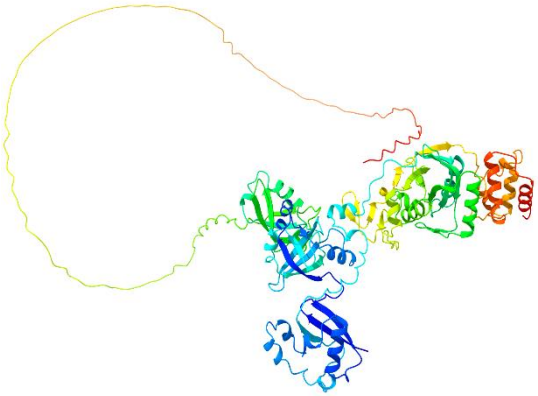
Figure 2.11 Alphafold2-multimer protein structure predictions for yeast heterodimers of Cdc48 and Npl4/Ufd1, and Npl4 and Ufd1. Alphafold2 sequence alignments and templates were produced using MMseqs2 and HHsearch. 3D models were rendered in ChimaeraX and coloured by a rainbow gradient according to amino acid position, from the N terminus (blue)

to the C terminus (red). Models were rotated and positioned to an approximately similar orientation for image capture.

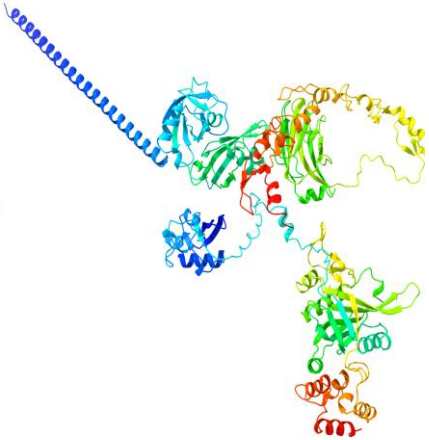
Figure 2.12 (overleaf) Alphafold2-multimer protein structure predictions for *Arabidopsis* heterodimers of Npl4 and Ufd1 isoform combinations. Alphafold2 sequence alignments and templates were produced using MMseqs2 and HHsearch. 3D models were rendered in ChimaeraX and coloured by a rainbow gradient according to amino acid position, from the N terminus (blue) to the C terminus (red). Models were rotated and positioned to an approximately similar orientation for image capture. Ufd1 is found on the lefthand side.



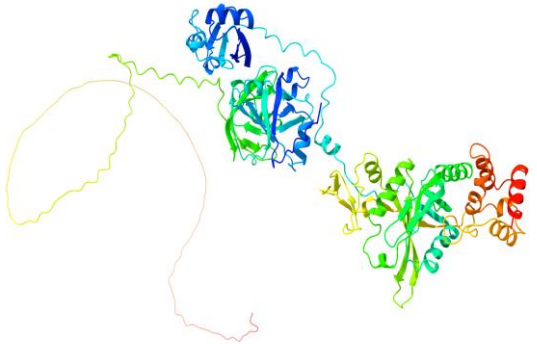
Npl4 and Ufd1A



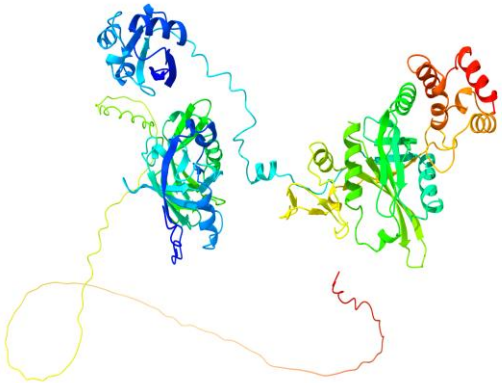
Npl4 and Ufd1B



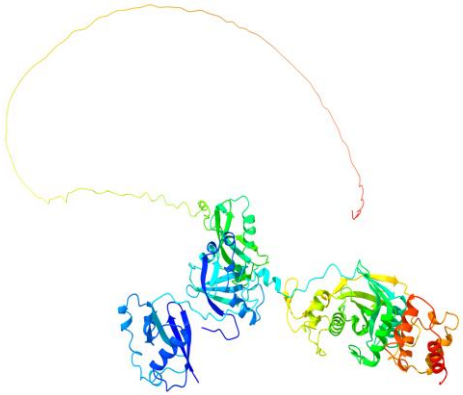
Npl4 and Ufd1C



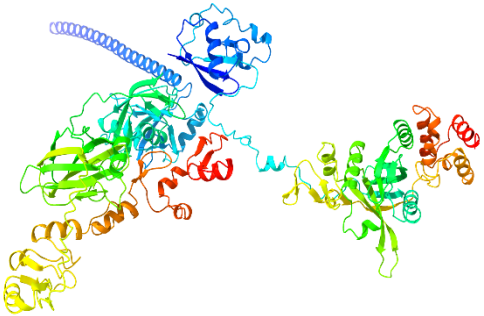
Npl4 and Ufd1D



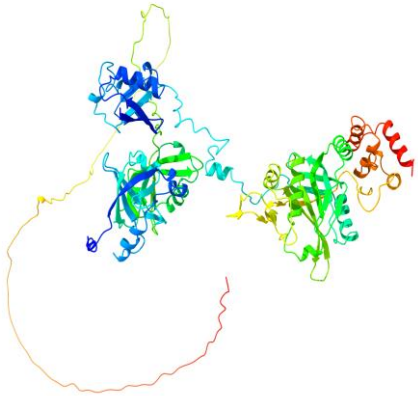
Npl4L and Ufd1A



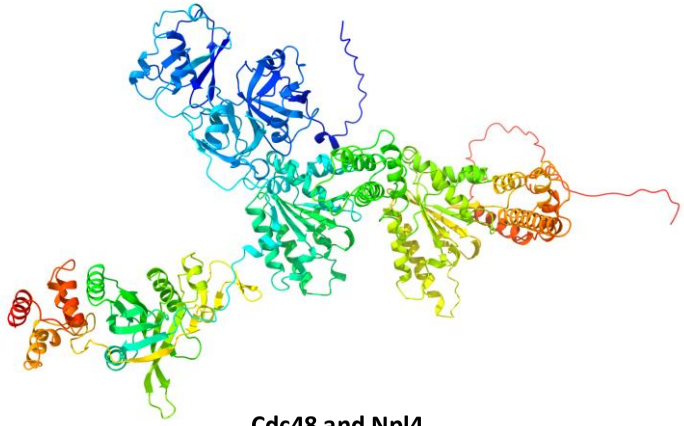
Npl4L and Ufd1B



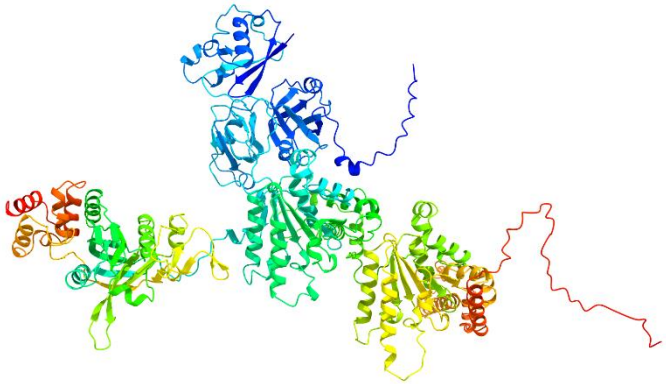
Npl4L and Ufd1C



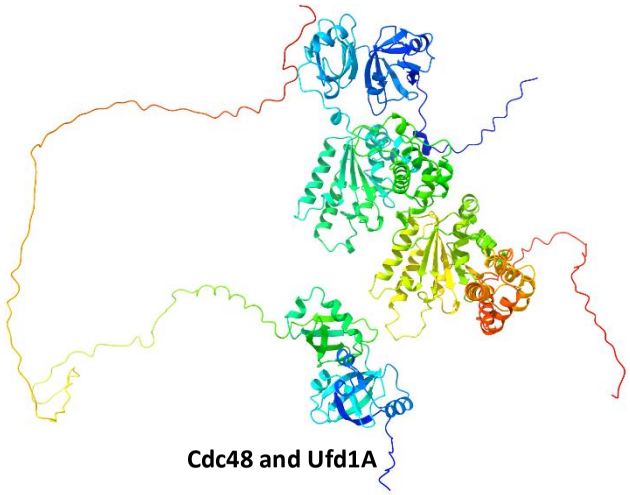
Npl4L and Ufd1D



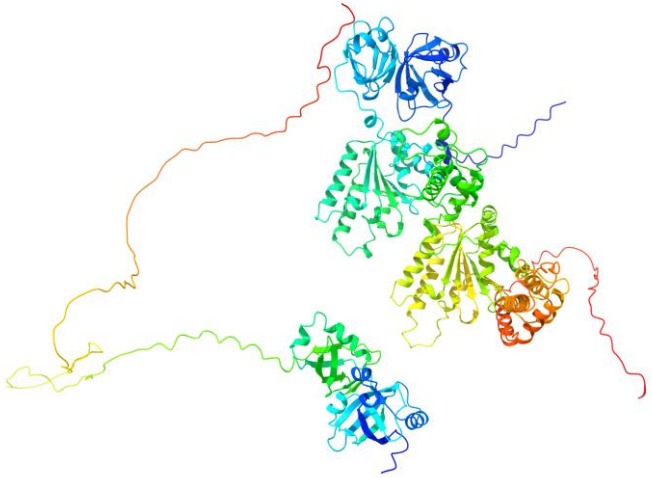
Cdc48 and Npl4



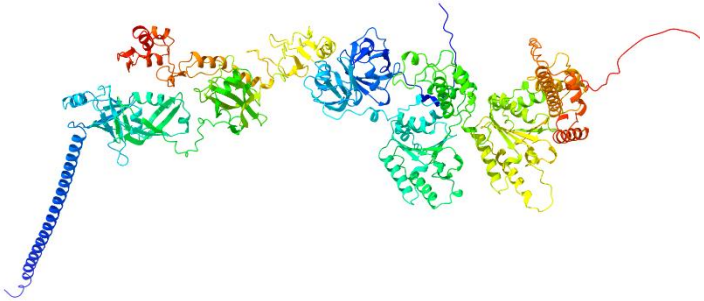
Cdc48 and Npl4L



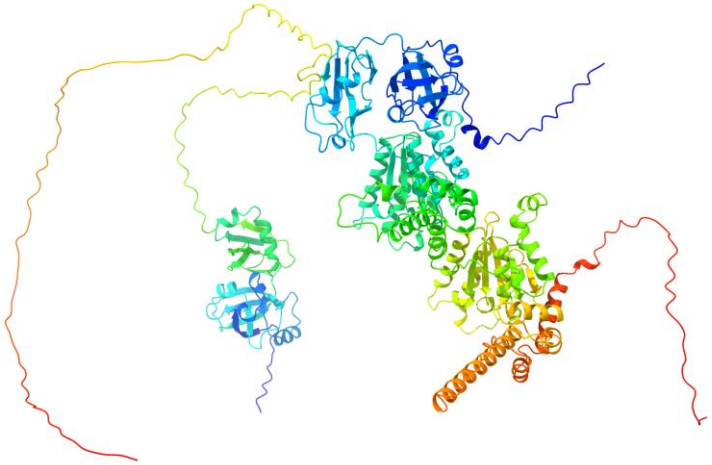
Cdc48 and Ufd1A



Cdc48 and Ufd1B



Cdc48 and Ufd1C



Cdc48 and Ufd1D

Figure 2.13 Alphafold2-multimer protein structure predictions for *Arabidopsis* heterodimers of Npl4 or Ufd1 isoforms with Cdc48. Alphafold2 sequence alignments and templates were produced using MMseqs2 and HHsearch. 3D models were rendered in ChimaeraX and coloured by a rainbow gradient according to amino acid position, from the N terminus (blue) to the C terminus (red). Models were rotated and positioned to an approximately similar orientation for image capture. Npl4 or Ufd1 is found on the lefthand side.

2.3.6 Ufd1C Possesses a C-terminal C3HC4 Zinc Finger Domain

A key finding in the domain analysis was that Ufd1C possesses a C-terminal zinc finger. Zinc finger proteins are extremely diverse, with numerous functions in cellular processes e.g. transcriptional regulation. Annotation from Interpro labelled Ufd1C as including a C3HC4 type zinc finger motif. Like zinc fingers generally, C3HC4 proteins can be of diverse functions e.g. in peroxisome biogenesis and the N-end rule pathway (Wu *et al.*, 2014).

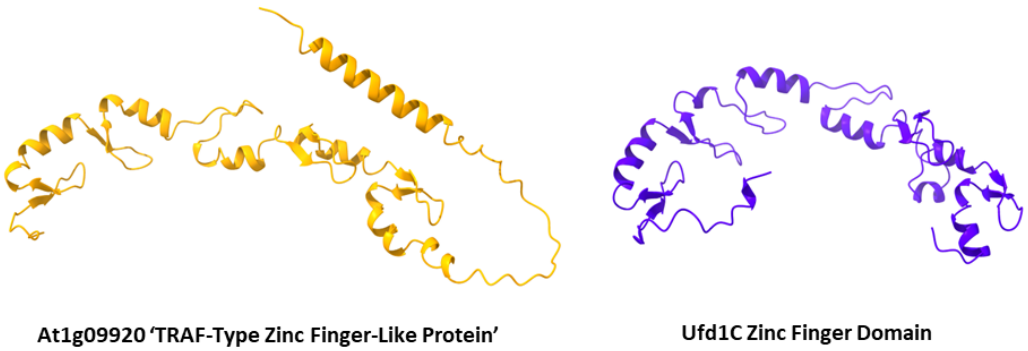
Perhaps coincidentally, the E3 ligase in CHLORAD, Sp1, is also a C3HC4 type E3 ligase. I first sought to more concretely determine whether Ufd1C truly possesses a C3HC4 domain. To accomplish this, I created multiple sequence alignments using a range of algorithms available from the European Bioinformatics Institute. In these alignments I included the zinc finger region of Ufd1C, as well as Sp1 and its *Arabidopsis* homologues Spl1 and Spl2, as well as the human orthologue of Sp1, MULAN (Ling *et al.*, 2012). I was able to successfully identify the C3HC4 residues using the algorithms T-COFFEE, MAFFT, MUSCLE, and KALIGN, but not Clustal Omega. From the consensus, I conclude that it is likely that Ufd1C possess a C3HC4 zinc finger based on conserved amino acid residues and predicted structure (Fig. 2.14a). The structure has no similarity to the C3HC4 RING domain of the E3 ligase Sp1 however.

To examine potential homologous zinc finger sequences, I performed BLASTP searching against model organisms with the Ufd1C zinc finger domain. Through this, I identified the human protein TRAFD1 (tumour necrosis factor receptor associated factor domain containing 1) (Witwicka *et al.*, 2015). I also then identified the homologue of this protein in *Arabidopsis*, AT1G09920: which is an uncharacterised TRAF type zinc finger protein. I then loaded the Alphafold2 strucutural database predictions for AT1G09920 and Ufd1C's zinc finger into ChimaeraX. I used the matchmaker function to align the structures, which produced a RMSD (root mean square deviation of atomic position) between 23 pruned atom pairs of 1.223 angstroms – demonstrating high conservation of structure (Fig. 2.14b). It

therefore appears that Ufd1C has a C-terminal TRAF type zinc finger. This is supported by phylogenetic reconstruction (Fig 2.7), which finds related TRAFD1 proteins in multiple eukaryotic model species through BLASTP searching. The functional implications of this are unclear. Unfortunately, TRAF type zinc fingers are not appreciably well-studied, especially in relation to the protein's function. In humans, TRAFD1 is a negative regulator of toll like receptors in the immune system (Witwicka *et al.*, 2015). Here in the TRAFD1 the zinc finger is involved in protein binding. It seems very plausible that this domain in Ufd1C is also involved in protein binding in some way – there may even be evidence to support that with Cdc48 (Fig. 2.13). It may also have some function in ubiquitin binding or modification, as these are additional common functions of Cdc48 adapter proteins.

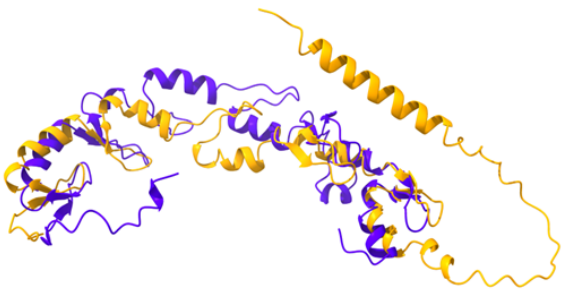
a)

b)



At1g09920 'TRAF-Type Zinc Finger-Like Protein'

Ufd1C Zinc Finger Domain



Matched Structures

Figure 2.14 Ufd1C zinc finger characterisation. **a)** Multiple sequence alignment using the KALIGN algorithm of the zinc finger domains of Arabidopsis Ufd1C, Sp1, SpL1, Spl2, and human MULAN. Residues aligning to the conserved C3HC3 motif are marked with a red asterisk. Aligned residues are highlighted in grey with a consensus of >50%. Colour bars above the consensus sequence represent the occurrence of each residue, with darker shades meaning greater occurrence. **b)** Alphafold2 database protein structures of At1g09920 (orange) and the zinc finger region of Ufd1C at its C-terminus (blue). The Matchmaker function of ChimaeraX was then used to superimpose structures, following pairwise sequence alignment (Needleman-Wunsch algorithm and BLOSUM-62 similarity matrix).

Chapter III

Results Chapter II: Protein Localisation and Interactions

3.1 Abstract

CHLORAD involves the selective ubiquitination of TOC proteins at the chloroplast OEM. To determine whether Npl4 and Ufd1 participate with Cdc48 in CHLORAD, I examined the subcellular localisation and interactions of these proteins. Initially, YFP fusion proteins of Npl4 and Ufd1 were observed using transiently expressing cells. Ufd1B was observed to solely localise to the nucleus and it is thus not capable of participating in CHLORAD – it was therefore removed from future experiments. All other Npl4 and Ufd1 proteins, however, showed similar localisation in the cell (validated using stable transgenic lines expressing YFP constructs), and may be broadly described as nucleocytoplasmic – this is in line with predictions in Chapter II and the observations in the literature. To examine where these interactions took place, I used bimolecular fluorescence complementation experiments. These showed surprising patterns of localisation with Cdc48, in specific and punctuated areas of the cell – the exact compartments are not clear, though there remains a lower level of cytoplasmic signal and punctuated fluorescence at the chloroplast OEM with all adapter proteins. Interactions within the heterodimer of Npl4 and Ufd1, and between Npl4 or Ufd1 with Sp1, Toc33 and Toc159 all showed stronger fluorescence in the cytoplasm and at the chloroplast OEM, also supporting the involvement of these proteins in CHLORAD. Physical interactions were validated using co-immunoprecipitations. These confirmed interactions within the heterodimer, with Cdc48, Sp1, Toc33 and SP2. Native interaction with TOC proteins was not detectable and it may be related to a signal threshold issue with Western blotting. There may appear to be broad redundancy between the observed interactions of the adapter proteins. However, Ufd1A and Ufd1D showed weaker interactions with Toc33 and Sp2. It may therefore be plausible that the Ufd1 isoform Ufd1C — which possesses a

markedly different domain architecture as observed in Chapter II — has greater likelihood of involvement in CHLORAD than the other Ufd1 isoforms.

3.2 Introduction

The current model of CHLORAD (Ling *et al.*, 2019) involves the degradation of TOC proteins in the chloroplast outer envelope membrane, to regulate the protein import process. The TOCs, particularly the isoforms Toc159, Toc33, and Toc75, are therefore the main identified substrates in this process. Other components of CHLORAD include the E3 ligase Sp1 (which engages in autoubiquitination activity and is thus also a substrate of CHLORAD) as well as the β barrel protein Sp2, which forms the retrotranslocon.

My hypothesis is that Cdc48 operates in CHLORAD via the activity of the UPS adapter proteins Npl4 and Ufd1. In Chapter II, I began characterising these proteins by sequence analysis, which demonstrated that they are likely functional components of the UPS with the capability to cooperate with Cdc48. However, no evidence specifically pointed to a role in chloroplast biology or CHLORAD from that data. Here, I begin addressing that issue with experimental approaches. Cdc48 operates in a variety of cellular activities in different compartments. Thus, it is crucial to establish which of the *Arabidopsis* homologues of Ufd1 and Npl4 may participate in CHLORAD by determining which of them could function at the chloroplast OEM. I then examined the specific interaction with CHLORAD proteins using two independent experiments. Another consideration is the degree of redundancy between the isoforms. I therefore also aimed to examine evidence of specific functions for any pair of Npl4 and Ufd1 as the CHLORAD heterodimer. A hope was to eliminate some proteins from future analysis, to reduce the number of mutant lines and constructs that required generation.

3.3 Results and Discussion

3.3.1 Npl4 and Ufd1 Proteins Show Nucleocytoplasmic Localisation

To identify where the Npl4 and Ufd1 proteins localise within the cell, I decided to employ translational fusions to the yellow fluorescent protein (YFP). YFP vectors containing the CDS (coding sequence) region of each of the Npl4 and Ufd1 isoforms were used to transfect isolated *Arabidopsis* mesophyll protoplasts, which were then viewed under a confocal microscope.

With confidence, I can say that all six of the adapter proteins show localisation to the nucleus (Fig. 3.1). With the exception of Ufd1B, I observed that all of the proteins are additionally present in the cytoplasm, and arguably at the plasma membrane (though this may simply be the junction between the cytoplasm and plasma membrane) (Fig. 3.1). There is experimental evidence to suggest that Npl4, Npl4L, Ufd1A and Ufd1B act in the nucleus to engage Cdc48 in the process of chromatin-associated protein degradation (Mérai *et al.*, 2014). Chromatin is a dynamic complex, interacting with various proteins during processes such as DNA repair, transcription and DNA replication (Vaz *et al.*, 2013). UPS components can be tightly associated with chromatin during periods of high protein demand. The observation that Ufd1C and Ufd1D also localise to the nucleus means that they could participate in similar processes, suggesting a degree of redundancy; alternatively, these two proteins may participate in parallel UPS processes in the nucleus. A consistent observation was that Ufd1D-YFP produced lower fluorescent signals in my experiments than the other proteins. The fluorescence intensity of the images was adjusted in imaging processing using gain, so that the signal of localisation was at a comparable level for all samples (Fig. 3.1). Given that all proteins are driven by the same overexpression promoter in these constructs (CamV35s; p35s), this might indicate a degree of post-translational regulation of this protein. It seems less plausible that there could be stability issues with the YFP fusion, given the sequence similarity to Ufd1A and that this protein shows higher fluorescence intensity.

The exclusive localisation of Ufd1B to the nucleus may suggest that it possesses a function that the other Ufd1 proteins do not, or that it only participates in the subset of Ufd1 functions in the nucleus (Mérai *et al.*, 2014) (Fig. 3.1). This was a reproducible result, and given the localisation is more specific than that the other isoforms, it seems less likely to be an overexpression artifact. This data would also support phylogenetic grouping of Ufd1B in a paralogous group, having acquired different functions or regulation to Ufd1A, C and D. It is somewhat surprising on the basis of protein structure and domain architecture, however, given it is highly similar to Ufd1A and D. Additional BiFC data (not presented) also reproduced the nuclear localisation of Ufd1B with Cdc48. Ufd1B is therefore not physically capable of interacting with CHLORAD, on this basis I decided to cease experiments containing this protein from hereon.

While the protoplast is a useful tool for these experiments, transient overexpression in such a system can be prone to artefactual results. To reproduce this data in a more physiologically relevant system, I generated stable transgenic lines. *Arabidopsis* plants were transformed by *Agrobacterium* mediated floral dipping. Lines were then screened on selective media, and expression of the transgene was validated by fluorescence microscopy and western blotting. The proportion of resistant seed was exceptionally low for these lines. For instance, only one line was identified for Npl4-YFP. This was unfortunately destroyed during the pandemic. At least one line existed for the remainder of the adapter proteins, however. These were propagated until homozygous, and the T4 generation was used for confocal fluorescence microscopy (Fig. 3.2). Broadly, this data reproduces the localisation observed in protoplasts. Fluorescent signal can be observed in the nucleus and cytoplasm. There is arguably some plasma membrane and endoplasmic reticulum specific localisation in both protoplasts and plants also (Hawes *et al.*, 2001) – but this is hard to specifically ascribe without more data. The lower fluorescent signal for Ufd1D was also observed here. It is worth noting that these proteins are also driven by the p35S overexpression promoter. The lack of Npl4-YFP may not be too great an issue, given their seeming redundancy observed

elsewhere (Li *et al.*, 2022; Mérai *et al.*, 2014), and that the other proteins also replicate the previously observed data in protoplasts. It is arguable that Ufd1B transgenic lines should also have been generated, to confirm the sole nuclear localisation. However, I believed that I could confidently rule out its involvement in CHLORAD based on the existing localisation data and BiFC interactions presented later.

Cytoplasmic localisation of the other Npl4 and Ufd1 proteins means they could be involved in a range of UPS processes with Cdc48, such as ERAD. It is not convincing evidence of involvement in CHLORAD, but it does mean that they meet the minimum requirement that these proteins can exist around the chloroplast OEM.

An important element in determining whether Npl4 and Ufd1 participate in CHLORAD, is to investigate whether the proteins do in fact localise to the chloroplast OEM. This is a difficult task, as Npl4 and Ufd1 are soluble proteins, and are thus free to diffuse through the cytosol without a permanent association to membranes that could be observed in free chloroplasts by microscopy. To pursue this aim, I sought to build on previous observations of a CDC48 mutant which retained fluorescence at the chloroplast OEM following breakage of the protoplast cell membrane (Ling *et al.*, 2019). The 'trap' mutant of Cdc48 contains a single amino acid substitution (E581Q) in the D2 ring ATPase domain: this inhibits ATP hydrolysis and 'traps' substrates as they are unable to be extracted or unfolded from their complex or membrane (Dalal *et al.*, 2004). I generated this mutant in frame with a C-terminal fusion of the fluorophore mScarlet. The excitation and emission profiles of this fluorophore allow it to be simultaneously viewed with YFP and chlorophyll, without crosstalk. I then co-transfected constructs overexpressing $Cdc48^{E581Q}$-mScarlet with Npl4 and Ufd1 YFP fusions. A selection of data from this experiment is presented here to explore the observed issue (Fig. 3.3). Protoplasts were broken by applying light pressure to the cover slide. However, the fluorescent signal observed around cell-free chloroplasts could be attributed to autofluorescence, as it was observed in the negative control in both YFP and mScarlet detection windows – though not within cells, which were truly negative (Fig. 3.3). Prior to cell

breakage, fluorescence could be observed in the nucleocytoplasmic compartment for Cdc48, Npl4 and Ufd1(Fig. 3.3). When the cells were broken, it was clearly visible that the fluorescence signal quickly dissipated into the solution (Fig. 3.3). I was therefore unable to determine association with the chloroplast OEM. This is a startling result and questions the previous report that the Cdc48 ‘trap’ mutant can be used to demonstrate sustained chloroplast OEM localisation (Ling *et al.*, 2019).

My interpretations likely would have been aided by the use of additional materials and methods to produce more confident conclusions of localisation. This could have included co-localisation analysis with stains such as 4′,6-diamidino-2-phenylindole, which binds to DNA (to more accurately determine nuclear localisation) (Tarnowski *et al.*, 1991). Additionally, I might also have plant lines expressing organelle marker proteins with fluorescent tags — these may have been used for protoplast isolation and transfection, and similarly have been used to examine co-localisation of signal.

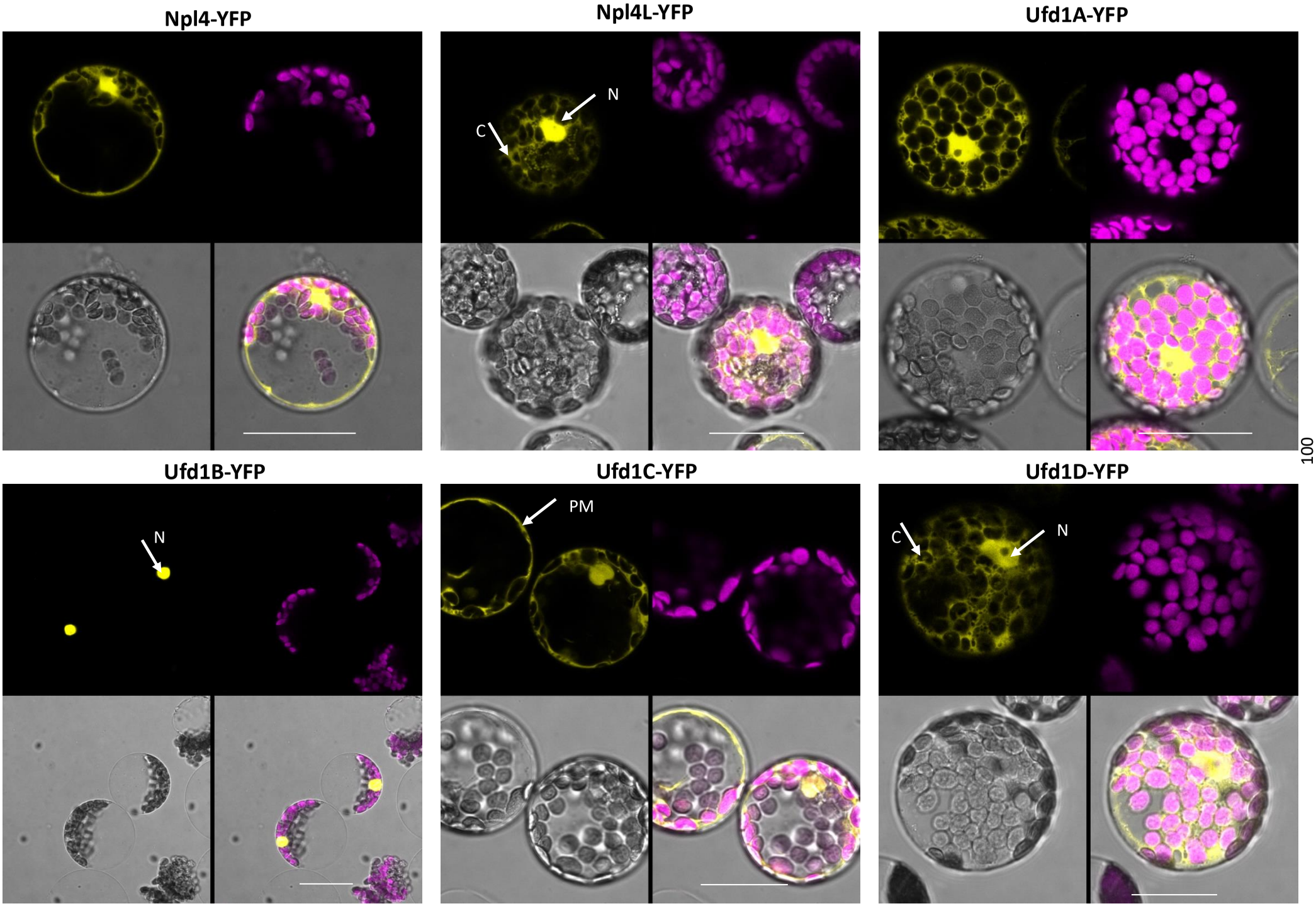

Npl4-YFP
Npl4L-YFP
N
C
Ufd1A-YFP
Ufd1B-YFP
N
Ufd1C-YFP
PM
Ufd1D-YFP
C
N

Figure 3.1 Subcellular localisation analysis of the *Arabidopsis* adapter proteins by confocal microscopy, demonstrating the sole nuclear localisation of Ufd1B. Micrographs of *Arabidopsis* mesophyll protoplasts transfected with Npl4 and Ufd1 isoforms fused to full-length YFP at the C-terminal end, under the control of the 35s promoter. Each panel shows, clockwise from the top left: YFP fluorescence (525 nm - 550 nm), chlorophyll fluorescence (650 nm - 750 nm), brightfield, and all fields merged. Scale bars are displayed as a white line at the bottom of the merged panel, set to a length of 50 µm. White arrows indicate points of pertinent fluorescence: N, nucleus; C, cytosol; PM, plasma membrane.

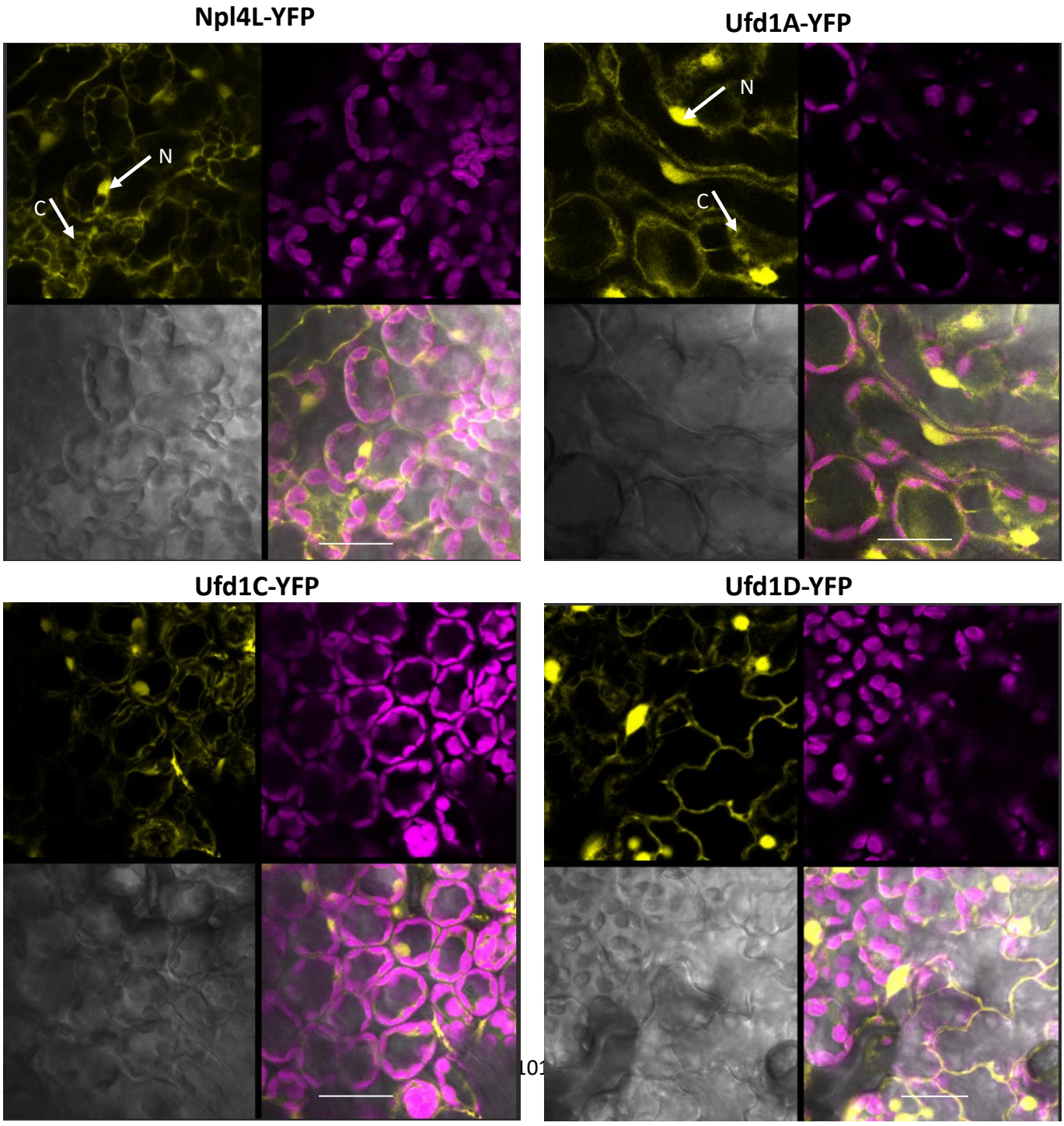


Figure 3.2 Subcellular localisation analysis of the *Arabidopsis* adapter proteins by confocal microscopy *in planta*. Micrographs of the abaxial surface of *Arabidopsis* leaves immersed in perfluorodecalin to enhance the optical quality, showing mesophyll tissue. Transgenic lines contained Npl4 and Ufd1 isoforms fused to full-length YFP at the C-terminal end, under the control of the 35s promoter. Each panel shows, clockwise from the top left: YFP fluorescence (525 nm - 550 nm), chlorophyll fluorescence (650 nm - 750 nm), brightfield, and all fields merged. Scale bars are displayed as a white line at the bottom of the merged panel, set to a length of 50 µm. White arrows indicate points of pertinent fluorescence: N, nucleus; C, cytosol.

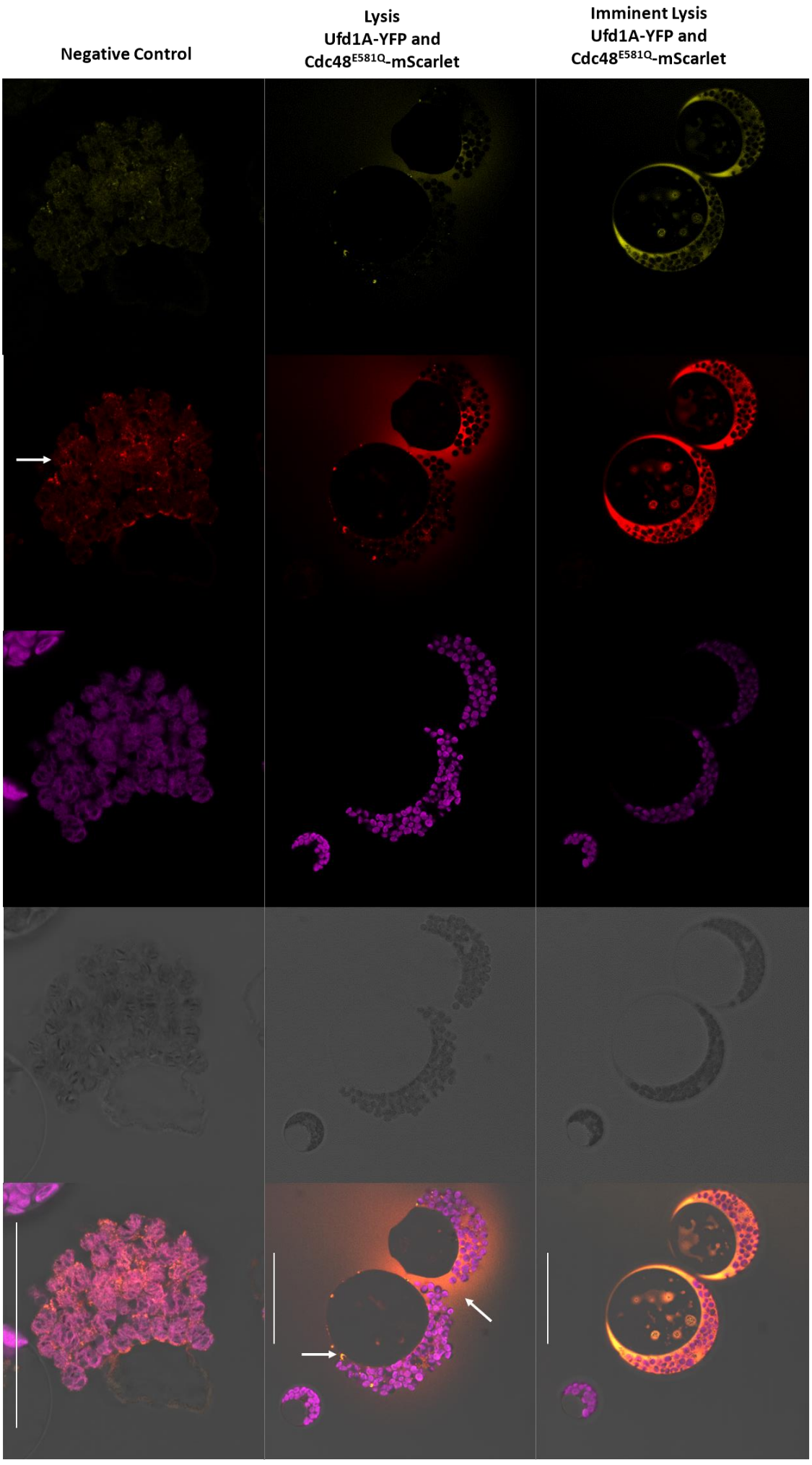

Negative Control
Lysis
Ufd1A-YFP and
Cdc48E581Q-mScarlet
Imminent Lysis
Ufd1A-YFP and
Cdc48E581Q-mScarlet
YFP
mScarlet
Chlorophyll
Brightfield
Merge

Fig. 3.3 The Cdc48 'trap' mutant is unable to demonstrate sustained chloroplast OEM localisation. Micrographs of typical transfected and co-transfected *Arabidopsis* mesophyll protoplasts are shown. Ufd1A was expressed in a translation fusion with the fluorophore eYFP, while Cdc48^{E581Q} was placed in translational fusion with the fluorophore mScarlet (to avoid cross talk with the YFP and chlorophyll channels). The upper and middle panels show co-transfections of Ufd1A-YFP and Cdc48^{E581Q}-mScarlet the same cells at different time points: immediately pre-lysis and immediately post-lysis. The lower panel shows a negative control transfection containing no plasmid DNA. Each channel is displayed in vertical columns, from left to right: YFP fluorescence (525 nm - 540 nm), mScarlet fluorescence (580 nm - 610 nm), chlorophyll fluorescence (650 nm - 750 nm), brightfield, and all fields merged. White arrows indicate points of pertinent fluorescence. Scale bars are displayed as a white line at the bottom of the merged panel, set to a length of 50 µm.

3.3.2 Bimolecular Fluorescence Complementation shows that Npl4 and Ufd1 May Interact with CHLORAD Proteins

Bimolecular fluorescent complementation (BiFC) is a method of testing the interaction between proteins in living cells. To implement this method, constructs were generated containing the genes of interest fused in-frame at the C-terminus with sequences encoding complementary N- or C-terminal fragments of YFP. If there is an interaction between the proteins of interest, the YFP fragments will also interact to form a functional fluorophore and a fluorescent signal can be detected. I aimed to test the interaction of each of the *Arabidopsis* Npl4 and Ufd1 homologues by BiFC in transfected *Arabidopsis* protoplasts using the following components and substrates of CHLORAD: Toc159, Toc33, Cdc48 and SP1. The structure of OMP85 β-barrel proteins render them infeasible targets for BiFC, hence Toc75 and Sp2 were not examined (Ling *et al.*, 2019). In addition, I aimed to similarly test the interactions between the adapters as potential partners within the NU heterodimer. The data presented here are representative of at least two experiments for all interactions.

Due to the exclusively nuclear localisation of Ufd1B, a smaller subset of potential interactors was tested for this component (e.g. Cdc48) and the data are not presented here; any observed signal was again solely nuclear. This reinforces the previous observation that Ufd1B may solely function within the nucleus, in unknown processes.

Initially, I observed that Npl4, Npl4L and Ufd1A, C and D were capable of interacting with Cdc48 (Fig.3.4). These data are curiously different to the fluorescence signals observed in other BiFC experiments, as well as the localisation data for each individual protein. In the other sets of experiments, the distribution of fluorescent signal is far more broad, seemingly occurring in the entirety of the nucleocytoplasmic compartment — here fluorescent signal is detected in more specific patches in the cell. I thought this may be due to mislocalisation or non-native interactions driven by overexpressing highly abundant proteins (Kudla & Bock, 2016). Cdc48 is already highly abundant in the cell, estimated to make up 1% of cytosolic proteins (Broad *et al.*, 2016). I thus performed a titration of plasmid concentrations (ranging from 5 µg to 100 ng of plasmid per construct), but this had no effect on the observed localisations – aside from eliminating fluorescence on the lower end. Plasmid titration is a common approach in protoplast transfection, used to assess whether a signal is the result of an overexpression artifact (Ling *et al.*, 2019); reduced plasmid quantity correlates with a reduction in the abundance of the encoded protein in the transfected cell. A weaker element of the signal in all panels is broadly cytoplasmic, but this is weakest for the Npl4 proteins (Fig. 3.4). The most puzzling signal is that demonstrated in the second panel for Npl4, where large spots of fluorescence are present. These would appear too large to be mitochondria or especially peroxisomes. With all proteins, the pattern would support nuclear and ER localisation, and this is probably best exemplified by Npl4L and Ufd1C. When it comes to localisation of the interaction around the chloroplast, the observed data do not necessarily rule out a role in CHLORAD for all isoforms. Firstly, there is the weak cytosolic background signal, which means adapter proteins can interact at the chloroplast. Secondly, dots of fluorescence around the chloroplast OEM could be observed with each set of proteins – best

shown with the top panels of Npl4 and Npl4L (Fig. 3.4). In some cases these dots were of varying size and did not overlap with the chlorophyll channel, which led me to speculate that these could be peroxisomes.

What is surprising, is how different the observed fluorescence then is within the heterodimer pairs (Fig. 3.5). Here the patterns are more similar to that seen in the YFP localisation data, as the proteins appear to be more broadly nucleocytoplasmic. This might suggest that the heterocomplex can exist and migrate freely and is only more specifically localised when in cooperation with Cdc48 as a functional UPS complex.

The first test of interaction with a CHLORAD specific substrate and component, came with Sp1. It should be noted that proteasome inhibition was required for all interactions to be strongly visible with CHLORAD substrates; this was accomplished by incubating protoplasts with bortezomib (Ling *et al.*, 2019). Sp1 is known to be post-translationally regulated via self-ubiquitination, and so UPS inhibitor application has previously been employed to aid observation of the protein under confocal microscopy (Ling *et al.*, 2012). Sp1 was observed to interact with all Npl4 and Ufd1 proteins (Fig.3.6). These interactions were primarily observed at the chloroplast OEM, where Sp1 is known to be specifically localised (Ling *et al.*, 2017). However, some mislocalisation was also observed at the plasma membrane, and arguably the peroxisome and mitochondria (with Ufd1A), in the case of the Ufd1 proteins.

Interactions with both TOC proteins, Toc33 and Toc159, were visible at the chloroplast OEM as well as in the cytoplasm, plasma membrane, and possibly the cytoskeleton (Fig. 3.7; Fig. 3.8). It is not uncommon to see fluorescent signals for some TOC proteins in the cytoplasm when conducting experiments of this nature, though the reason for this not entirely clear. It could also be a mistargeting artefact. It may have been possible to utilise ruptured chloroplasts here, to determine OEM localisation, though this was unfortunately not attempted. It is known that the A domain of Toc159 is readily cleaved, the construct used in these experiments contains only the G and M domains, with terminus of the fluorophore fused to the N-terminal section of the G domain; hence it is impossible that the A domain can

account for the aberrant signal of Toc159 (Fig.3.8). Another possibility is that it could reflect interactions taking place following substrate extraction from the OEM. I believe this is less likely, as one would assume that it occurs highly proximally to the chloroplast OEM, after which Npl4 and Ufd1 separate from the ubiquitinated substrate to allow it to be degraded by the cytosolic proteasome.

This BiFC data demonstrates that Npl4 and Ufd1 proteins can function in CHLORAD. Interactions were sustained with multiple substrates, and interactions within the heterocomplexes could facilitate activity at the chloroplast OEM. There is sufficient evidence here to speculate about redundancy between isoforms, as there were few observed differences between Npl4 and Ufd1 sets across experiments. Most surprisingly perhaps, this holds true for Ufd1C. From the data presented in Chapter II, it might have been assumed that a more strikingly different set of results would be acquired. Ufd1C may therefore be a functional component of the Npl4-Ufd1-Cdc48 complex and could participate in CHLORAD redundantly with the other Ufd1 proteins.

BiFC does contain inherent limitations: overexpressing a protein can cause artefactual signals within the cell, such as mislocalisation. Reconstitution of the fluorophore is irreversible, and can occur spontaneously when the concentration of the protein exceeds a threshold (Kudla & Bock, 2016) – made more likely by the use of constitutive promoters in BiFC vectors. As a result, BiFC should only be used with appropriate controls and in combination with further independent methods to verify protein-protein interactions.

A suitable negative control to confirm the specificity of the detected interactions was used for the OEM. Fluorescence was confirmed by self-interaction tests. The OEM protein Oep7 (outer envelope protein 7) was used and has been shown to be an adequate negative control for other CHLORAD BiFC data (Ling *et al.*, 2019). No fluorescence was observed in these transfections with Oep7 (Fig. 3.9), confirming the exclusivity of the previously observed interaction with BiFC.

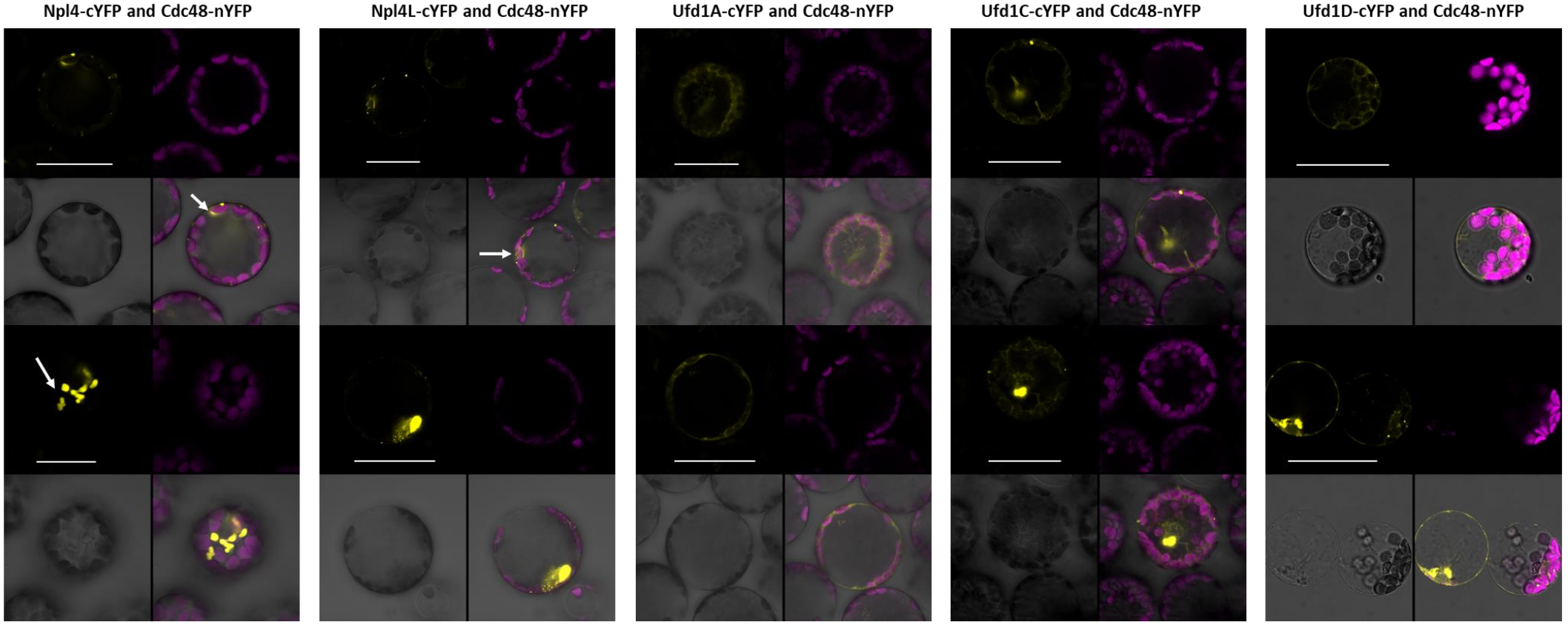

Npl4-cYFP and Cdc48-nYFP
Npl4L-cYFP and Cdc48-nYFP
Ufd1A-cYFP and Cdc48-nYFP
Ufd1C-cYFP and Cdc48-nYFP
Ufd1D-cYFP and Cdc48-nYFP

Fig. 3.4 Bimolecular fluorescence complementation analysis of the interactions of the adapter proteins within Cdc48. Micrographs of typical co-transfected *Arabidopsis* mesophyll protoplasts are shown. Each co-transfection expressed one protein with expressed with a translational fusion of the nYFP fragment, with the other expressing the cYFP fragment. Each panel shows, clockwise from the top left: YFP fluorescence (525 nm - 550 nm), chlorophyll fluorescence (650 nm - 750 nm), brightfield, and all fields merged. Two panels of independent cells are shown per transfected pair, stacked on top of each other. Scale bars are displayed as a white line at the bottom of the YFP fluorescent panel, set to a length of 50 µm. White arrows indicate points of pertinent fluorescence.

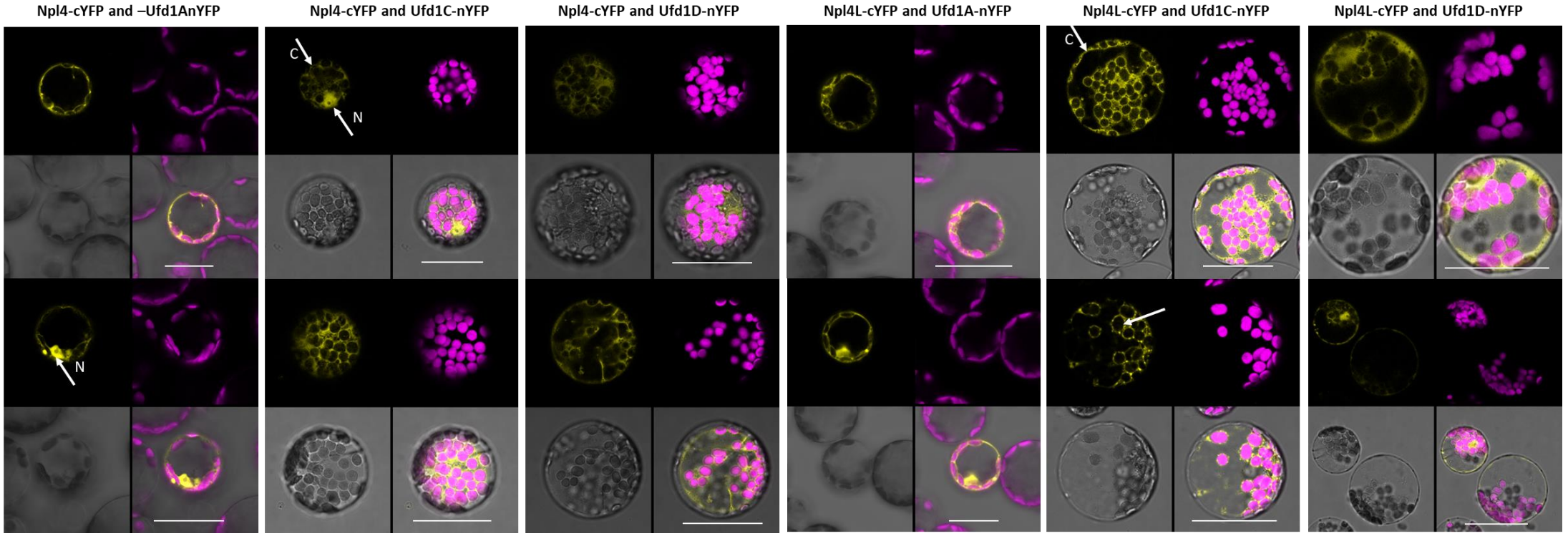

Npl4-cYFP and –Ufd1AnYFP
Npl4-cYFP and Ufd1C-nYFP
Npl4-cYFP and Ufd1D-nYFP
Npl4L-cYFP and Ufd1A-nYFP
Npl4L-cYFP and Ufd1C-nYFP
Npl4L-cYFP and Ufd1D-nYFP
C
N
C
N

Fig. 3.5 Bimolecular fluorescence complementation analysis of the interactions of the adapter proteins within the heterodimer. Micrographs of typical co-transfected *Arabidopsis* mesophyll protoplasts are shown. Each co-transfection expressed one protein with expressed with a translational fusion of the nYFP fragment, with the other expressing the cYFP fragment. Each panel shows, clockwise from the top left: YFP fluorescence (525 nm - 550 nm), chlorophyll fluorescence (650 nm - 750 nm), brightfield, and all fields merged. Two panels of independent cells are shown per transfected pair, stacked on top of each other. Scale bars are displayed as a white line at the bottom of the merged panel, set to a length of 50 µm. White arrows indicate points of pertinent fluorescence: N, nucleus; C, cytosol.

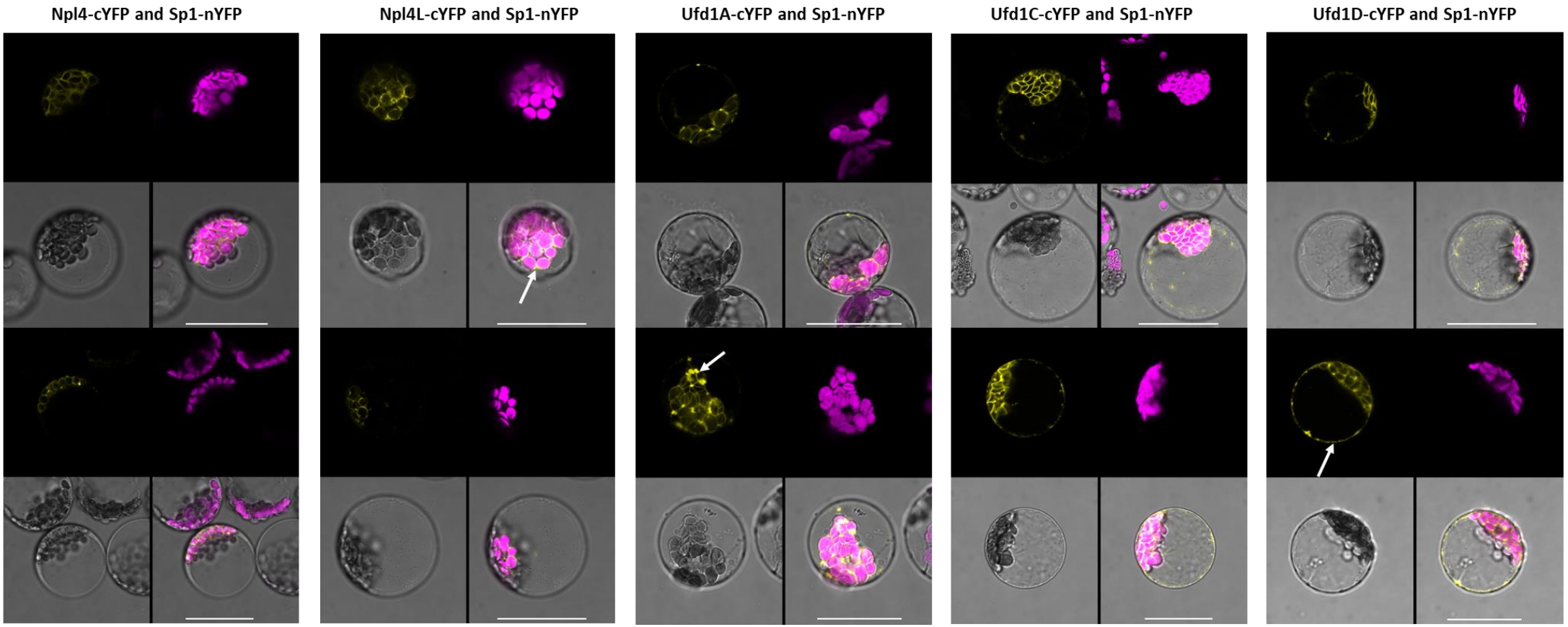

Npl4-cYFP and Sp1-nYFP
Npl4L-cYFP and Sp1-nYFP
Ufd1A-cYFP and Sp1-nYFP
Ufd1C-cYFP and Sp1-nYFP
Ufd1D-cYFP and Sp1-nYFP

Fig. 3.6 Bimolecular fluorescence complementation analysis of the interactions of the adapter proteins with Sp1. Micrographs of typical co-transfected *Arabidopsis* mesophyll protoplasts are shown. Each co-transfection expressed one protein with expressed with a translational fusion of the nYFP fragment, with the other expressing the cYFP fragment. Each panel shows, clockwise from the top left: YFP fluorescence (525 nm - 550 nm), chlorophyll fluorescence (650 nm - 750 nm), brightfield, and all fields merged. Two panels of independent cells are shown per transfected pair, stacked on top of each other. Scale bars are displayed as a white line at the bottom of the merged panel, set to a length of 50 µm. White arrows indicate points of pertinent fluorescence.

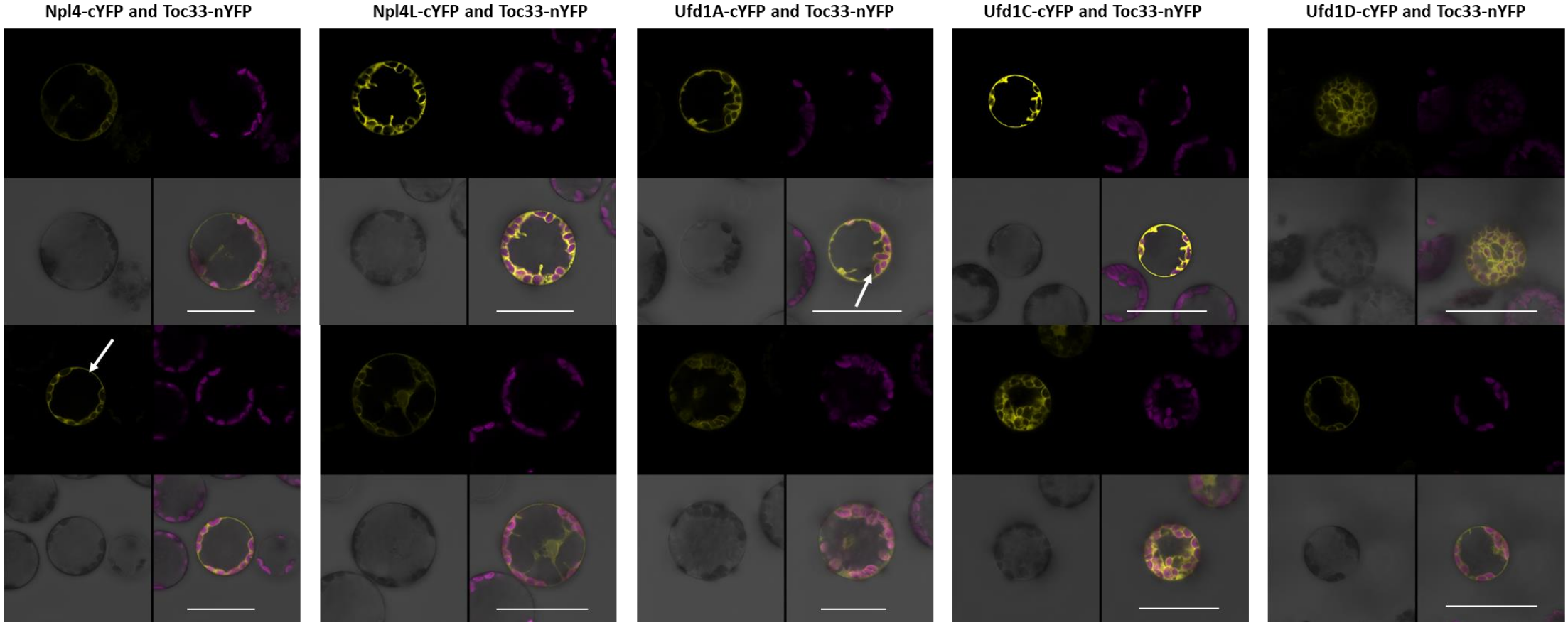

Npl4-cYFP and Toc33-nYFP
Npl4L-cYFP and Toc33-nYFP
Ufd1A-cYFP and Toc33-nYFP
Ufd1C-cYFP and Toc33-nYFP
Ufd1D-cYFP and Toc33-nYFP

Fig. 3.7 Bimolecular fluorescence complementation analysis of the interactions of the adapter proteins with Toc33. Micrographs of typical co-transfected *Arabidopsis* mesophyll protoplasts are shown. Each co-transfection expressed one protein with expressed with a translational fusion of the nYFP fragment, with the other expressing the cYFP fragment. Each panel shows, clockwise from the top left: YFP fluorescence (525 nm - 550 nm), chlorophyll fluorescence (650 nm - 750 nm), brightfield, and all fields merged. Two panels of independent cells are shown per transfected pair, stacked on top of each other. Scale bars are displayed as a white line at the bottom of the merged panel, set to a length of 50 µm. White arrows indicate points of pertinent fluorescence.

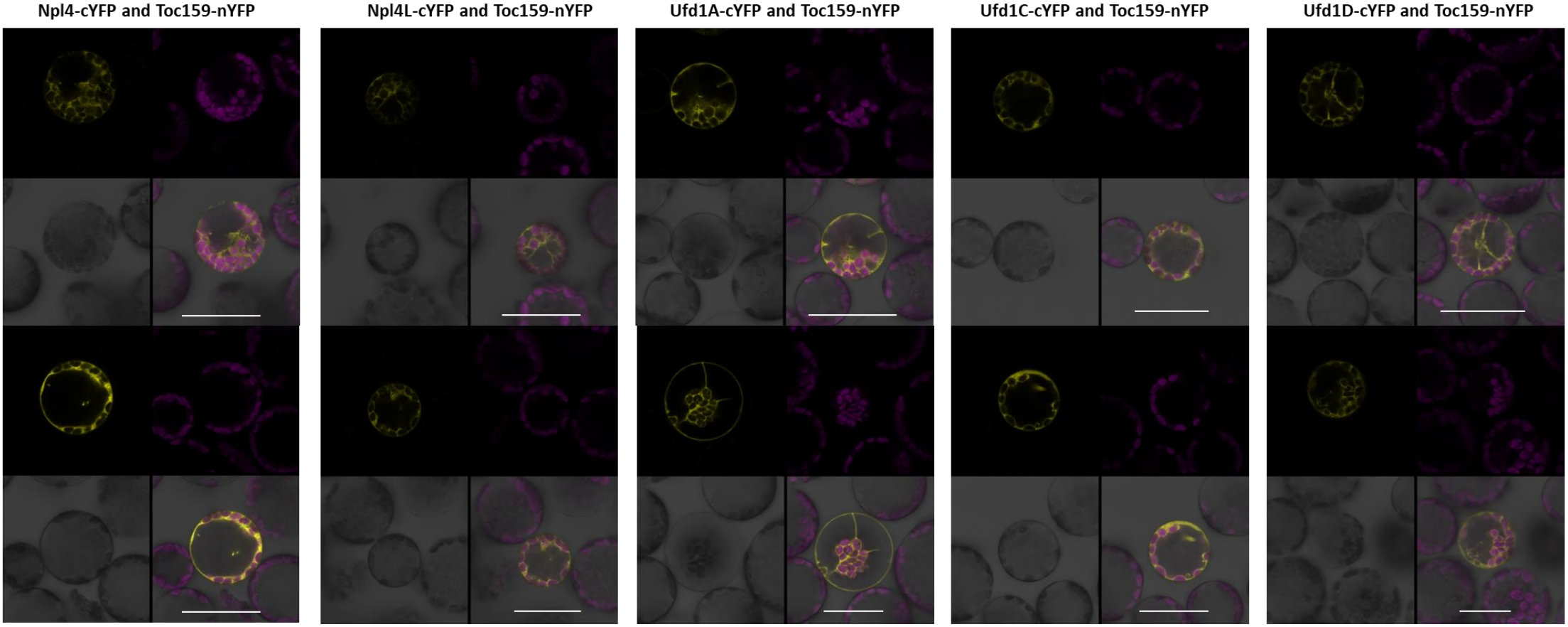

Npl4-cYFP and Toc159-nYFP
Npl4L-cYFP and Toc159-nYFP
Ufd1A-cYFP and Toc159-nYFP
Ufd1C-cYFP and Toc159-nYFP
Ufd1D-cYFP and Toc159-nYFP

Fig. 3.8 Bimolecular fluorescence complementation analysis of the interactions of the adapter proteins with Toc159. Micrographs of typical co-transfected *Arabidopsis* mesophyll protoplasts are shown. Each co-transfection expressed one protein with expressed with a translational fusion of the nYFP fragment, with the other expressing the cYFP fragment. Each panel shows, clockwise from the top left: YFP fluorescence (525 nm - 550 nm), chlorophyll fluorescence (650 nm - 750 nm), brightfield, and all fields merged. Two panels of independent cells are shown per transfected pair, stacked on top of each other. Scale bars are displayed as a white line at the bottom of the merged panel, set to a length of 50 µm.

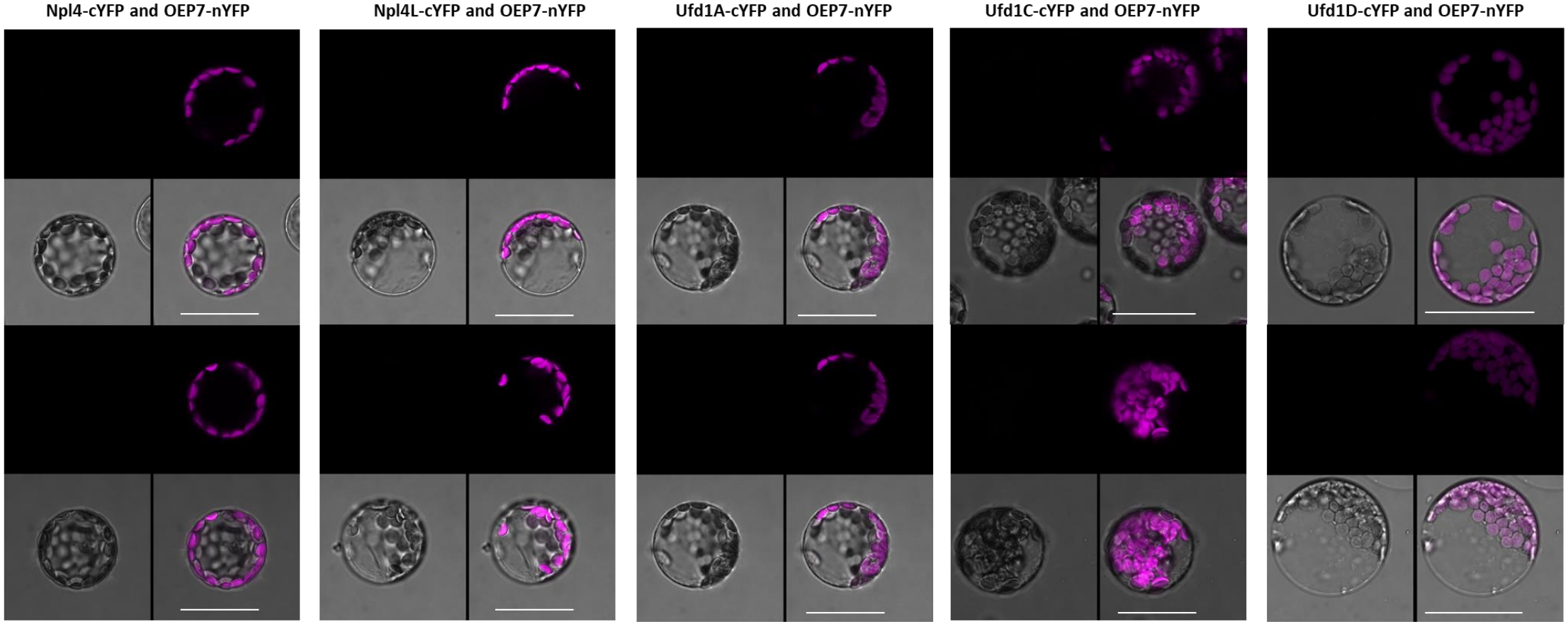

Npl4-cYFP and OEP7-nYFP
Npl4L-cYFP and OEP7-nYFP
Ufd1A-cYFP and OEP7-nYFP
Ufd1C-cYFP and OEP7-nYFP
Ufd1D-cYFP and OEP7-nYFP

Fig. 3.9 Bimolecular fluorescence complementation analysis of the interactions of the adapter proteins with the negative control OEP7. Micrographs of typical co-transfected *Arabidopsis* mesophyll protoplasts are shown. Each co-transfection expressed one protein with expressed with a translational fusion of the nYFP fragment, with the other expressing the cYFP fragment. Each panel shows, clockwise from the top left: YFP fluorescence (525 nm - 550 nm), chlorophyll fluorescence (650 nm - 750 nm), brightfield, and all fields merged. Two panels of independent cells are shown per transfected pair, stacked on top of each other. Scale bars are displayed as a white line at the bottom of the merged panel, set to a length of 50 µm

3.3.3 Co-Immunoprecipitations Identify Physical Interactions with CHLORAD Proteins Sp1, Sp2, Toc33 and Cdc48

In order to validate the interactions observed by BiFC, an additional independent test of the interaction was required. Co-immunoprecipitation (Co-IP) is a widely used and robust technique to investigate the physical interactions between proteins *in vivo*. Isolated *Arabidopsis* mesophyll protoplasts were again used here, transfected with the same YFP constructs used for subcellular localisation. In addition to replicating the observed interactions seen by BiFC, I also intended to examine whether Sp2 can interact with Npl4 and Ufd1, as Sp1 and Sp2 are predicted to form a stable complex (Ling *et al.*, 2019).

Transfected protoplasts were incubated with 5 µM Bortezomib to inhibit the proteasome for four hours and increase the abundance of UPS substrates (such as Sp1). Protoplasts were lysed with IP buffer containing a low concentration of a non-denaturing detergent to maintain native protein complexes; lysates were incubated with epitope specific agarose or magnetic agarose antibody conjugates to pull down protein complexes and eluted under denaturing conditions.

Initially, I attempted to validate interactions with Cdc48. To accomplish this, I used HA epitope tagged Cdc48 in a construct, co-transfected with Npl4 isoforms with YFP fusions and Ufd1 isoforms with Myc epitope tags (Fig. 3.10). The Ufd1 isoforms were expressed with YFP fusions to a lower level in trials. There was therefore a concern that the protein would not be visible in the IP, even if there was a real interaction. To increase the chance of detection, Myc tagged isoforms which showed stronger signal intensity were used (this may be an antibody related phenomenon. Here, Cdc48-HA was the bait protein immunoprecipitated with anti-HA beads. For experimental controls, I transfected protoplasts solely with Cdc48-HA, as well as blotted with the non-CHLORAD substrate Tic40 (Fig. 3.9). This experiment confirmed predicted interactions and those seen by BiFC: all isoforms of Npl4 and Ufd1 are capable of interacting with Cdc48. There is not appreciable signal enrichment in the IP for Cdc48. Noticeably a similar level of enhancement is observed with the Npl4 isoforms, but not with the Ufd1 isoforms. This may suggest a stronger interaction between Npl4 and Cdc48, than between Cdc48 and Ufd1. The interactions appeared to be specific, as there was no pulldown of the control protein Tic40 or any non-specific bands in the single transfection. During this experiment, I also attempted to examine interactions with TOC proteins natively. Somewhat surprisingly, Cdc48 was not able to pulldown Toc159 and Toc33, as previously observed (Ling *et al.*, 2019). It does not appear to be due to expression of the heterocomplex as a whole, as there is also no pulldown in the single Cdc48-HA lane (Fig. 3.10). It may be a signal detection issue. The quantity of Toc159 and Toc33 observed in the total lysate in the published experiment, is of far higher quantity, and the IP represents a very small fraction of that protein that is actually pulled down. There is also a difference between the banding of Toc159 in the total lysate of lanes 1-3, compared to lanes 4-7 (Fig. 3.10); this is likely due to transfer efficiency during the western blot, as this section was on a different membrane, and this was reflected in the poorer transfer of molecular weight marker bands. Alternatively, it might be related to the expression of Npl4, though this is highly speculative.

Interactions within the heterodimer using Co-IP, also validated the interactions from BiFC (Fig. 3.11). Here, Npl4 or Npl4L were expressed as YFP fusions in co-transfections with Ufd1 isoforms containing Myc peptide translational fusions. Anti-GFP beads were used to pull down YFP (due to the conserved sequence between GFP and YFP), and membranes were probed with anti-myc to detect the bait. Single transfections of Npl4/L-YFP showed no non-specific interactions. Two TIC proteins, Tic40 and Tic110 were blotted here as controls, and neither showed an interaction with Npl4 or Npl4L (Fig. 3.11). Despite their structural differences, all Ufd1 proteins appear to be equally capable of interacting with Npl4 and Npl4L. It is also apparent with the Myc fusions of Ufd1 proteins, that there is a degree of band smearing. This may indicate post translational modifications of these proteins – indeed, this is already known via phosphorylation of Ufd1A at least (Galvão *et al.*, 2008). Toc159 was also blotted here, but no native interaction was observed. As before with Cdc48, this may be a signal threshold issue.

To determine whether the CHLORAD substrate Toc33 could physically interact with Npl4 and Ufd1 proteins, I employed a transient overexpression system to increase the detectable quantity of Toc33. Toc33-HA was co-transfected with Npl4 and Ufd1 proteins in translational fusion with YFP. YFP was used as bait with anti-GFP beads, again because of the low detectable quantities of YFP tagged proteins in the total lysates. Significant enrichment of Npl4 and Ufd1 can be observed in the IPs here (Fig. 3.12a). The Toc33-HA signal was detected in all IP samples, including the control. To see if there were significant differences between the Npl4 and Ufd1 proteins with the control, I quantified the protein abundance in the IP lanes from two biological replicates (Fig. 3.12b). Using an ANOVA test followed by a Tukey post-hoc test, I was able to determine that only Npl4, Npl4L and Ufd1C showed significant differences ($\alpha = 0.05$) to the intensity of the Toc33-HA pulldown in the negative control. This is most strongly observed with Npl4, while Npl4L and Ufd1C show a similar level of pulldown. These results do not convincingly replicate the findings of BiFC therefore. However, they may support some redundancy between Npl4 and Npl4L, and possibly that

Npl4 has stronger preference for involvement in CHLORAD. The lane order was kept consistent in loading, so it may be possible that there was consistently some cross-contamination of Npl4 from the total lysate. This is an intriguing result for Ufd1C, which may strengthen the argument for its involvement in CHLORAD over Ufd1A and D. Once again, I was not able to observe a pulldown with Toc159 natively, however. In combination, these data may support higher protein expression and/or detection being necessary to validate the interaction with TOCs.

I next investigated interactions with the E3 ligase Sp1 (Fig. 3.12a). The arrangement of this experiment is much the same as for the Toc33-HA pulldown. Here, YFP isoforms of Npl4 and Ufd1 were used as bait proteins, with transiently co-expressed Sp1 tagged with HA. However, in this experiment I also used a co-transfection as a control using the protein Tic40-YFP - as a stringent control for the interaction with the epitope tag. As before, I detected pulldown of the prey protein (Sp1-HA) in all lanes of the IP (Fig. 3.13a). This time I quantified both the protein abundance in the IPs of Sp1-HA and with each of the YFP fusion proteins, from two biological replicates (Fig. 3.13b). This demonstrated that the control Tic40-YFP expression was far higher than that observed with the Npl4 and Ufd1 proteins. Thus, as a ratio of IP:IP of x-YFP to Sp1-HA, the Tic40-YFP ratio was at least 10-fold weaker than that of the Npl4 and Ufd1 proteins. Statistically comparing the means with an ANOVA test, showed that there were significant differences between the IP intensities of Npl4 and Ufd1 protein pulldowns of Sp1-HA, compared to Tic40-YFP ($\alpha = 0.05$); with at least 2-fold higher intensities with the adapter proteins (Fig. 3.13b). I therefore tentatively conclude that interactions with Sp1 are validated by Co-IP. There are also fewer differences between the adapter proteins here with this CHLORAD substrate, when compared to Toc33.

Initial attempts to test the interaction with SP2, used transient expression similar to other experiments; this did not prove successful. After contacting the individual who worked on Sp2, it emerged that the chloroplastic localisation of transiently expressed Sp2 could not be verified. I therefore utilised stable transgenic (T3) lines overexpressing Sp2-Myc, for which

the localisation was confirmed (Ling *et al.*, 2019). All adapter proteins can interact with Sp2 by Co-IP (Fig. 3.14). This experiment was most similar in setup to the test of Sp1 interaction: Npl4 and Ufd1 YFP fusions were bait, and I used a transfection as negative control – in this instance it was YFP-HA. Sp2 commonly shows multiple banding in blots, which may be indicative of post translational modifications or degradation. There is an extremely minor contaminating band in the control IP, which I did not think necessitated quantification. It is notable that the Npl4 isoforms and Ufd1C consistently pulldown a greater fraction of Sp2-Myc, which reflects the observation also seen with Toc33-HA. Collectively however, I have observed general redundancy between Npl4 and Ufd1 isoforms. It may be argued that Ufd1C has preferential interaction with select CHLORAD clients, compared to Ufd1A or D. This is unlikely to be based on structure or the strength of membrane association of the substrates, as Sp1 is more similar to Toc33 in this regard (as they both possess simple alpha helical transmembrane domain/s, compared to the β-barrel of Sp2). It could however reflect differences in ubiquitination patterns. Sp2 is not a known substrate of CHLORAD however, so physical interaction is likely mediated transiently by some unknown mechanism or recruiting protein - as Sp2 has no predicted domains which may interface with the Cdc48 complex.

Finally, I would like to state that I initially endeavoured to validate physical interactions further using stable transgenic lines expressing TAP fusions. These proteins would have been TAP purified and submitted for mass spectrometry, followed by proteomic analysis. The aims of this were to firstly validate interactions by a third technique, and secondly to establish a substrate repertoire for each Npl4 and Ufd1 isoform. This may have aided in the assessment of redundancy and which proteins have clearer involvement in CHLORAD or chloroplast related processes. Unfortunately, this aspect was not completed due to financial constraints and difficulties imposed by the Covid-19 pandemic.

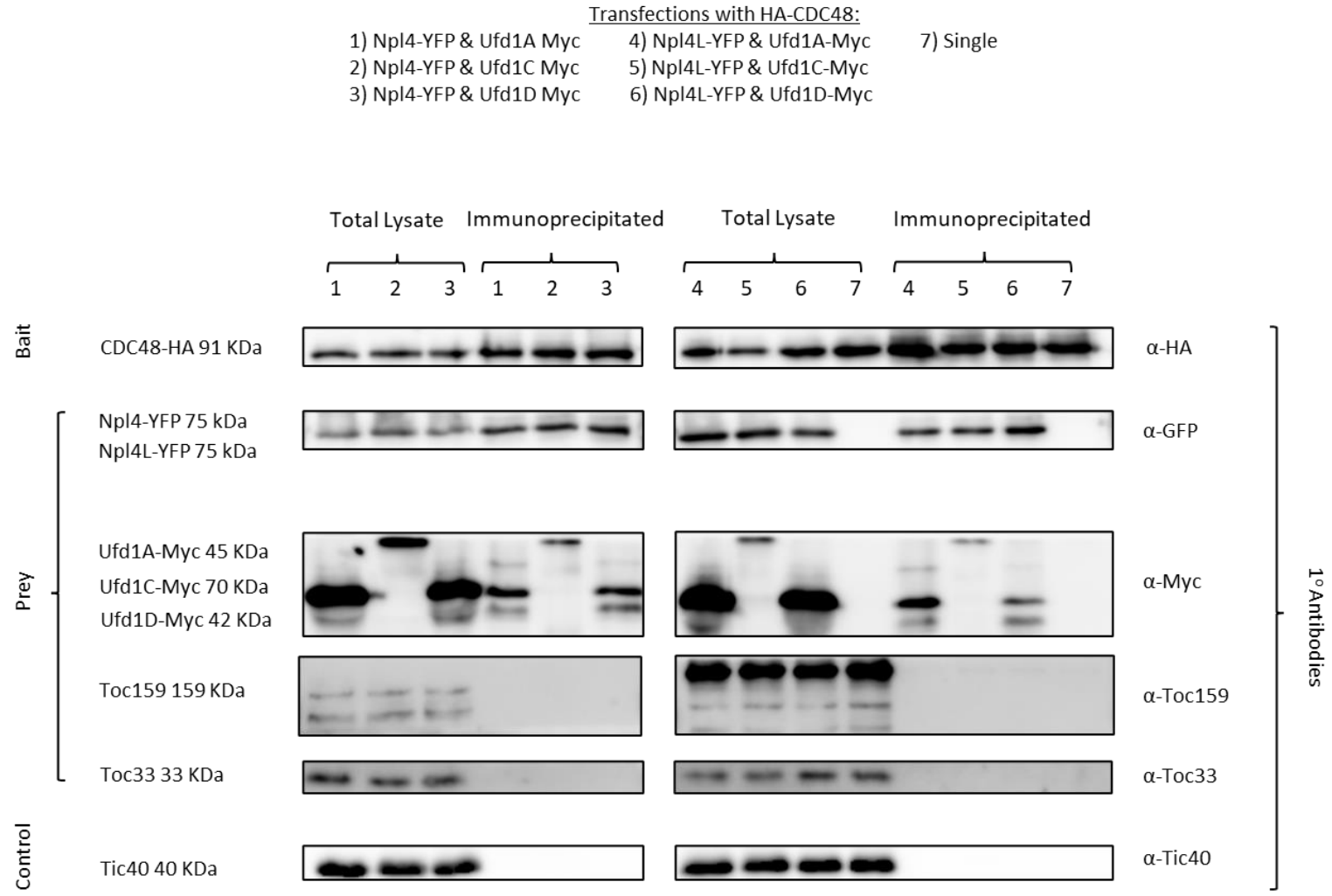


Fig. 3.10 Co-Immunoprecipitation analysis of the interactions of the adapter proteins with Cdc48-HA. Western blots from proteins extracted from co-transfected *Arabidopsis* mesophyll protoplasts. Npl4 isoforms expressed C-terminal YFP fusions, and were co-transfected with Ufd1 isoforms with C-terminal 6xMyc peptides, as well as Cdc48 translationally fused to 3xHA peptides at the C terminus; a single transfection of Cdc48-HA acted as a control. The total lysate was taken after cell lysis and debris pelleting. Anti-HA beads were used for immunoprecipitation and incubated with the lysate for one hour at 4 °C. Non-CHLORAD substrate proteins (TICs) were used as additional controls. Proteins probed and their masses are shown on the left hand side, with the antibody used for detection on the right hand side. The lane order is shown by the key above the figure.

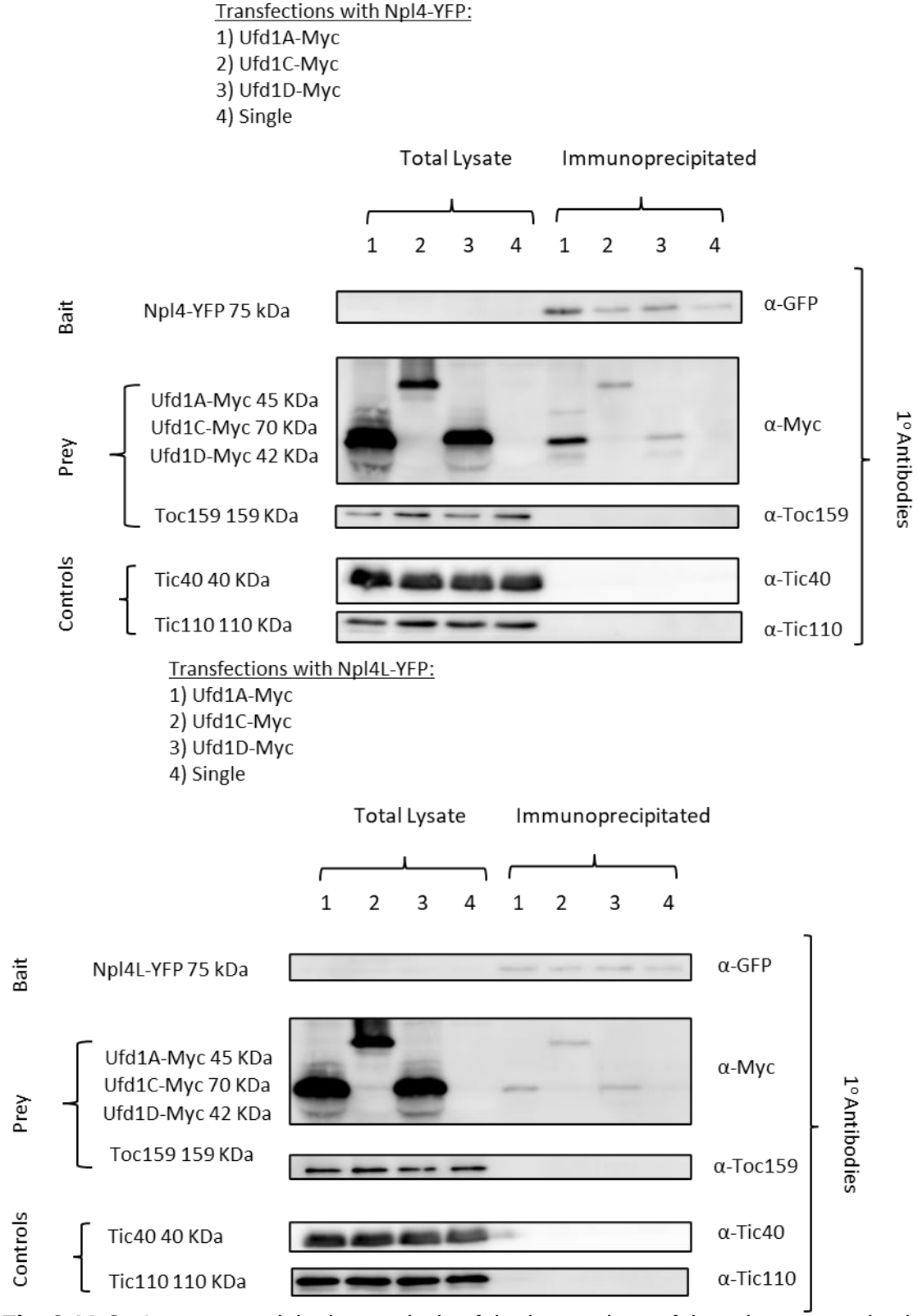
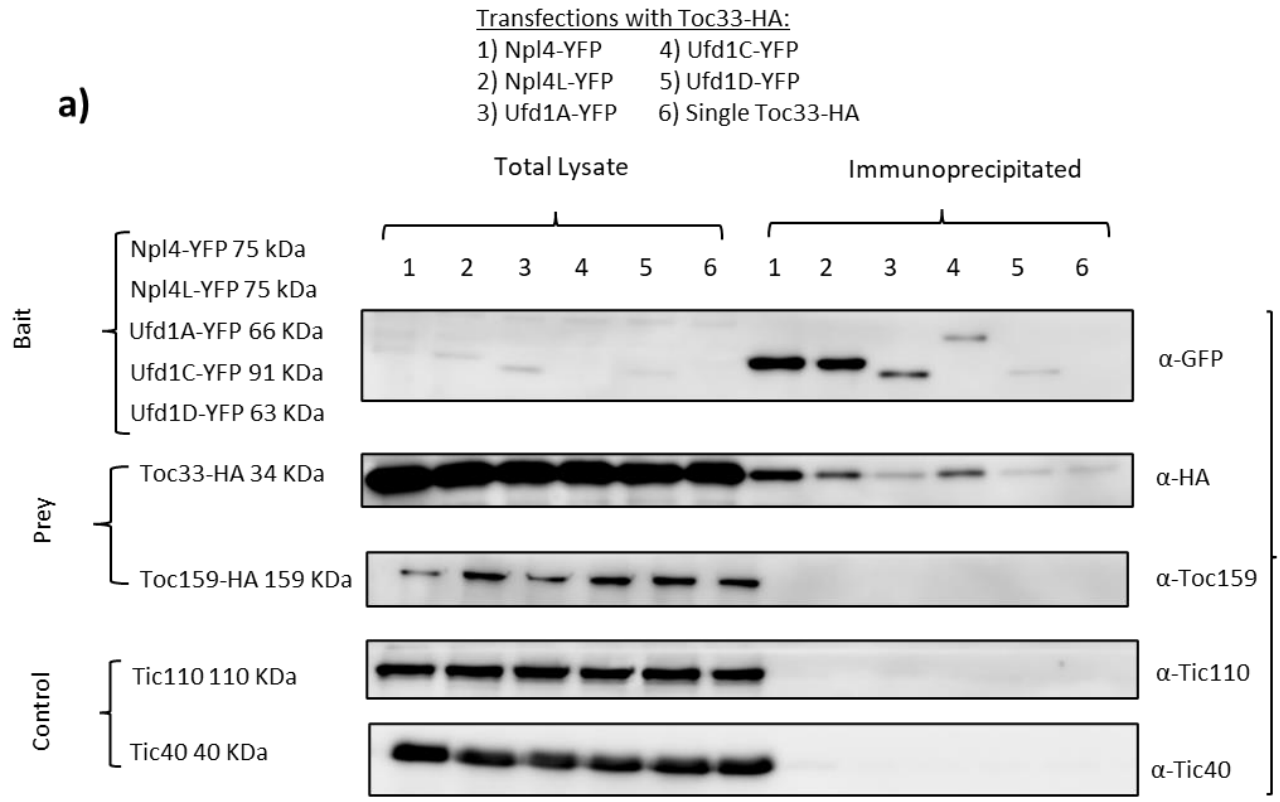


Fig. 3.11 Co-Immunoprecipitation analysis of the interactions of the adapter proteins in heterodimers. Western blots from proteins extracted from co-transfected *Arabidopsis* mesophyll protoplasts. Npl4 isoforms expressed C-terminal YFP fusions, and were co-transfected with Ufd1 isoforms with C-terminal 6xMyc peptides; a single transfection of Npl4-YFP or Npl4L-YFP acted as a control. The total lysate was taken after cell lysis and debris pelleting. Anti-GFP beads were used for immunoprecipitation and incubated with the lysate

for one hour at 4 °C. Non-CHLORAD substrate proteins (TICs) were used as additional controls. Proteins probed and their masses are shown on the left hand side, with the antibody used for detection on the right hand side. The lane order is shown by the key above the figure.



b)

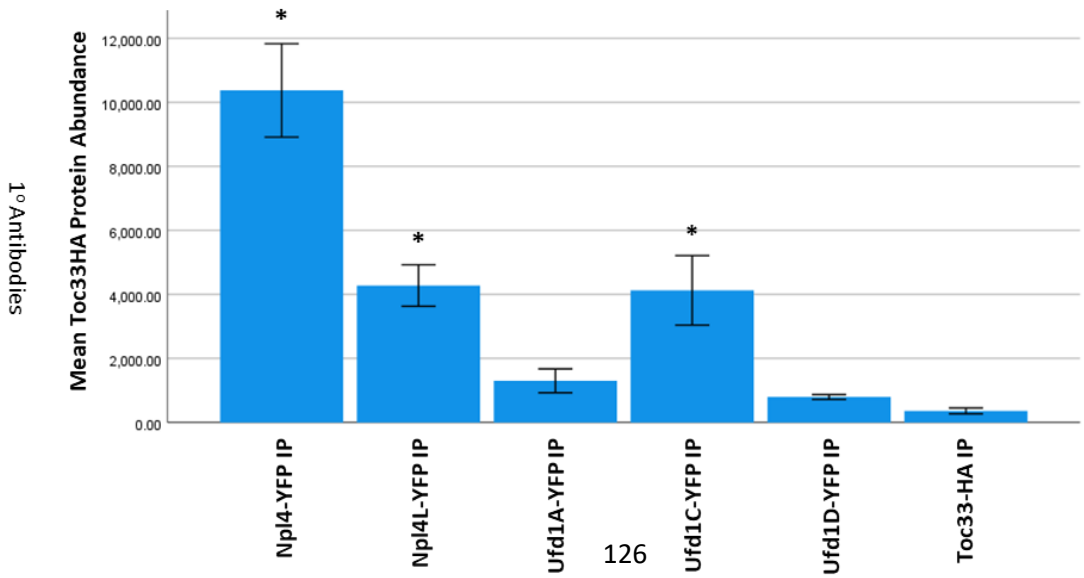
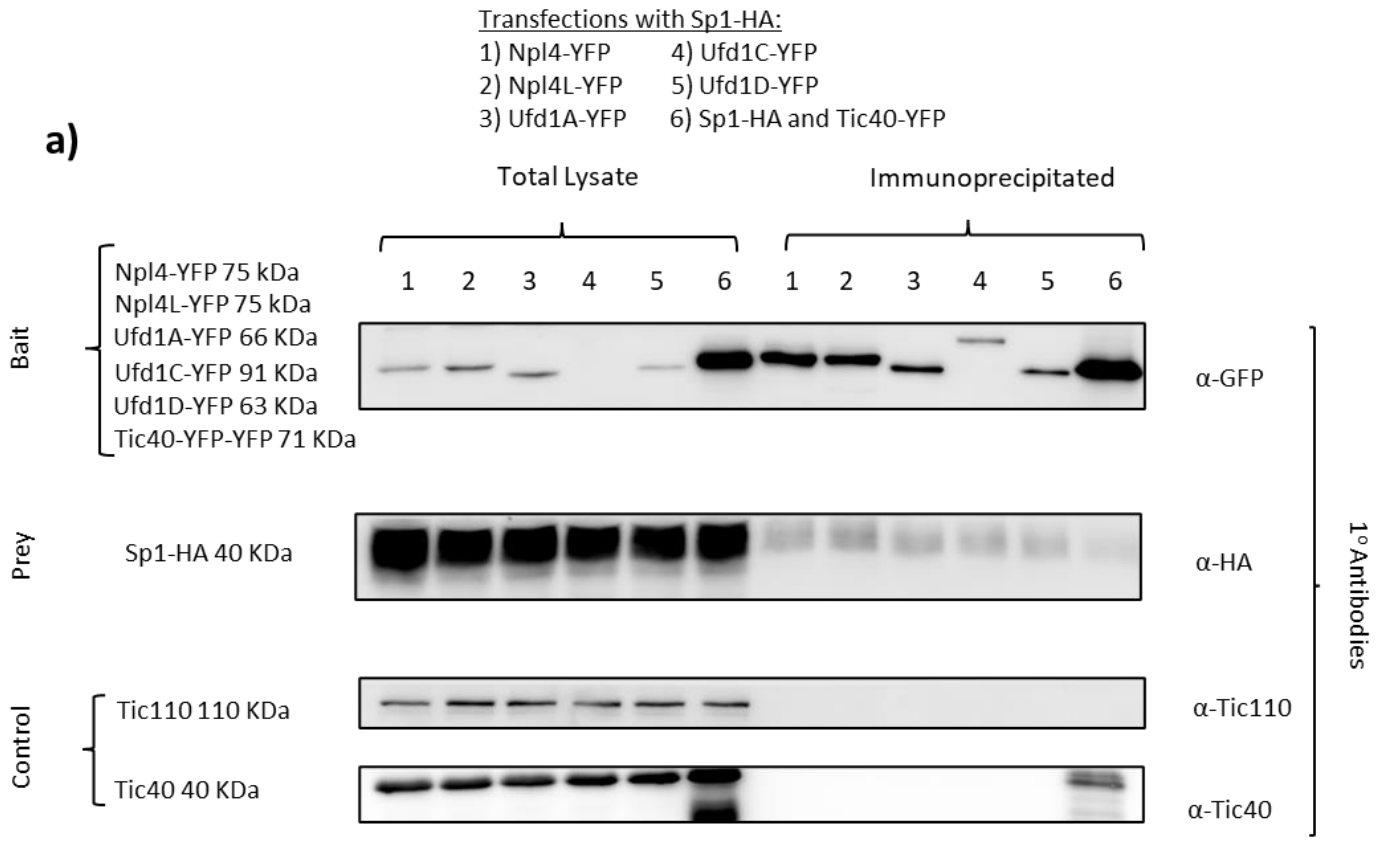


Fig. 3.12 Co-Immunoprecipitation analysis of the interactions of the adapter proteins with Toc33-HA. **a)** Western blots from proteins extracted from co-transfected *Arabidopsis* mesophyll protoplasts. Npl4 and Ufd1 isoforms expressed C-terminal YFP fusions, and were co-transfected with Toc33 in a translational fusion with 3xHA peptides; a single transfection of Toc33-HA acted as a control. The total lysate was taken after cell lysis and debris pelleting. Anti-GFP beads were used for immunoprecipitation and incubated with the lysate for one hour at 4 °C. Non-CHLORAD substrate proteins (TICs) were used as additional controls. Proteins probed and their masses are shown on the left hand side, with the antibody used for detection on the right hand side. The lane order is shown by the key above the figure. **b)** A simple bar graph of protein abundance of Toc33-HA in immunoprecipitated samples. Protein abundance was quantified using the gel analysis tool in ImageJ. Bars labelled with an asterisk show samples with significant (α = 0.05) differences to the protein abundance in single Toc33-HA transfection. Error bars are ± 1 Standard Error of the mean.



b)

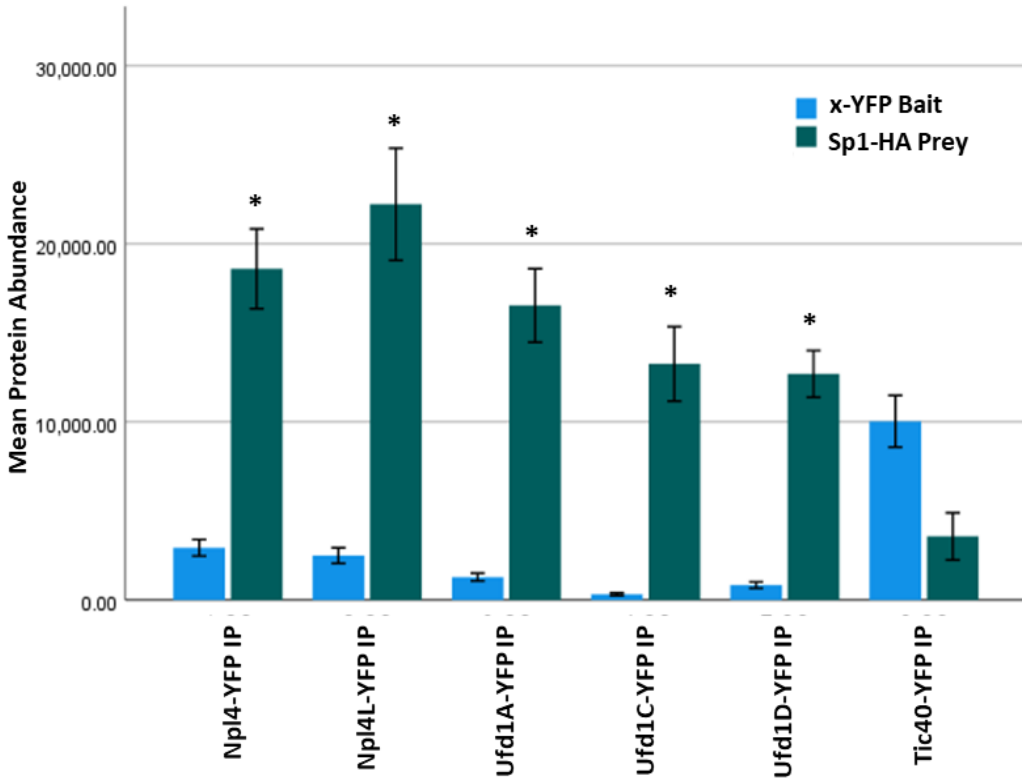


Fig. 3.13 Co-Immunoprecipitation analysis of the interactions of the adapter proteins with Sp1-HA. **a)** Western blots from proteins extracted from co-transfected *Arabidopsis* mesophyll protoplasts. Npl4 and Ufd1 isoforms expressed C-terminal YFP fusions, and were co-transfected with Sp1-HA in a translational fusion with 3xHA peptides; a co-transfection with Tic40-YFP. The total lysate was taken after cell lysis and debris pelleting. Anti-GFP beads were used for immunoprecipitation and incubated with the lysate for one hour at 4 °C. Non-CHLORAD substrate proteins (TICs) were used as additional controls. Proteins probed and their masses are shown on the left hand side, with the antibody used for detection on the right hand side. The lane order is shown by the key above the figure. **b)** A compound bar graph of protein abundance of protein in immunoprecipitated samples with x-YFP or SP1-HA. Protein abundance was quantified using the gel analysis tool in ImageJ. Bars labelled with an asterisk show samples with significant (α = 0.05) differences to the protein abundance of Sp1-HA, in the control Tic40-YFP co-transfection. Error bars are ± 1 Standard Error of the mean.

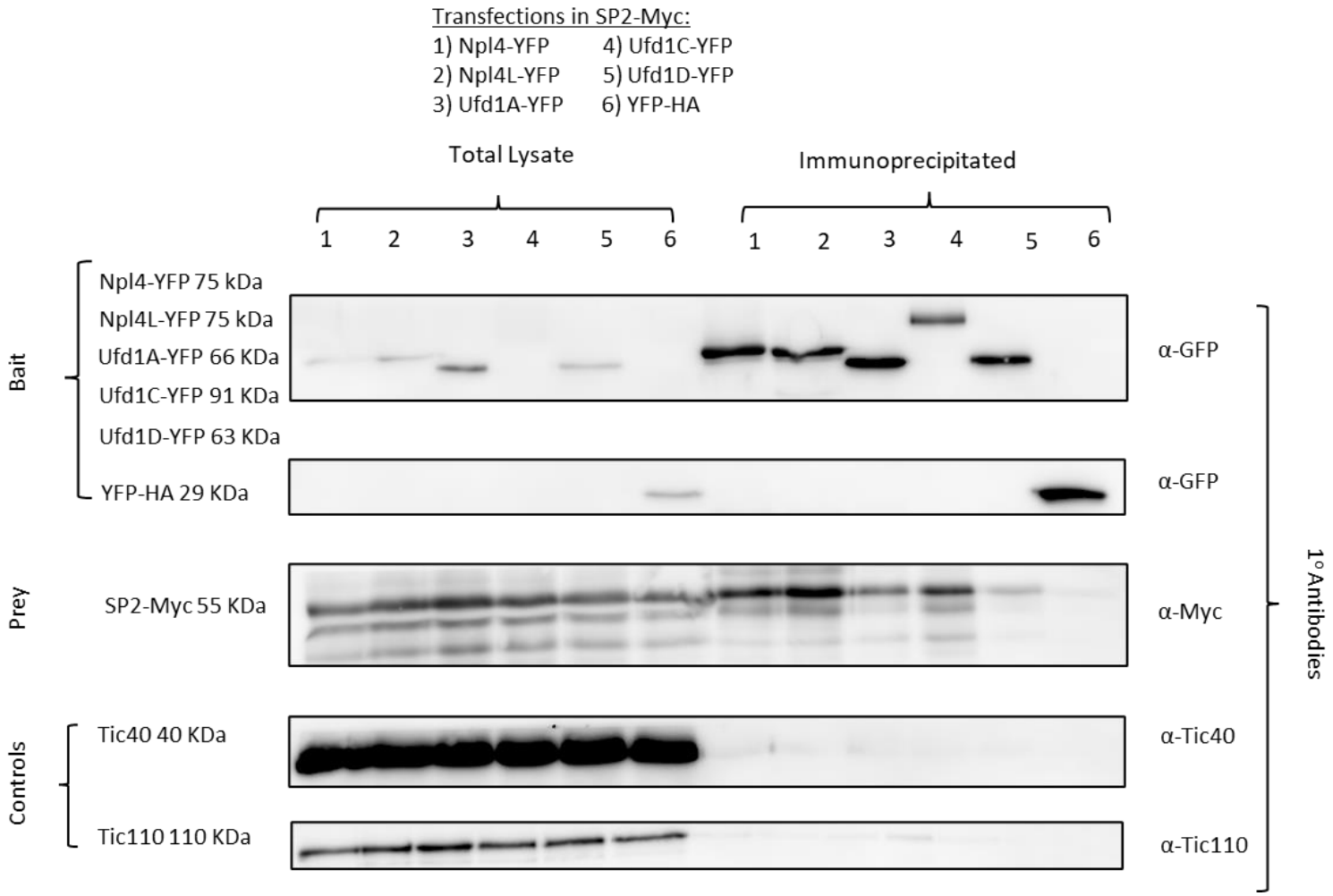


Fig. 3.14 Co-Immunoprecipitation analysis of the interactions of the adapter proteins with Sp2-Myc. Western blots from proteins extracted from transfected *Arabidopsis* mesophyll protoplasts, isolated from a stably expressing Sp2-Myc line (T3 generation). Npl4 and Ufd1 isoforms expressed C-terminal YFP fusions, and were transfected into isolated protoplasts; a co-transfection with YFP-HA. The total lysate was taken after cell lysis and debris pelleting. Anti-GFP beads were used for immunoprecipitation and incubated with the lysate for one hour at 4 °C. Non-CHLORAD substrate proteins (TICs) were used as additional controls. Proteins probed and their masses are shown on the left hand side, with the antibody used for detection on the right hand side. The lane order is shown by the key above the figure.

Chapter IV

Results Chapter III: Characterisation of Mutant Lines

4.1 Abstract

Genetic analysis has been at the heart of characterising the functional significance of CHLORAD. To this end, I aimed to pursue investigations of Npl4 and Ufd1 mutant lines in plant development and in their response to stress, in addition to basic general characterisation. Unfortunately, this chapter is severely limited by a lack of valuable double mutants to TOC and CHLORAD knock-out genotypes. The suppressive effect of *npl4* and *ufd1* mutants could therefore not be examined by genetic interaction, which would have supported a functional role in CHLORAD. Instead, I utilised T-DNA lines and generated overexpressing (OX) lines with YFP or TAP fusions, to conduct experimental analysis. Gene expression analysis reveals that the T-DNA lines are genuine knockouts and that several Np4 and Ufd1 genotypes with YFP fusions are practical overexpressing lines. Examination of the role of these mutants in development, by de-etiolation and senescence, did not reveal any distinct phenotype. However, a strong phenotype was detected for OX lines of Npl4 and Ufd1D with TAP tags, especially in response to osmotic stress. This manifested in overaccumulation phenotype of the Toc75 protein. This effect is seemingly independent of gene expression and may be the product of dominant negative mutation generated by the TAP tag. Overall, it is difficult to draw many functional conclusions from this chapter. The results are largely negative, but this does not preclude an involvement of Npl4 and Ufd1 in CHLORAD. The topics of development and stress response will need to be assessed in the future with a combination of appropriate mutant lines.

4.2 Introduction

The establishment of the CHLORAD model began with the identification of the Sp1 protein and its ability to selectively ubiquitinate the protein import apparatus. Activation of the pathway has been shown to have an adaptive function, through the assessment of single

and double mutants of *sp1* and *sp1 ppi1,* showing that it can aid measures of chloroplast biology during de-etiolation and senescence (Ling *et al.*, 2012). It later emerged that CHLORAD has a functional role in assisting in the response to abiotic stress (Ling & Jarvis, 2015a). Sp2 would appear to largely phenocopy the trend observed with Sp1 in mutant analysis; Cdc48 also shows some similar measures, though the analysis is more complex due to the lack of viable knock-out lines (Ling *et al.*, 2019).

In Chapter III, I observed that the adapter proteins of Cdc48 are capable of physically interacting with the TOCs (as CHLORAD substrates) and as well as CHLORAD proteins Cdc48, Sp1 and Sp2. To examine what functional role Npl4 and Ufd1 proteins play in CHLORAD, I sought to examine mutants in these genes. If they function upstream of Sp1 and Sp2, and prevent the removal of TOC proteins, we might expect to see a similar physiologically beneficial role with these mutants. To that end, I began basic characterisation looking at measures related to chloroplast biology and examined the impact of stress and developmental transitions on these mutant lines.

4.4 Results and Discussion

4.4.1 Mutant Lines and Crosses

It is important to begin with a description of the lines used in this analysis and how they create limitations within the experiments that have been performed. During the conception of the project, my aim was to perform experimentation using a series of: single knock-out (KO) and overexpression (OX) lines; double and triple mutants to counter isoform redundancy; and crossed lines, to TOC and CHLORAD mutants. One of the strongest sources of evidence previously for the involvement of Sp1, Sp2 and Cdc48 in CHLORAD has come from genetic interactions (Ling *et al.*, 2019; Ling *et al.*, 2012; Ling & Jarvis, 2015b). The Toc33 knock-out mutant, *ppi1*, possesses a chlorotic phenotype, exhibits defects in chloroplast biogenesis and ultrastructure, and reduced abundance of TOC proteins (Jarvis *et al.*, 1998). Through genetic interactions, CHLORAD mutants are able to supress the

phenotype of *ppi* (Ling *et al.*, 2019; Ling *et al.*, 2012). This results in a recovery or greening phenotype in the double mutant plant. For example, both *sp1 ppi1* and *sp2 ppi1* are able to suppress the *ppi1* phenotype — this results in improvements in chloroplast biogenesis, increased chlorophyll content, and an increase in the abundance of TOC proteins (Ling *et al.*, 2019). This is because lack of a negative regulator of TOC protein abundance allows for higher accumulation of proteins such as Toc75 and Toc159, which seemingly allow for greater protein import and a restoration of the phenotype to near WT levels. Crosses of *ppi1* to *npl4-1*, for example, might have revealed whether a similar upstream effect of inhibited protein turnover was visible.

It is with great regret, that I must report that these lines were not produced in time to be included in the experimental analysis in this chapter. I had great difficulties with crossing (dead flower heads, sterile seed, WT progeny) and repeated attempts failed throughout this project. Near the end of this project, I was able to complete some initial crossing with the aid of further demonstration. The produced lines are, at the time of writing, in the F2 generation and require screening of the seed to select the desired segregated mutants. These crosses took the form of single mutants in *Npl4* and *Ufd1* genes to *ppi1-1* and *tic40-4* (as a control), generating the lines: *npl4-1 ppi1-1, npl4l-4 ppi1-1, ufd1C-2 ppi1-1, ufd1D-2 ppi1-1, npl4-1 tic40-4, npl4l-4 tic40-4, ufd1C-2 tic40-4, ufd1D-2 tic40-4.* In these cases, both *ppi1-1* and *tic40-4* were the mother plant: successful crossing could therefore be observed by green progeny in the heterozygous F1 generation, as both *ppi1-1* and *tic40-4* have chlorotic phenotypes. In addition, double mutants to examine redundancy are also at the F2 generation, these consist of *npl4-1 npl4L-4,* and *ufd1C-2 ufd1D-2.* It is frustrating that the double mutant lines could not be included in this analysis, as they have been shown to be necessary to observe functional phenotypes due to redundancy (Li *et al.*, 2022; Mérai *et al.*, 2014).

The knock-out (KO) lines in this chapter are all T-DNA lines acquired from the NASC (Nottingham Arabidopsis Stock Centre). Maps of the genetic loci and T-DNA insertion sites and their respective codes are shown in Figure 4.1. Also displayed on the diagram are a number of T-DNA lines which were not included in this analysis. This is primarily because they are either not relevant T-DNA insertion mutants (such as *npl4L-3* in the 3' untranslated region), there were issues with seed viability, or the plants failed to show an insertion site during genotyping PCR (e.g. *npl4-2*). Nonetheless, multiple knock out lines are ideal, in order to validate any observed phenotype; for this reason and given the limited options with T-DNA lines, I designed guide RNAs (gRNAs) and planned to use CRISPR (clustered regularly interspaced palindromic repeats)-Cas9 mediated gene editing to generate KO lines. Two gRNAs were designed for each gene around the 5' section (Fig. 4.1), ideally excising a segment of the gene during repair of the double strand break. WT (wild type, Col-0 *Arabidopsis thaliana*), *ppi1-1* and *tic40-4* were all transformed by floral dipping with a CRISPR-Cas9 binary vector system (Tsutsui & Higashiyama, 2017). Unfortunately, screening of all seed yielded less than 20 antibiotic resistant plants in total. Genotyping of these plants, first by PCR, then by sequencing the PCR product, revealed that no mutations had taken place. I decided instead to concentrate my analysis on the existing lines in physiological experiments.

The knock-out line *npl4L-1* (SAIL_702_D11) (Fig 4.1) was featured in experiments performed by Li *et al*. (2022) – where they label it as *npl4b*. However, I did not choose this line for my analysis; this was due to the position of the insertion site, which sits at approximately 300 bp upstream of the first exon (Fig 4.1) and is likely in the promoter region. Due to this, I was concerned it was not a true knock-out line. This concern was vindicated by Li *et al*. (2022), who discovered this line was only a knock-down by sqRT-PCR. Coupled with the redundancy concern between Npl4 and Npl4L, it is likely that *npl4L-1* would not have been an informative or relevant single mutant for this analysis.

Also featured in Li *et al*. (2022), was the line SALK_015581, which they label as *ufd1b*. This is a knock-out line in the gene At2g21270, labelled in this study as *Ufd1A*, for which no knock-out line was used in the experiments described in this chapter. Unfortunately, this line was not identified during the earlier phase of this project. This was partially due to an oversight, as well as an entry error in NASC. When searching for SALK_015581 (identified on SIGnAL) on the NASC database, an entry pertaining to locus At5g48330 (RUG2) is produced. To correctly find the insertion line relating to At2g21270, the correct code is SALK_015581C. Unfortunately, by the time I identified this issue in the publication from Li *et al*. (2022), it was too late to double check this entry and order the line for experimental analysis in this project. Also unfortunate, is the fact that there are no other T-DNA insertion lines positioned internally of At2g21270. The line SALK_015677 (Fig 4.1) was acquired, but all plants screened by PCR were found to be WT. As the sequence read from the insertion also mapped to a region just upstream of the first exon, this line was abandoned. It was hoped that a knock-out could be acquired for *Ufd1A* via CRISPR-Cas9 gene editing in place of a T-DNA line.

Overexpressing lines (OX) were generated by *Agrobacterium* mediated floral dipping. These include Npl4 and Ufd1 lines with C-terminal YFP fusions and C-terminal TAP (Table 4.1). In both constructs the gene of interest is under the constitutive overexpression promoter CamV35S. A complete series of Npl4 and Ufd1 lines with CTAP or YFP fusions was not generated; the missing were not identified by screening and these are an Npl4-YFP line and a Ufd1C-CTAP line (Table 4.1). Most genotypes do have multiple OX lines therefore, which may aid in determining the reproducibility in any observed phenotype. In the following experiments in this chapter all lines were homozygous, screened by western blotting (at least) for expression of the transgene; the YFP lines were at the T4 generation, whereas the TAP lines were at the T3 generation.

In the subsequent sections of this chapter, these twelve homozygous KO and OX lines are analysed (Table 4.1), in comparison to WT *Arabidopsis* plants. It was observed that none of these single mutant lines show any visible phenotype during development (Fig. 4.2). Redundancy between homologues might be the cause for this observation, with UPS activities sustained through compensation.

Gene	**Knock-Out Lines**	**Overexpression Lines**
NPL4 (At2g47970)	*npl4-1*	Npl4-CTAP
NPL4L (At3g63000)	*npl4L-4*	Npl4L-YFP Npl4L-CTAP
UFD1A (At2g21270)	-	Ufd1A-YFP Ufd1A-CTAP
UFD1C (At4g15420)	*ufd1C-2*	Ufd1C-YFP
UFD1D (At4g38930)	*ufd1D-2*	Ufd1D-YFP Ufd1D-CTAP

Table 4.1 A table displaying the 12 mutant lines used in experimental analysis in Chapter IV. For each gene, there are a T-DNA insertion line and/or an overexpressing line/s. T-DNA insertion lines were knock-out lines and sourced from NASC – plants were screened for the presence of T-DNA at the gene of interest insertion site and only homozygous lines were used in experiments. Overexpression lines place the gene of interest under the control of the constitutive CamV35S promoter; these transgenic lines were propagated until at least the T3 generation and screened for homozygosity. All overexpression lines created translational fusions at the C-terminus to either a YFP or TAP tag.

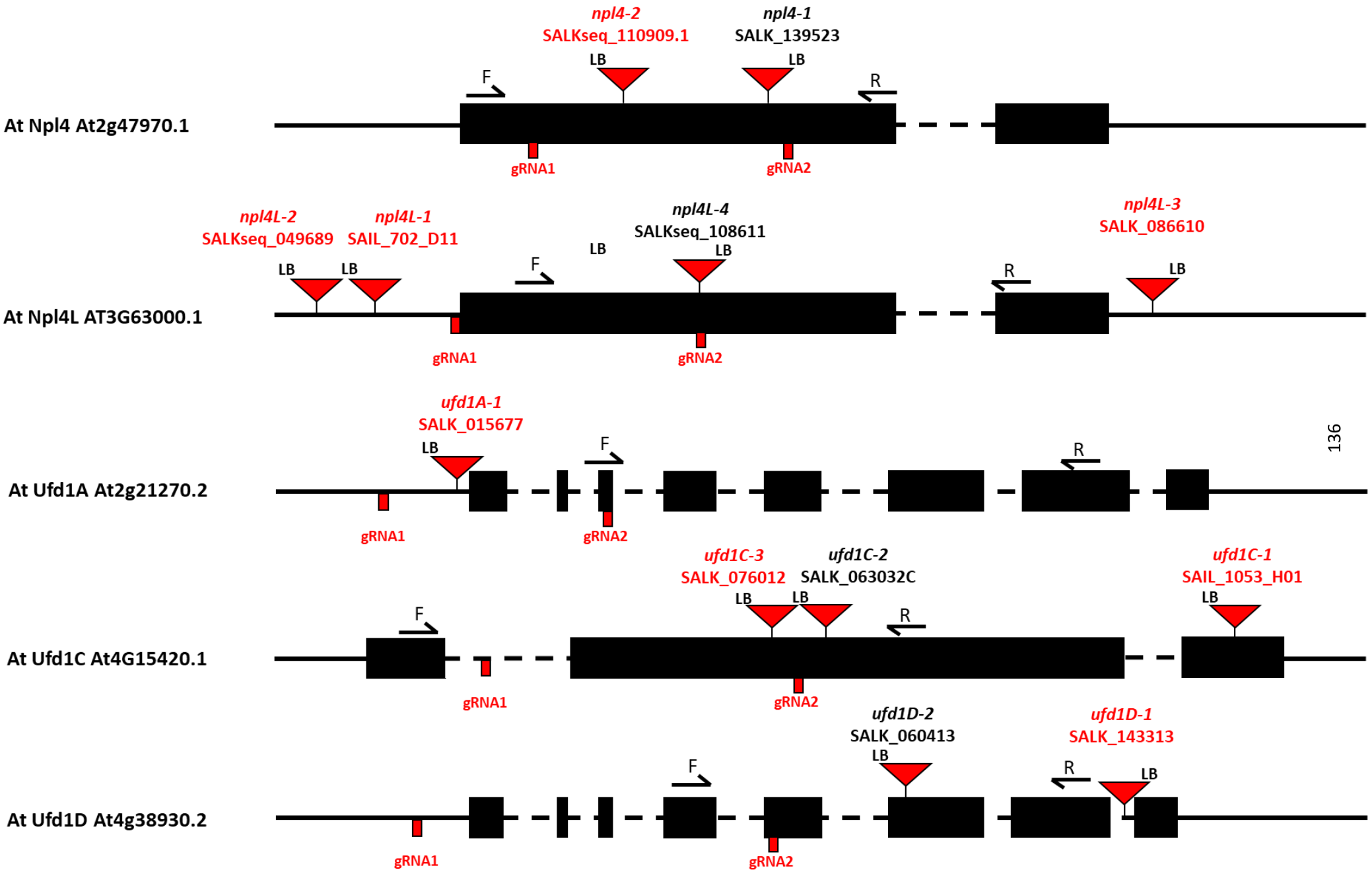

npl4-2
SALKseq_110909.1
npl4-1
SALK_139523
LB
LB
F
R
At Npl4 At2g47970.1
gRNA1
gRNA2
npl4L-2
SALKseq_049689
npl4L-1
SAIL_702_D11
npl4L-4
SALKseq_108611
npl4L-3
SALK_086610
LB
LB
LB
LB
F
R
LB
At Npl4L AT3G63000.1
gRNA1
gRNA2
ufd1A-1
SALK_015677
LB
F
R
136
At Ufd1A At2g21270.2
gRNA1
gRNA2
ufd1C-3
SALK_076012
ufd1C-2
SALK_063032C
ufd1C-1
SAIL_1053_H01
LB
LB
LB
F
R
At Ufd1C At4G15420.1
gRNA1
gRNA2
ufd1D-2
SALK_060413
ufd1D-1
SALK_143313
LB
F
R
LB
At Ufd1D At4g38930.2
gRNA1
gRNA2

Figure 4.1 Diagram showing the genomic loci of Npl4 and Ufd1 genes in *Arabidopsis thaliana*. Protein coding exons for each gene model are shown with black boxes, with the 5' to 3' orientation being from left to right: the start and stop codons are found at the start of the left-most box and end of the right-most box respectively for each gene. Intron sequences are shown by dashed lines. The position of T-DNA insertions in acquired lines are shown by red triangles; the position of the left border (LB) annotation highlights the orientation of the insertion: left side meaning a reverse read insertion site and *vice versa*. The name given to each T-DNA insertion line is displayed above the triangle, as well as the identifying code from SIGnAL. The approximate binding site of primers used for sqRT-PCR is displayed by symbols annotated with F (forward primer) and R (reverse primer), producing a roughly 700 bp amplicon from mRNA. The position of complementary gRNA sequences used to generate CRISPR-Cas9 lines are shown by red boxes, with two designed for each gene using CRISPR-P v2.0. The identity of lines used in experimental analysis is marked by black text (lines with red text were not pursued), with only 4 total lines across the 5 genes. The length of each line is scaled to 2,500 bp.

WT

npl4-1

npl4L-4

ufd1C-2

ufd1D-2

Npl4L-YFP

Ufd1A-YFP

Ufd1C-YFP

Ufd1D-YFP

Npl4-CTAP

Npl4L-CTAP

Ufd1A-CTAP

Ufd1D-CTAP

Figure 4.2 Representative images showing the phenotypes of *Arabidopsis thaliana* Npl4 and Ufd1 mutant lines, and WT. Plants were germinated under standard long day conditions on soil and photographed at 14 days old.

4.4.2 Gene Expression

To examine the relevancy of these lines in subsequent experiments, I first investigated gene expression to demonstrate that they are relevant KO and OX lines. RNA was isolated from 10-day old seedlings, and reverse transcribed to cDNA. Gene specific and a control consisting of Actin2 (ACT2) primer set was then used to perform semi-quantitative PCR. Actin 2 is a well-established control gene, to examine gene expression, as it's expression is consistent and unimpacted by tissue type, development, or external signals (Bustin, 2002). As expression was consistently even, it was not viewed as necessary to seek additional controls, though they may have been beneficial. Agarose gel electrophoresis was used to resolve DNA bands (Fig. 4.3). Bands were then quantified using ImageJ. Included in this analysis are additional CTAP OX lines for Npl4 and Ufd1D, which will be of relevance in section 4.4.4.

Following quantification of band intensities from sqRT-PCR, gene expression was statistically investigated on the normalised data (to ACT2) (Fig. 4.4). Gene expression data was normally distributed for each genotype, according to Shapiro Wilk tests ($p > 0.05$), and possessed homogenous variance according to Levene's test ($F = 1.540$, df1 = 19, df2 = 40, $p = 0.123$). Parametric analysis was thus pursued using a one-way ANOVA test. There was a statistically significant difference between expression levels of the various genotypes ($F = 77.916$, df1 = 19, df2 = 40, $p < 0.001$). A Tukey post-hoc test was run to examine between which specific genotypes there existed significant differences; of most interest were potential differences with the WT control. Starting with the Npl4 primer set, all mutants demonstrated significant differences in gene expression ($p < 0.002$). There were also significant differences between *npl4-1* ($p < 0.001$) and the three Npl4 TAP lines, but not between Npl4-

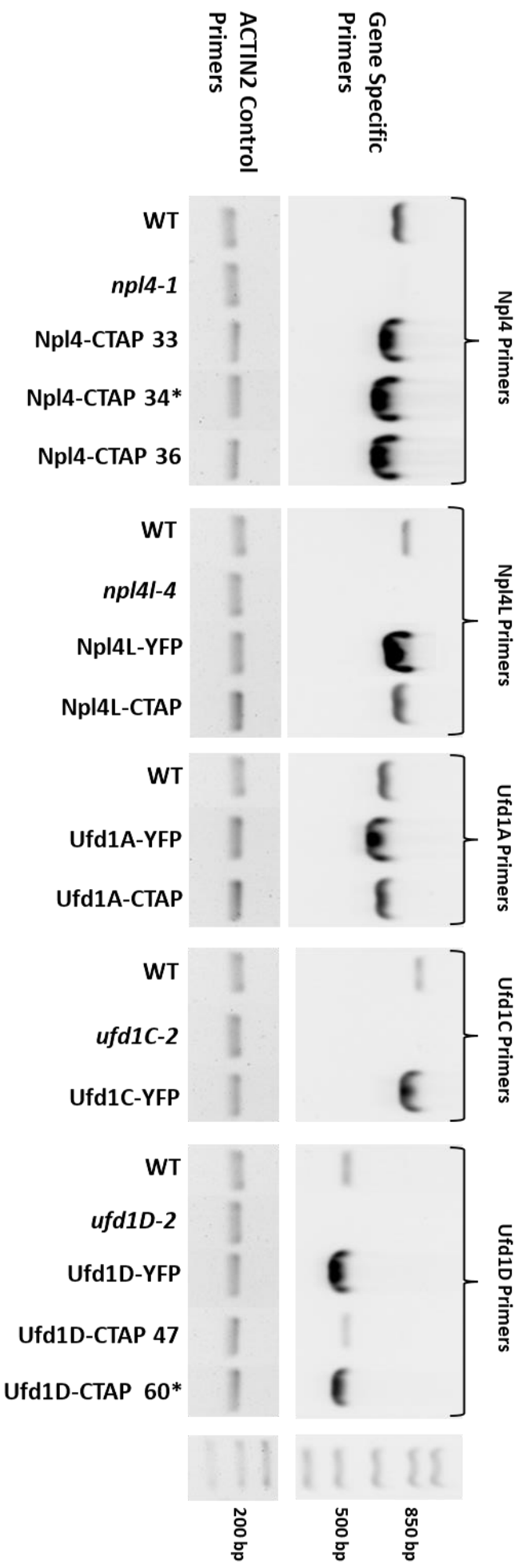
CTAP 34 and Npl4-CTAP 36. Surprisingly, there was no statistically significant difference between WT and *npl4L-4* gene expression ($p = 0.997$), despite expression in the knock-out being essentially 0. This was also observed for the other knock-out mutants *ufd1C-2* ($p = 1.000$), and *ufd1D-2* ($p = 1.000$), when compared to WT expression levels – indeed all knockout mutants aside from *npl4-1* showed the same significance relationship to the OX lines as WT. Npl4L-YFP was the sole Npl4L overexpressing line that was significantly different from WT ($p < 0.001$), as Npl4L-CTAP did not ($p = 0.089$); these lines were also significantly different to each other ($p < 0.001$). This pattern was also true for the Ufd1A OX lines, with only Ufd1A-YFP have significant differences to WT ($p < 0.001$), whereas Ufd1A-CTAP did not ($p = 0.999$). Expression of Ufd1C in Ufd1C-YFP was significantly different to WT ($p < 0.001$). Only Ufd1D-YFP and Ufd1D-CTAP 60 showed significant differences to WT ($p < 0.001$); Ufd1D-CTAP 47, which was not used for subsequent experiments, did not have significant differences to WT expression ($p = 1.000$), and the expression was approximately half that of WT.

Apparent from relative expression graphs is the fact that only several lines may represent valid overexpression mutants by sqRT-PCR (Fig. 4.4). Namely, these would be Npl4L-YFP, Ufd1C-YFP and Ufd1D-YFP. In general, sqRT-PCR may underrepresent expression levels due to lower sensitivity when compared to a method such as quantitative PCR. While the cycle number was first optimised, and kept constant, it may still be possible that samples with highly abundant cDNAs reach a saturation point at a lower cycle number. In general, CTAP lines had lower expression than YFP lines. Both protein structures should be relatively inert, but it can be common for modifications such as these to interfere with protein function. However, this is usually a result of the tag interfering with the activity of protein, due to proximity to the terminus it is placed in fusion with. As both tags are on the C-terminus this cannot account for the difference. It would also be surprising if this interference was consistent between Npl4 and Ufd1 proteins. It may simply be random chance, as these are

just a handful of comparisons; I might simply not have identified TAP lines with higher expression during screening by random chance.

Despite the lack of statistically significant differences, the level of expression from the KO lines should be practically of value. Expression from the KO lines was essentially equal to zero, the lack of significant difference may be viewed as that of the WT expression being also equal to zero therefore in the comparisons made. It may also be a reflection of the adjusted significant values used in the post-hoc test, as these values change based on the number of comparisons being made to correct for type I error (false positives).

Figure 4.3 (overleaf) DNA bands separated by agarose gel electrophoresis of semi-quantitative reverse transcription polymerase chain reaction. 1kb+ DNA ladder is marked on the right-hand side with relevant band sizes. Gene specific primers were used to amplify DNA from relevant genetic backgrounds. In each series, WT is followed by KO then OX lines. On the lower panel, a control primer pair (ACTIN2) was used on the samples to show equal loading of cDNA. Several OX CTAP lines are shown for *Npl4* and *Ufd1D* genes, the ones used in further experiments are marked with an asterisk.


ACTIN2 Control Primers
Gene Specific Primers
WT
npl4-1
Npl4-CTAP 33
Npl4-CTAP 34*
Npl4-CTAP 36
Npl4 Primers
WT
npl4l-4
Npl4L-YFP
Npl4L-CTAP
Npl4L Primers
WT
Ufd1A-YFP
Ufd1A-CTAP
Ufd1A Primers
WT
ufd1C-2
Ufd1C-YFP
Ufd1C Primers
WT
ufd1D-2
Ufd1D-YFP
Ufd1D-CTAP 47
Ufd1D-CTAP 60*
Ufd1D Primers
200 bp
500 bp
850 bp

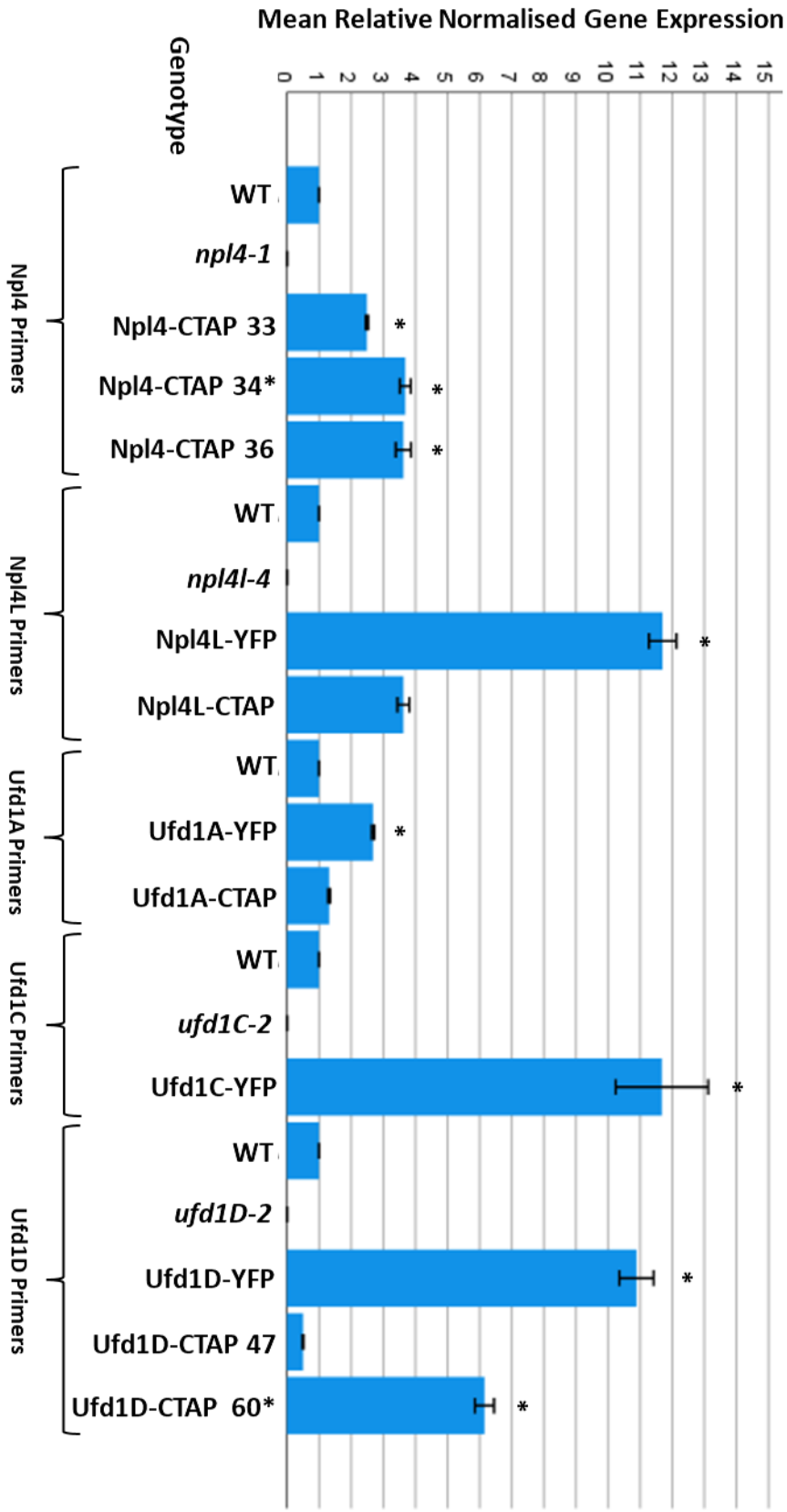

Mean Relative Normalised Gene Expression
Genotype
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Npl4 Primers
WT
npl4-1
Npl4-CTAP 33
Npl4-CTAP 34*
Npl4-CTAP 36
Npl4L Primers
WT
npl4l-4
Npl4L-YFP
Npl4L-CTAP
Ufd1A Primers
WT
Ufd1A-YFP
Ufd1A-CTAP
Ufd1C Primers
WT
ufd1C-2
Ufd1C-YFP
Ufd1D Primers
WT
ufd1D-2
Ufd1D-YFP
Ufd1D-CTAP 47
Ufd1D-CTAP 60*
*

Figure 4.4 Simple bar chart of quantified band intensity from semi-quantitative reverse transcription polymerase chain reaction. Gene specific bands are normalised to ACTIN2 primer products, and divided by the WT intensity to produce relative ratios of expression. CTAP genotype names labelled with an asterisk were used in later experiments. Bars labelled with an asterisk show genotypes with significant (α = 0.05) differences to WT. Error bars are ± 1 Standard Error of the mean.

4.4.3 De-Etiolation

Etioplasts are chloroplast precursors that form in the darkness and are rapidly converted into chloroplasts upon exposure to light. CHLORAD has a function in the fine tuning of this process, as *sp1* mutants showed stunted development and viability following transfer of seedlings from the dark to light (Ling *et al.*, 2012). Mutant lines were either germinated in the dark and allowed to grow for 5 days (Fig 4.5), or after this period they were then subject to continuous light treatment for 24 hours to rapidly induce de-etiolation (Fig. 4.6). The effect on the protein import apparatus was surveyed via western blotting (Fig. 4.7). Proteins were harvested from tissue immediately following photography. A standard SDS-PAGE and Western blotting protocol was used to separate and transfer proteins to nitrocellulose membranes. Membranes were then probed with native antibodies against Toc159, Toc75, and Toc33 (as CHLORAD substrates) and controls of Tic40 and H3 (Fig. 4.7). Bands were then quantified using ImageJ and statistically analysed with SPSS.

Non-parametric statistical analysis was pursued as the data was not normally distributed. Kruskal Wallis tests were run for samples to compare distributions across genotypes. Within the dark treated samples, there were no significant differences across the groups for the abundance of both Toc33 (H = 20.785, df = 12, p = 0.054). and Tic40 (H = 15.385, df = 12, p = 0.221) proteins. Significant differences (α = 0.05) were found within Toc159 (H = 30.600,

df = 12, p = 0.002) and Toc75 (H = 27.949, df = 12, p = 0.006) abundance. Dunn-Bonferroni post hoc tests were then run to examine the specific differences between genotypes. The Bonferroni adjusted significant results found no significant difference between genotypes.

I then analysed the samples treated with one day of light to induce de-etiolation. Here, Kruskal Wallis H tests showed that there were no significant differences across the groups for the abundance of Toc75 (H = 18.523, df = 12, p = 0.101), Toc33 (H = 20.687, df = 12, p = 0.055). and Tic40 (H = 10.595, df = 12, p = 0.564) proteins. Significant differences were once again found for Toc159 abundance (H = 23.923, df = 12, p = 0.021). As before, however, the Bonferroni adjusted significant results found no significant differences with WT.

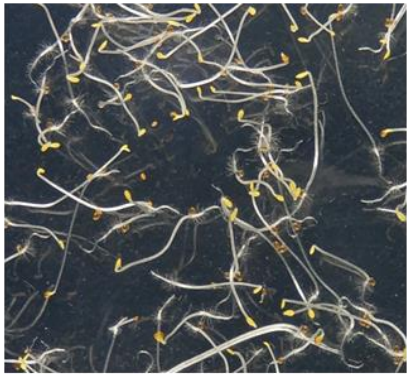
There were few visual differences between lines for the plants grown under exclusively dark conditions (Fig. 4.5). There may be more ungerminated seeds in the KO lines, but this wasn't addressed statistically. Almost all plants show typical signs of etiolation: elongated hypocotyls and undeveloped cotyledons. Following exposure to light, de-etiolation is readily apparent (Fig. 4.6), as cotyledons begin to emerge in the majority of plants. Yet again, there were few differences by eye between genotypes, following continuous light treat.

The de-etiolation treatment can be clearly observed having an effect: TOC protein abundance is elevated after exposure to light, especially for Toc33 (Fig. 4.7). However, there were no observed significant differences in response to this treatment across the mutant genotypes. It would therefore appear that individual KO and OX lines do have not a significant role in CHLORAD with regards to de-etiolation.

Figure 4.5 (overleaf) An assessment of de-eiotlation I: *Arabidopsis thaliana* Npl4 and Ufd1 mutant lines, and WT were germinated in the dark for 5 days. Representative images are shown, captured immediately before samples were harvested for protein extracts.

5 Days Dark

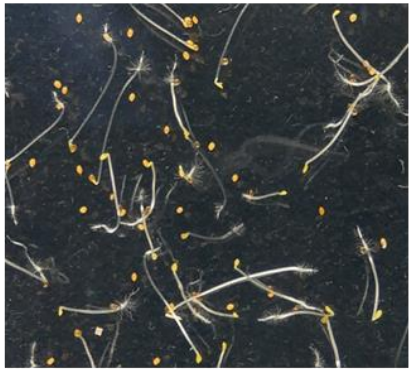
WT



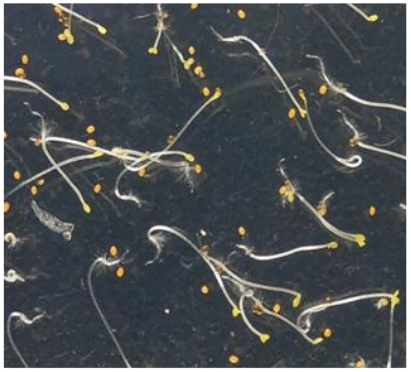
npl4-1

npl4L-4

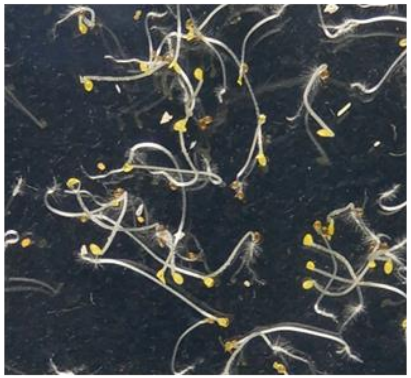
ufd1C-2



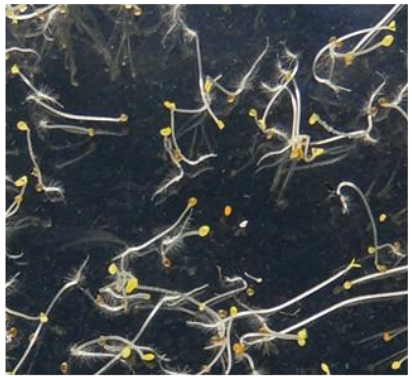
ufd1D-2



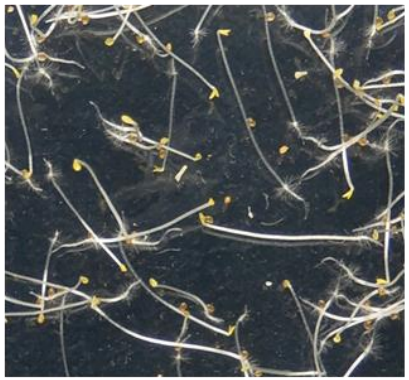
Npl4L-YFP



Ufd1A-YFP



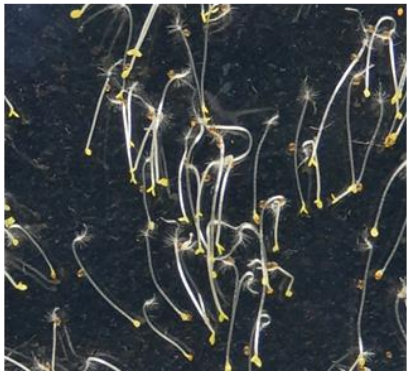
Ufd1C-YFP



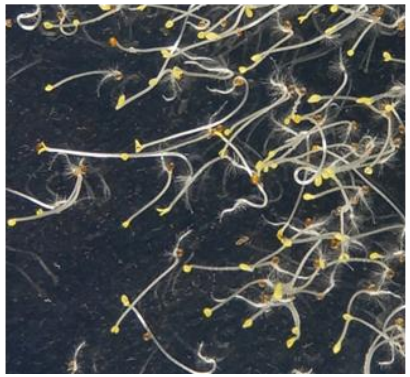
Ufd1D-YFP

Npl4-CTAP

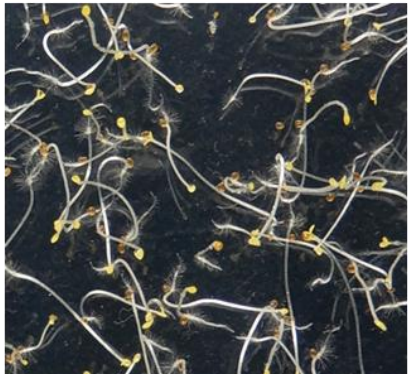
Npl4L-CTAP



Ufd1A-CTAP

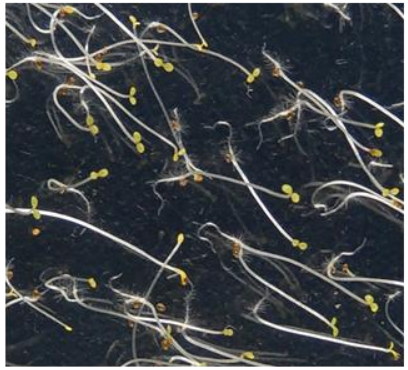


Ufd1D-CTAP

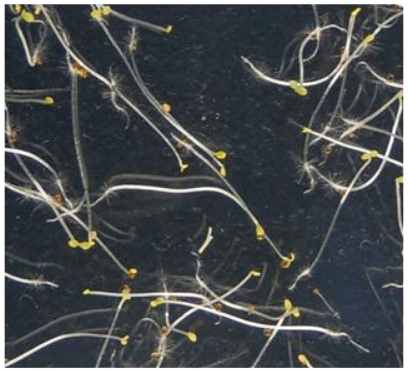


5 Days Dark, with One Day Light

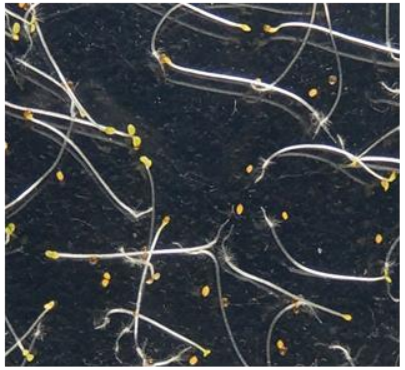
WT



npl4-1

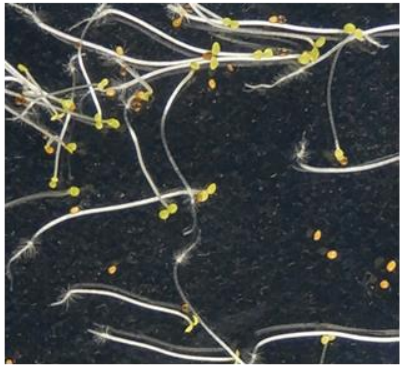


npl4L-4

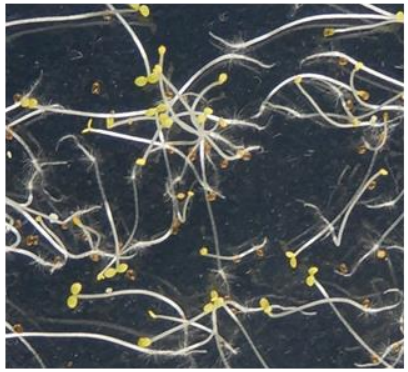


ufd1C-2

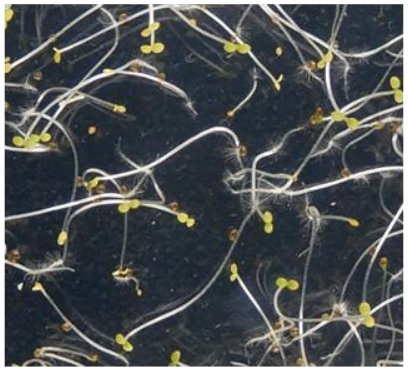
ufd1D-2



Npl4L-YFP

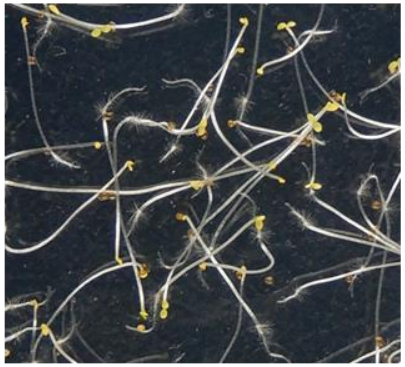


Ufd1A-YFP

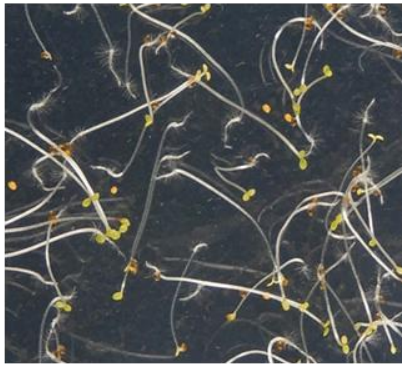


Ufd1C-YFP

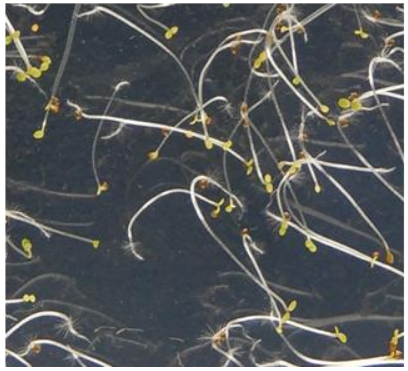
Ufd1D-YFP



Npl4-CTAP



Npl4L-CTAP



Ufd1A-CTAP

Ufd1D-CTAP

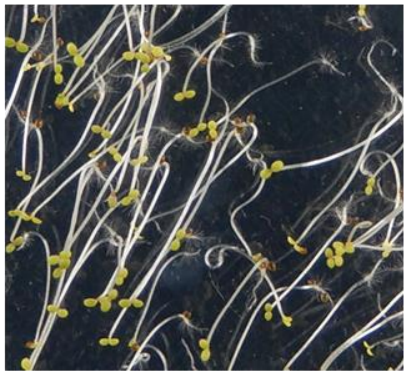


Figure 4.6 An assessment of de-eiotlation II: *Arabidopsis thaliana* Npl4 and Ufd1 mutant lines, and WT were germinated in the dark for 5 days, then exposed to continuous light for 24 hours. Representative images are shown, captured immediately before samples were harvested for protein extracts.

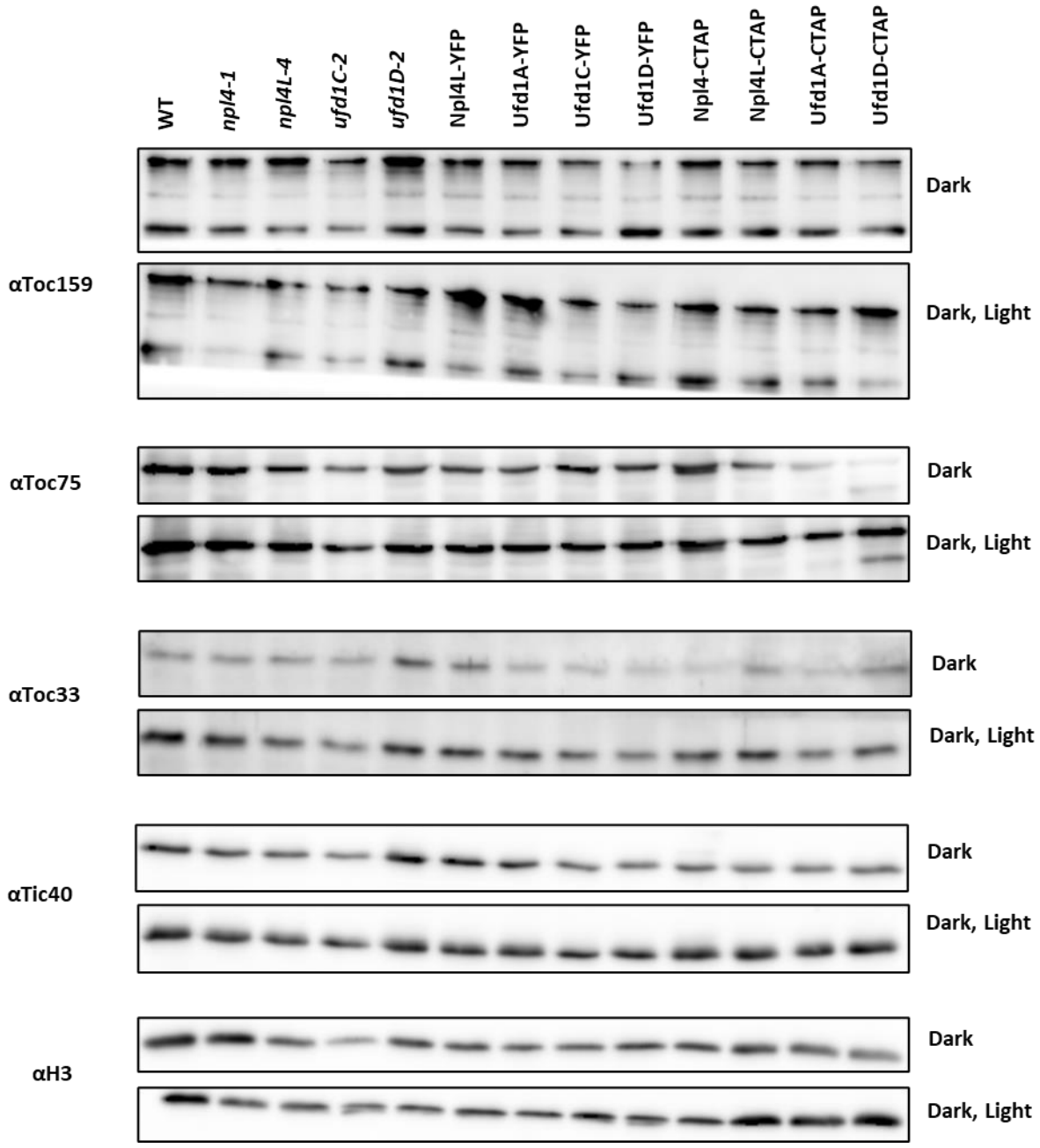
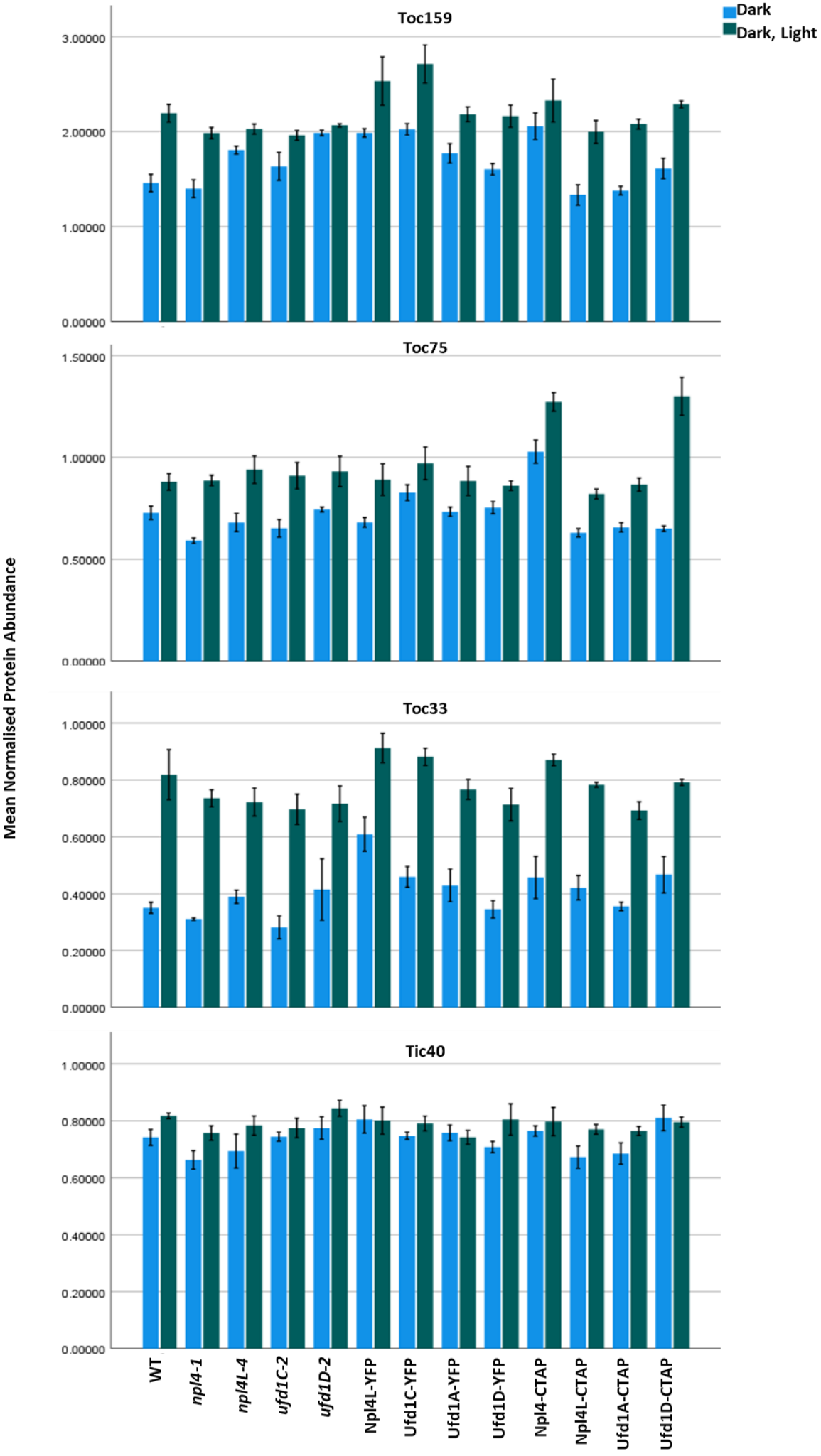


Figure 4.7 Western blots of protein extracts from de-etiolated seedlings. Seedlings were exposed either to 5 days of dark treatment, or 5 days of dark treatment and one day of continuous light treatment. Genotypes for each line are shown above. TOC proteins were blotted against the membranes as representative CHLORAD substrates. Tic40 was blotted as a control.

Figure 4.8 (overleaf) Multiple bar graphs showing quantified protein abundance from the de-etiolation experiment. Seedlings were exposed either to 5 days of dark treatment, or 5 days of dark treatment and one day of high light – this is shown by the key in the top-right. Genotypes are displayed on the x axis; protein abundance normalized to H3 are shown on the y axis. Error bars are ± 1 Standard Error of the mean.


Dark
Dark, Light
Toc159
Toc75
Toc33
Tic40
Mean Normalised Protein Abundance
WT
npl4-1
npl4L-4
ufd1C-2
ufd1D-2
Npl4L-YFP
Ufd1C-YFP
Ufd1A-YFP
Ufd1D-YFP
Npl4-CTAP
Npl4L-CTAP
Ufd1A-CTAP
Ufd1D-CTAP
Genotype

4.4.4 TOC Abundance and Mannitol Treatment

One of the most consistent phenotypes for CHLORAD mutants, is the ability to impact the abundance of TOC proteins. This was a crucial piece of evidence in demonstrating that Cdc48 has a functional role in CHLORAD. The effect, however, was only observed under stress conditions (osmotic, to be precise) not in untreated plants. Conversely, mutants of Sp1 and Sp2 (OX lines) could impact TOC abundance in untreated conditions; this may be partially explained by the fact that Cdc48 knock-out mutants are embryo lethal, so an inducible dominant negative mutant must be used which could generate a milder phenotype.

To investigate this phenotype, I started by germinated lines on standard MS media. When 10 days old, I transferred plants to either untreated MS (control) or to MS supplemented with 400 mM mannitol, as a source of osmotic stress. Osmotic stress treatment resulted in the accumulation of anthocyanin compounds in most treated plants, regardless of their genotype, as well as some chlorosis (not shown). Following treatment, I harvested tissue and extracted proteins. I then performed SDS-PAGE and western blotting on these samples; membranes were probed with antibodies against Toc159, Toc75, Toc33, Tic40 and H3 (Fig. 4.9). Protein bands were then quantified using ImageJ and statistically analysed using SPSS (Fig. 4.10).

Protein abundance data was normally distributed by Shapiro Wilk tests (df = 3, $p > 0.05$), across protein types and genotypes. Levene's test also accepted the null hypothesis of homogeneity of variance for each measured protein ($p > 0.489$). I therefore proceeded to analyse the data parametrically: I decided to use two one-way ANOVAs, to compare protein abundance in both untreated and mannitol treated samples separately. This was because I am more interested in the effects between genotypes, than I am of the treatment itself.

Beginning with the untreated samples, there were no statistical differences between genotypes with Toc159 ($F = 0.548$, df = 12, $p = 0.863$), Toc33 ($F = 540$, df= 12, $p = 0.540$) and Tic40 ($F = 0.990$, df = 12, $p = 0.863$). However, there were significant differences between genotypes measured with Toc75 abundance ($F = 0.922$, df = 12, $p < 0.001$). For

post-hoc analysis, I used Dunnet's t (2-sided) test, to specifically compare with WT. This showed only one significant difference, with Npl4-CTAP ($p < 0.001$) – despite the graphical difference with Ufd1D-CTAP (Fig.4.10), the difference in abundance was not significant ($p = 0.158$).

Next looking at mannitol treated samples, here there were more significant differences, with Toc159 ($F = 4.315$, df = 12, $p < 0{,}001$), Toc33 ($F = 2.749$, df= 12, $p = 0.015$) and Toc75 ($F = 0.272$, df = 12, $p < 0.001$). Only Tic40 showed no significant differences ($F = 0.272$, df = 12, $p = 0.989$). Dunnet's t (2-sided) test was again used as post hoc test to compare mutant genotypes to WT. For Toc159 and Toc33, no significant differences ($\alpha = 0.05$) were found, meaning that the difference across protein abundance existed between mutant genotypes. This was not the case for Toc75, where both Npl4-CTAP and Ufd1D-CTAP showed significance ($p < 0.001$).

This physiological data provides an interesting strand of evidence about Npl4 and Ufd1 involvement in CHLORAD. There is a clear Toc75 overaccumulation observed with the overexpressing lines Npl4-CTAP and Ufd1D-CTAP. This is quite apparent on the western blots (Fig. 4.9), though with Ufd1D-CTAP it curiously manifests in two bands. In this case, the second band is of a lower molecular weight, which may indicate some degree of proteolysis. It is unclear why a protein that over accumulates shows greater abundances in a truncated or degraded form; it could reflect some form of modification to limit the activity of aberrantly accumulating protein, or a misfolded state. Stress treatment was not required to induce this phenotype in either case, though it was statistically significant with mannitol treatment for Ufd1D-CTAP. What is arguably most interesting, is that the effect is the opposite of what one might expect. If the adapter proteins function to recruit Cdc48, enabling substrate retrotranslocation, one might expect overexpression would reduce the abundance of substrate. This is certainly what is observed with Sp1 OX and Sp2 OX. More curiously also, it appears to not be correlated to expression, at least in the case with Ufd1D, where the YFP line has nearly two times higher expression. To see if this phenotype was a

consequence of the TAP tag, I analysed additional lines of Npl4 and Ufd1 with CTAP tags. I did not observe the phenotype with Ufd1D-CTAP 47, though this may be related to weak expression (Fig. 4.4). Yet, I did observe it with Npl4-CTAP 33 and 36, which show comparable expression to the Npl4-CTAP used in the majority of this chapter. The TAP tag may therefore create a dominant negative phenotype. Why this phenotype is only observed with Toc75 is another curious question. It has been observed that there is typically one TOC protein that is selectively more affected than others in previous analysis: for instance, Toc33 with Cdc48 dominant negative and Toc75 with *sp1 ppi1.* It is unclear why this is or why there is little consistency to this pattern between examined genotypes.

Overall, the osmotic stress can be observed having a noticeable impact on the abundance of TOC proteins generally, with the abundance of Toc159 and Toc33 roughly halving (Fig. 4.10). Nonetheless outside of Npl4-CTAP and Ufd1D-CTAP with Toc75, there are no other major differences. Lack of reciprocation with the KO lines might be explained by isoform redundancy. But given this is not consistent between Ufd1D-YFP and UfdD-CTAP, suggests that it is not an overexpression related issue, but a tag related one. However, even if this is the case the impact still implies a functional link. Taken with direct evidence of physical interaction, it is informative data for a role of Npl4 and Ufd1D in CHLORAD. Further examination with double mutants will no doubt be crucial to investigate this phenotype further.

Figure 4.9 (overleaf) Western blots of protein extracts from untreated (control) and stress treated (mannitol) plants. Seedlings were germinated on MS plates and transferred to either another MS plate or an MS plate containing 300 mM mannitol for a further 3 days. Genotypes for each line are shown above. TOC proteins were blotted against the membranes as representative CHLORAD substrates. Tic40 was blotted as a control.

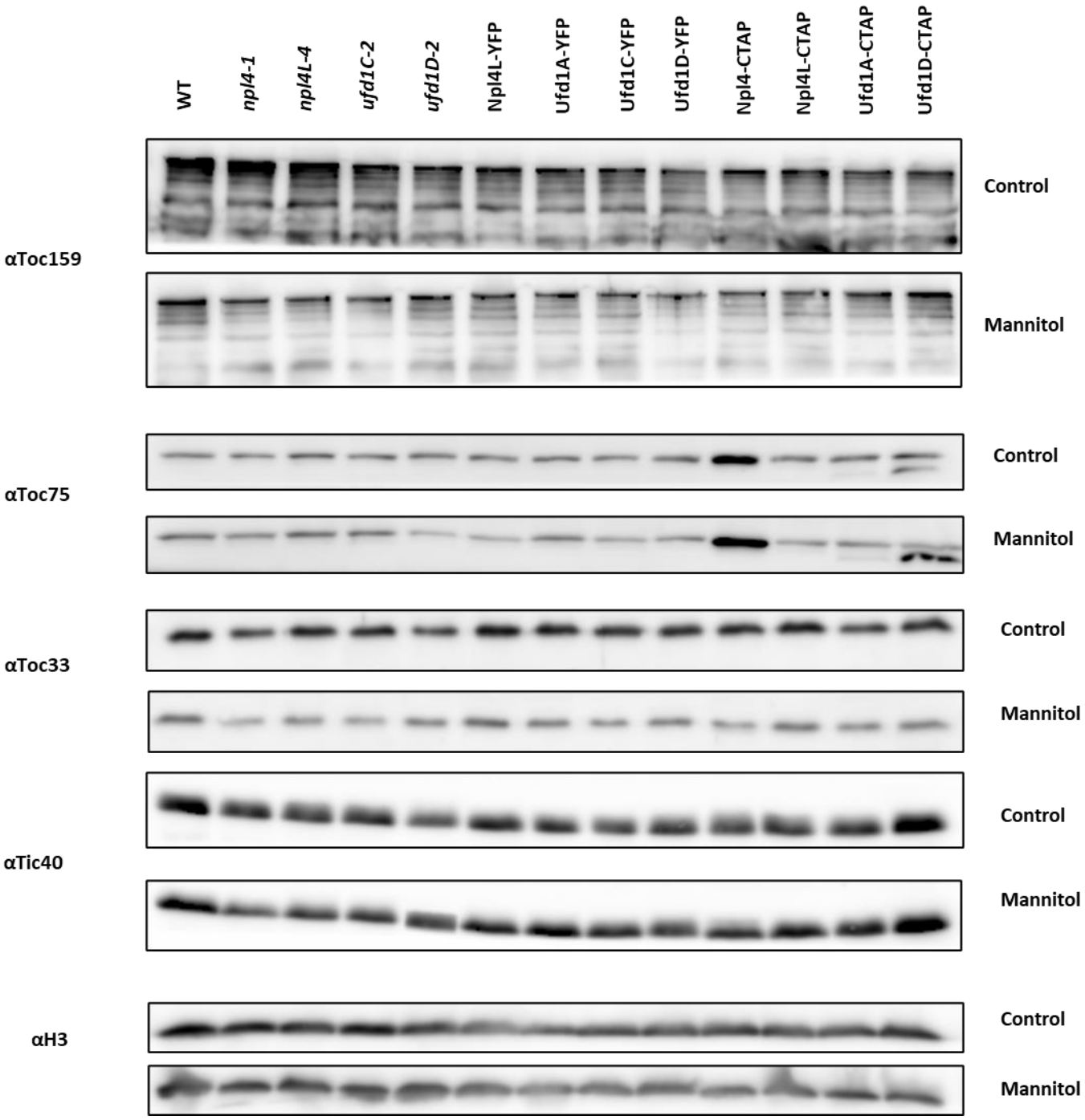
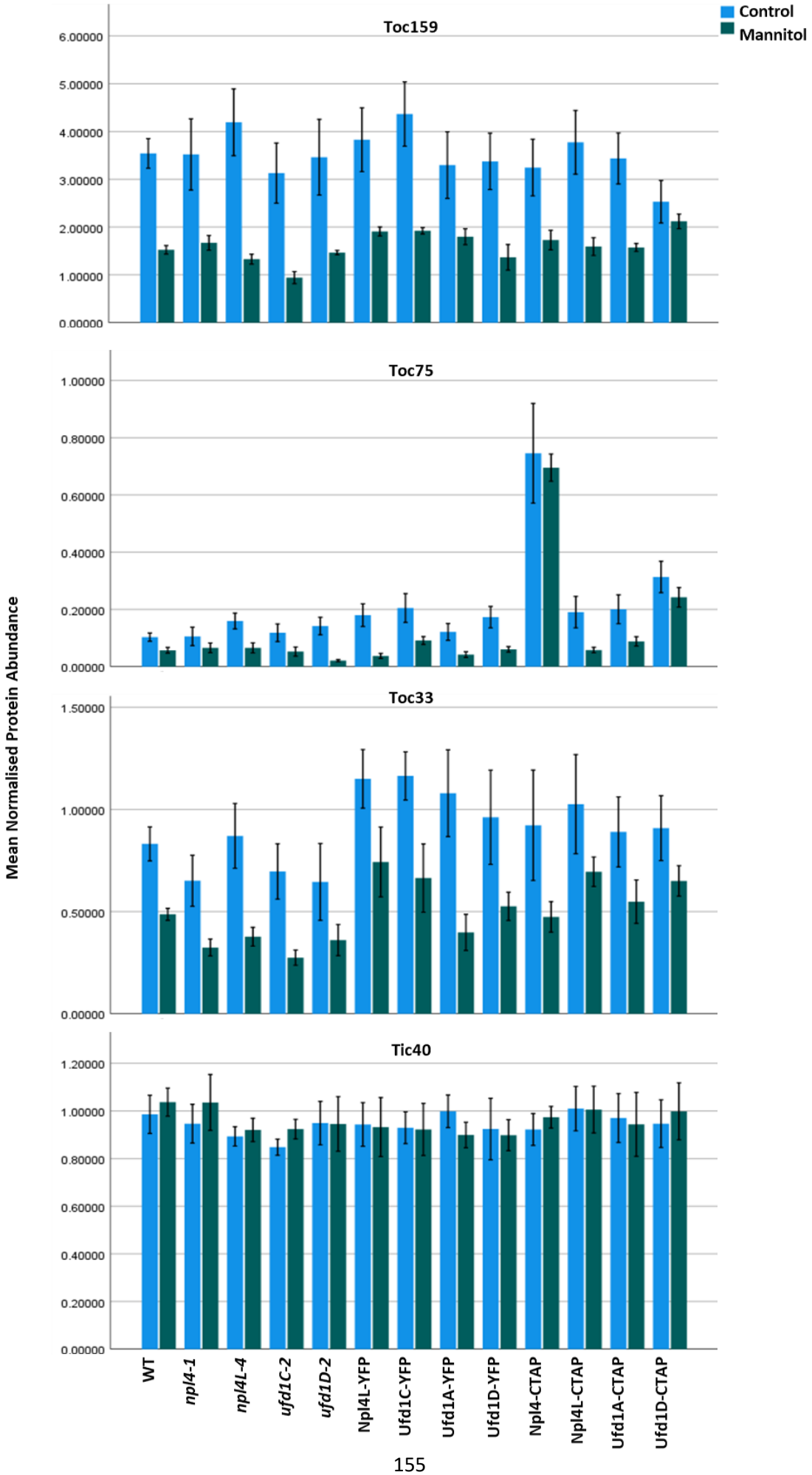


Figure 4.10 (overleaf) Multiple bar graphs showing quantified protein abundance from the from untreated (control) and stress treated (mannitol) plants. Seedlings were germinated on MS plates and transferred to either another MS plate or an MS plate containing 300 mM mannitol for a further 3 days this is shown by the key in the top-right. Genotypes are displayed on the x axis; protein abundance normalized to H3 are shown on the y axis. Error bars are ± 1 Standard Error of the mean.


Control
Mannitol
Toc159
6.00000
5.00000
4.00000
3.00000
2.00000
1.00000
0.00000
Toc75
1.00000
0.80000
0.60000
0.40000
0.20000
0.00000
Toc33
1.50000
1.00000
0.50000
0.00000
Tic40
1.20000
1.00000
0.80000
0.60000
0.40000
0.20000
0.00000
Mean Normalised Protein Abundance
WT
npl4-1
npl4L-4
ufd1C-2
ufd1D-2
Npl4L-YFP
Ufd1C-YFP
Ufd1A-YFP
Ufd1D-YFP
Npl4-CTAP
Npl4L-CTAP
Ufd1A-CTAP
Ufd1D-CTAP
Genotype

4.4.5 Salt Treatment

Plants sown on media containing high salt concentrations were scored by two measures: germination efficiency and development stage. These were used, as previously reported (Ling & Jarvis, 2015b), to examine the ability of plants to overcome the stress associated with high salinity at early stages of development, where CHLORAD aids in overcoming the ROS stress created by highly saline conditions. Plants of the aforementioned mutant lines were sown directly onto MS plates containing 150 mM NaCl and allowed to grow for 14 days (Fig. 4.11). Measurements were recorded by counting two developmental indices: germination efficiency and the proportion of 'developed' plants. Here, a 'developed' plant means one that at least has fully emerged cotyledons which are vibrant (not pale) green.

Counts of germinated plants and of ungerminated seeds were analysed statistically by using the Chi-Square independence test: this found that there is a significant associated between genotype and germination efficiency ($\chi^2 = 179.103$, df = 12, $p < 0.001$). Examining the strength of the association with Phi and Cramer's V test, finds what may be described as a very strong (Akoglu, 2018) association between the variables ($\phi_c = 0.290$, $p < 0.001$). Post-hoc pairwise Z-tests were then used to examine intergroup differences between genotypes based on column proportions, with Bonferroni corrections for p values ($\alpha = 0.05$). There were significant differences between WT and the three single KO lines *npl4L-4*, *ufd1C-2*, and *ufd1D-2* for germinated seeds. In general, significant differences were observed between *npl4L-4*, *ufd1C-2*, and *ufd1D-2* lines and all other genotypes for both germinated seed proportions.

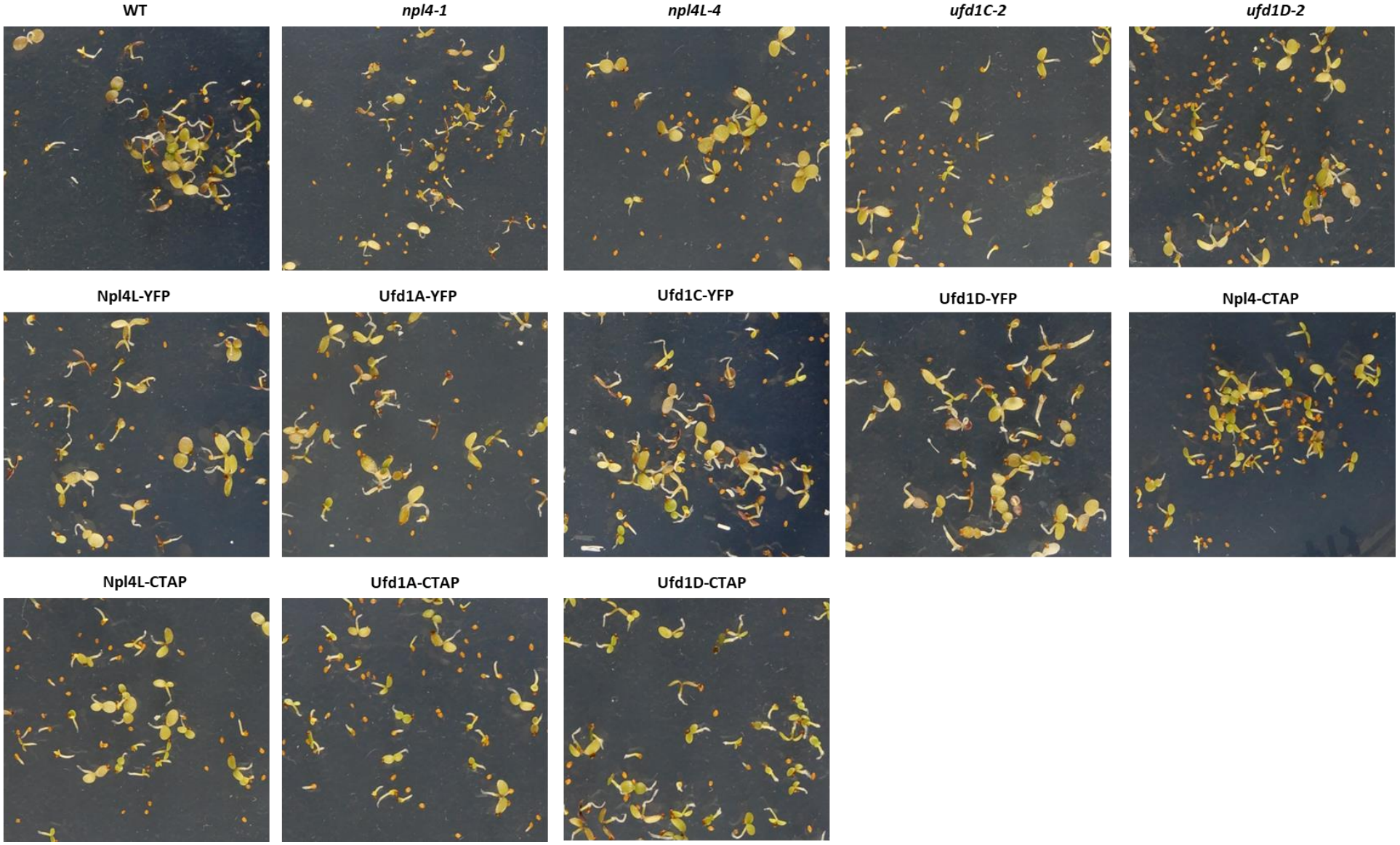
Counts of developed plants, those with expanded cotyledons, were also analysed by the Chi-Square independence test. There was a significant association between plant development and their genotype ($\chi^2 = 40.819$, df = 12, $p < 0.001$). This association was not to the same magnitude as that observed for germination efficiency, but Phi and Cramer's V test still found a strong association (Akoglu, 2018) between genotype and development ($\phi_c =$

0.180, $p < 0.001$). As before, post-hoc Z-tests were examined. These tests found only one developed significantly more frequently than WT, with Ufd1D-CTAP ($p = 0.009$); Ufd1D-CTAP also demonstrated significant differences with knock-out mutants *npl4-1* and *ufd1C-2*.

There are visible differences in germination efficiency with the KO mutants (Fig.4.11). The observed differences may not be due to a function in CHLORAD. I suggest this, as no differences in germination efficiency were observed in the analysis of Sp1 mutants germinated on salt. The decreased efficiency here, may be related to impacted UPS processes in other cellular functions. In scoring development, only Ufd1D-CTAP showed a significant difference, but the magnitude of this was small – especially compared to relative increase observed between WT and Sp1 OX, which was roughly 2-fold. It should be noted therefore that the baseline count I gave to WT was far higher than previously reported (70% vs 35%) (Ling & Jarvis, 2015a). The same strength of salt treatment was applied here, and the experiment was repeated multiple times. It may reflect either differences in counting, or perhaps the quality of the seed used. In either case, it may suggest that the original result examining SP1 in salt tolerance should be re-examined in this specific regard. This is not to dispute the role of Sp1 in stress tolerance generally, as osmotic and paraquat stress tolerance were demonstrated with multiple measures.

Figure 4.11 (overleaf) Salt stress treatment of *Arabidopsis thaliana* Npl4 and Ufd1 mutant lines, and WT. Plants were sown on MS media containing 150 mM NaCl and allowed to grow for 14 days under standard long day conditions. Representative images are shown, captured immediately before samples were harvested for protein extracts.

150 mM NaCl


WT
npl4-1
npl4L-4
ufd1C-2
ufd1D-2
Npl4L-YFP
Ufd1A-YFP
Ufd1C-YFP
Ufd1D-YFP
Npl4-CTAP
Npl4L-CTAP
Ufd1A-CTAP
Ufd1D-CTAP

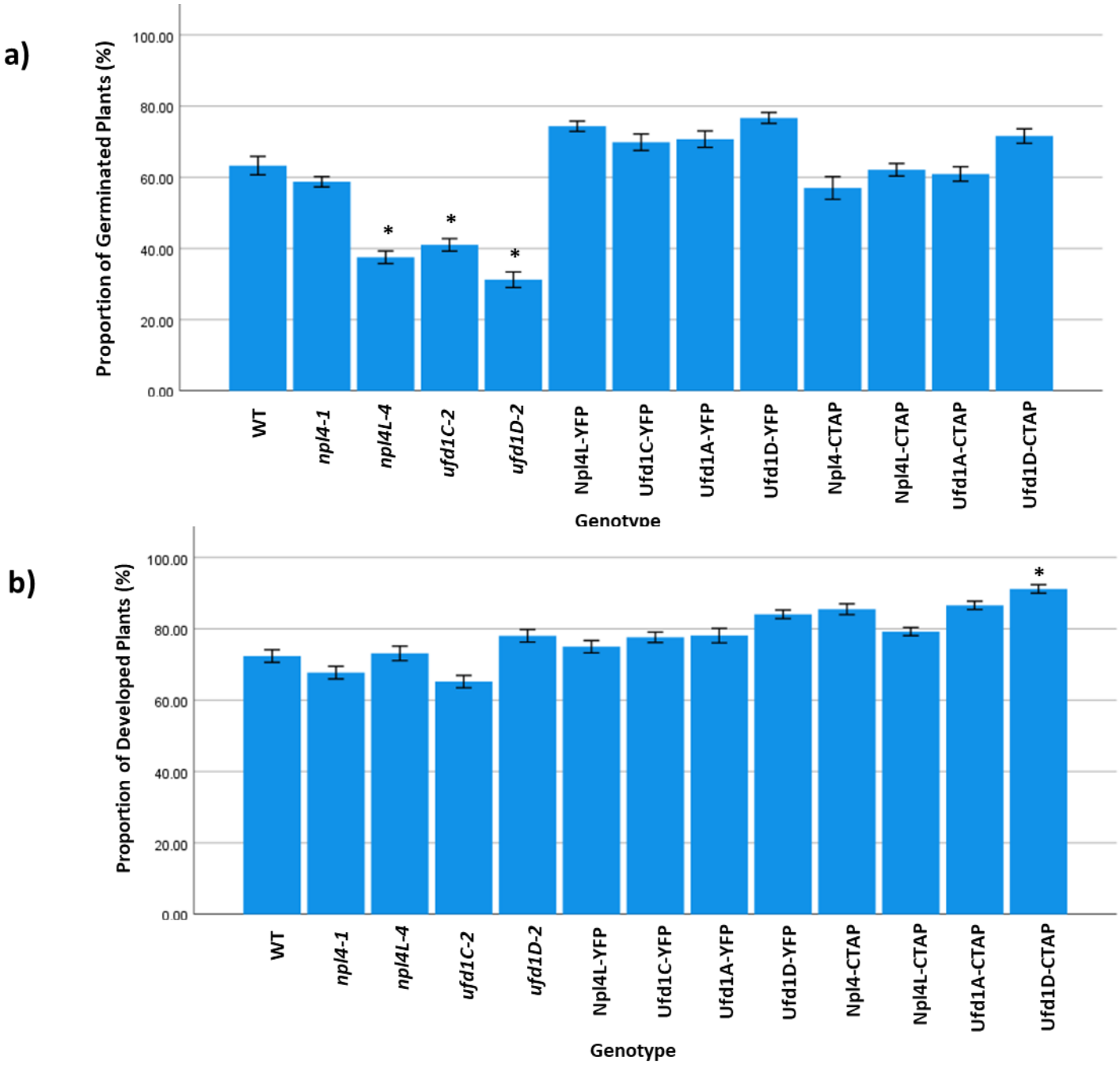


Figure 4.12 Simple bar charts of counted data from salt treated plants. a) The proportion of plants from different mutant lines which germinated on salt containing media. b) The proportion of developed plants (those with expanded cotyledons) from different mutant lines grown on salt containing media. Error bars are ± 1 Standard Error of the mean. Bars labelled with an asterisk show genotypes with significant (α = 0.05) differences to WT.

4.4.6 Chlorophyll Content, Chlorophyll Fluorescence and Senescence Induction

The chloroplast is the home of photosynthesis, that most essential of plant metabolic processes. CHLORAD has a functional role in regulating photosynthesis linked to ROS production from abiotic stress. As a regulator of protein import generally, CHLORAD mutants have reduced quantities of chlorophyll content, indicative of defects in chloroplast biogenesis. Despite the lack of visible differences between genotypes, the chlorophyll content of 28-day old plants was measured (Fig. 4.12a) using a SPAD meter. Plants were grown to this age under standard long day conditions in soil.

Average chlorophyll content was a typical value of approximately 1.5 nmol/g of fresh weight tissue in the WT, indicating that the plants were healthy and growing under normal conditions. I proceeded to analyse this data by first checking its normality. Shapiro Wilk tests (df = 6, $p > 0.05$) showed a normal distribution and Levene' test also accepted the null hypothesis of homogeneity of variance ($p > 0.432$). I therefore examined the data using a one-way ANOVA. This found that there surprisingly was a significant difference between the genotypes ($F = 3.966$, df = 12, $p < 0.001$). Dunnet's t (2-sided) test was again as a post hoc test to examine the differences with WT. Only *npl4-1* was significantly pale ($p = 0.008$) (Fig. 4.12a).

During leaf senescence, chloroplasts are converted into gerontoplasts. This process involves the disassembly of photosynthetic complexes and recycling of nutrients. As with de-etiolation, CHLORAD was shown to have functional significance with this developmental process: Sp1 OX plants demonstrated an accelerated reduction in chloroplast activity and *vice versa*. Chlorophyll fluorescence measurements are a standard measuring technique in assessing various aspects of photosystem II activity (Maxwell & Johnson, 2000). Modern instruments allow for non-invasive interrogation of photosynthesis activity *in vivo*. The classical measure is that of the maximum photochemical efficiency of photosystem II, Fv/Fm.

I treated a selection of leaves from 28-day old plants by wrapping them in foil. This restricted access to light which acts as a key signal to engage senescence. After 5 days, I uncovered them and recorded Fv/Fm values of senescence treated and untreated leaves (a measure of the general photosynthesis efficiency of the mutant background).

The Fv/Fm data (Fig. 4.12b) for untreated and senescence treated plants was initially analysed using the Shapiro-Wilk test, Levene's test of homogeneity of variance, and frequency histograms. This initial examination of the data was used to determine whether it was suitable for parametric statistical comparisons using variations of the ANOVA test. to examine the distribution of the data. Not all dependent variables (7/26) failed to reject the null hypothesis ($p < 0.05$) for the Shaprio Wilk test, meaning the data did not conform to a normal distribution. While Levene's test was satisfied for the senescence treated genotypes ($F = 0.746$, df1 = 12, df2 = 60, $p = 0.702$), the variance between samples was not equal for the untreated genotypes ($F = 2.463$, df1 = 12, df2 = 65, $p = 0.010$). Non-parametric analysis was thus pursued to examine the Fv/Fm data.

Kruskal Wallis H tests showed that there was a significant difference within the untreated samples ($H = 38.248$, df = 12, $p < 0.001$). The Bonferroni adjusted significant results found no significant differences with WT. However, there were significant differences between Ufd1C-YFP and Npl4L-YFP ($p = 0.001$), *npl4L-4* ($p = 0.001$) and *ufd1C-2* ($p < 0.001$).

Kruskal Wallis H tests also showed that there was a significant difference within the senescence induced samples ($H = 29{,}981$, df = 12, $p = 0.003$). The Bonferroni adjusted significant results found significant differences between Ufd1C-YFP and WT ($p = 0.030$) and *ufd1D-2* ($p = 0.05$).

Like other single CHLORAD mutants, there are no general observed differences in chlorophyll content in the majority of lines. There is one detectable significant difference with *npl4-1,* though the magnitude is not very large and there was substantial variability in the

sample, shown by overlapping error bars with WT (Fig. 4.12a). In addition, given the lack of any visible difference I am cautious to conclude anything meaningful, and would recommend this analysis be repeated.

Ufd1C-YFP demonstrated the weakest Fv/Fm data when untreated and under senescence induction – and this was statistically different to WT, and less than half the value. Senescence induction generally was observed to have a large impact on photosynthetic efficiency (Fig. 4.12b). This would at least demonstrate that senescence is genuinely being induced: as the photosynthetic machinery is broken down, so does PSII efficiency decrease. This would then suggest that there are no major impacts of Npl4 and Ufd1 mutants on PSII efficiency, aside from Ufd1C-YFP. It is hard to say why this might be, especially because it does not tally with any other lines of evidence in this Chapter. It might therefore be regarded as anomalous, and need to be disproven through future repeated analysis with additional mutants.

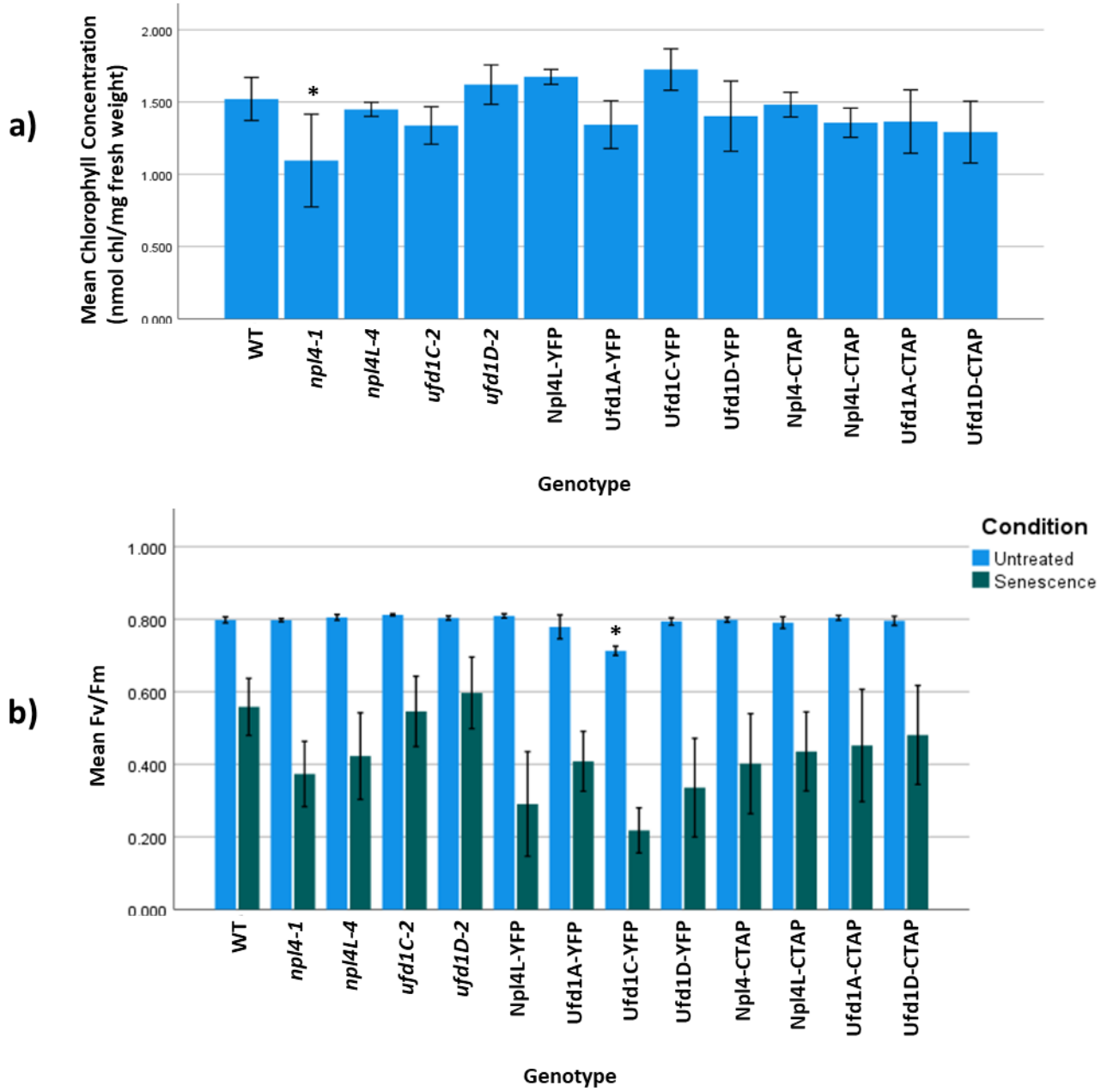


Figure 4.13 Bar charts of chlorophyll content and photosynthetic performance (Fv/Fm). a) A simple bar chart of the mean chlorophyll concentration of 28-day old plants grown under standard long day conditions. b) The proportion of developed plants (those with expanded cotyledons) from different mutant lines grown on salt containing media. Error bars are ± 1 Standard Error of the mean. Bars labelled with an asterisk show genotypes with significant (α = 0.05) differences to WT.

Chapter V

General Discussion

CDC48 is a conserved homohexameric member of the AAA+ (ATPases Associated with diverse cellular Activities) chaperone protein family (Bègue et al., 2017). It has a diverse suit of roles, being found to function not only in the UPS, but membrane fusion, DNA replication, immunity, and gene expression amongst others (Baek et al., 2013). The specificity of its action is determined by the adapter protein it interacts with. The Ufd1-Npl4 heterodimer has an established role in the recruitment of Cdc48 to ubiquitin proteasome mediated protein degradation, most evidenced in mammals and yeast, but also in plants (Bègue et al., 2017). My principle aims in this thesis were to characterize the Npl4 and Ufd1 proteins in *Arabidopsis thaliana*, and crucially determine whether they participated in chloroplast associated protein degradation.

Initially, I embarked on bioinformatic analysis to examine the data present within the protein sequences of Npl4 and Ufd1. I have found that the domain architecture of Npl4 is broadly conserved between *Arabidopsis*, yeast and humans. There is one major difference in the lack of zinc finger domain, which is seemingly absent from the Viridiplantae. This is a striking difference which will surely impact the mode of operation of Npl4 in the unfoldase activity of Cdc48. Literature on the role of yeast Npl4 on the substrate extraction process, found reduced activity when both zinc finger motifs in the domain were mutated (Bodnar *et al.*, 2018). In human Npl4 it was also recently observed that the zinc finger interaction with Cdc48 was crucial for its unfolding activity (Pan *et al.*, 2021). Thus, the mechanism of action of *Arabidopsis* Npl4, and potentially Npl4 in plants generally, may differ to that which is known from other organisms. There is some precedent for variation in the zinc finger domain of Npl4 sequences, as human Npl4 additionally has a C-terminal zinc finger domain; this has been shown to interact with polyubiquitinated substrates, though it is unclear if it is essential for the protein's function (Pan *et al.*, 2021).

There is considerably more divergence between the presented Ufd1 sequences. The three proteins Ufd1A, Ufd1B and Ufd1D are all fairly similar in domain architecture to yeast Ufd1 (Twomey *et al.*, 2019). The major difference between Arabidopsis and yeast Ufd1 is in the number of SHP boxes, as Arabidopsis proteins seemingly contain one instead of two — it is plausible that interactions with Cdc48 are still sustained by only one copy of the motif. Intriguingly, I observed a notable difference in the domain architecture of the larger Ufd1C protein: its SHP box and NBM are absent, and it has a C-terminal zinc finger domain. In the context of Cdc48 adapter proteins, there is precedence for zinc finger domains interacting with ubiquitin in human Npl4. However, zinc finger domains do have a myriad of functions, so I cannot make a reliable conclusion about why Ufd1C possesses this domain without experimental evidence. I believe I was able to correctly identify the closest homologous sequence to this zinc finger domain, categorizing it as a TRAF type – though this yielded little functional information. The related TRAFD1 protein in humans is involved in the immune response, through negatively regulating activity via toll like receptors {Witwicka, 2015 #278}. As CHLORAD is also known to intersect with the immune response (Sowden, thesis, unpublished), it is plausible that Ufd1C may perform a related function.

The information yielded from protein structure prediction using Alphafold2 provided much additional support for the domain annotations. Remarkably, conservation of structure allowed for the simulation of interactions between pairs of proteins, often bearing resemblance to structural data from yeast and human studies. A highlight of this data was the observed conservation of a ubiquitin binding groove in Arabidopsis Npl4 and Npl4L. This would suggest at least some aspects of the mechanism of operation {Twomey, 2019 #111}{Ji, 2022 #285} of Npl4 are conserved between plants and yeast.

The phylogenetic reconstructions are consistent with the theory of there being a degree of redundancy between the different Npl4 and Ufd1 isoforms {Li, 2022 #192} – most strongly in relation to Npl4 and Npl4L, and Ufd1A and Ufd1D. Ufd1C and Ufd1B reside within separate clades in the angiosperms in my phylogeny, indicating that these sequences diverged

relatively early during plant evolution and thus have potentially acquired new or distinct functions. The existence of multiple isoforms of Npl4 and Ufd1 indicate that it is not an uncommon feature across the angiosperms. Additional phylogenies using model organisms from across the eukaryotes, provides intriguing data regarding the relationships of Npl4 and Ufd1 proteins. It would appear that the lack of zinc finger domain in Npl4 is an exclusive characteristic of the Viridiplantae. It also highlights the degree of differences in the sequence of Ufd1C. It is grouped with TRAFD1 sequences from other eukaryotes, which may suggest that it is not a true Ufd1 protein. However, the UT3 domain is conserved within all of these sequences, suggesting interaction with ubiquitin is possible at least.

The initial set of experimental investigations allowed me to establish whether any of Ufd1 and Npl4 could participate in CHLORAD. My YFP localization and BiFC interaction data enabled me to rule out Ufd1B as being involved in CHLORAD; solely nuclear localisation was ever observed for this protein. A nuclear function is supported by the identification of Ufd1B in a study of chromatin associated processes {Mérai, 2014 #256}. I found that the Npl4, Npl4L and Ufd1A, Ufd1C and Ufd1D isoforms all clearly have the capacity to participate in the process: subcellular localisation was cytosolic in each case, and I observed interaction with TOC substrates and CHLORAD components.

The localisation of interactions between with surprisingly, and reproducibly, different with Cdc48 compared to all other BiFC interactions and YFP micrographs. This did not appear to be related to the level of expression. This would therefore suggest that Npl4 and Ufd1 are capable of passively moving through the nucleus and cytosol, but only exist as a functional complex with Cdc48 in specific places in the cell. It is possible that the recruitment of the Cdc48-Npl4-Ufd1 complex may be dependent on different cellular signals and highly regulated.

An emerging pattern when considering the interaction and mutant data, is that Npl4 is the strongest candidate for involvement in CHLORAD. However, Npl4L shows many of the same interactions and redundancy between the two isoforms may indeed by expected. The recent

paper identifying Npl4 and Ufd1 in involvement in the degradation of chloroplast proteins, also found that they functioned redundantly {Li, 2022 #192}. Surprisingly based on predicted structures and protein domain assignments, Ufd1C is arguably the most promising Ufd1 protein to be involved in CHLORAD. Involvement of Ufd1C in CHLORAD is perfectly plausible if it is a true Ufd1 protein: this is because the recruitment of CDC48 is largely grouped by function. For example, during cellular division CDC48 is recruited by SYP31 (Bègue et al., 2017), so it is likely that any Ufd1 protein could be utilised in a UPS process. Indeed, involvement in the ubiquitin associated protein degradation systems may non-specifically involve any Npl4 or Ufd1 proteins. In mammalian cells, the CDC48-NPL4-UFD1 complex is involved in ubiquitin related protein degradation at the ER, ribosome, and chromatin (Bruno et al., 2013). But how this may occur with Ufd1C is quite the mystery. The predicted features and interacting regions are markedly different to what is understood from the literature. Whereas, Ufd1A, Ufd1B and Ufd1D show similar predicted interactions, for example: the SHP box facilitates interactions between these proteins and Cdc48, whereas with Ufd1C it binds to uncharacterised predicted zinc finger. The accuracy of these simulations is unlikely to be appreciably high, as even with Npl4, all Ufd1 proteins fail to sustain interactions at between the MPN domain of Npl4 and the NBM of Ufd1. It does however mean that how Ufd1C could function in such a process as CHLORAD is somewhat of a mystery.

A question may also arise of whether there is an adaptive value in retaining so many copies of the same gene (which of course there needn't be (Gould and Lewontin, 1979)). Co-expression data shows little to no impact of perturbative effects or development, thus indicating that they may all function redundantly between isoforms. Another possibility is that this simply acts to increase the abundance of the recruiting complex and increase the efficiency of UPS related functions; CDC48 is highly abundant, estimated to make up 1% of cytosolic proteins (Broad et al., 2016). Additionally, it could be the case that increased copy numbers are retained as an adaptive feature from pathogenic selective pressure.

The primary aim from Chapter IV was to determine whether Npl4 and Ufd1 have any functional significance in CHLORAD. I must cautiously conclude that I failed to address this aim. I believe this project is fundamentally limited in this regard by lack of appropriate mutant lines to address this question. In order to establish a genetic link, it is key to examine whether there is a suppressive phenotype in the TOC mutant *ppi1.* This has formed the initial strand of evidence for CHLORAD proteins and may be considered the first barrier to entry for Npl4 and Ufd1 to be fit within the CHLORAD model.

What I instead performed, was a mutant analysis of acquired knockout lines and generated overexpression lines. I sought to examine any detectable physiological effects, particularly in relation to chloroplast biology. The experiments I performed had previously been demonstrated on repeated occasions to show functional significance of CHLORAD proteins in developmental and stress associated phenomena. The results I generated are largely neutral with no obvious patterns, suggesting little to no functional involvement of Npl4 or Ufd1 proteins in response to de-etiolation, senescence, osmotic or salt stress.

Reciprocal data between the OX and KO lines, should not perhaps be expected. Previous analysis has shown redundancy between pairs of Npl4 and Npl4L, and Ufd1A and Ufd1D (Li *et al.*, 2022; Mérai *et al.*, 2014). Given this, the OX lines might be of more value. Indeed, the most substantial result from this section was that from Npl4-CTAP and Ufd1D-CTAP. The specific increase in Toc75 protein abundance was 3-8 fold higher than the WT expression. I speculated that this is likely the result of dominant negative mutation, as the observation was only associated with TAP constructs and independent of expression. Nonetheless, it may signify a functional link through. The next question is how does that manifest in a way that fits with the rest of the observed data? Interaction with Toc75 was not examined, due to structural issues rendering it infeasible for the chosen experiments. Yet, interaction was seen with Toc159 and Toc33, yet no impact was produced on their abundance. Overaccumulation of Toc75 is also not present to the same degree in the blotting conducted during the de-etiolation experiment, thus there may be a developmental aspect as well. It is

a confusing scenario, and the answer may lie in that it is an indirect effect, caused by a secondary process. This is certainly a phenotype that should be investigated further.

Limitations and Future Directions

It is arguable how useful early bioinformatic analysis was in the start of this project databases in this project. Systems biology was once described as “low input, high throughput, no output science” by Nobel prize winner Sydney Brenner. Mathematical models are frequently unable to account for biological complexity or are based on atomistic data. Nonetheless, taken with caution, they can aid hypothesis testing. Indeed I was able to validate some predicted protein interactions, such as with Cdc48. My phylogenetics approach was not optimal, but instead a compromise for a fast-computing times. Our final tree topology may therefore be of limited accuracy, and would benefit from reconstruction using additional techniques. While Alphafold structures show very high accuracy, they are still not exactly comparable to experimentally determined crystal structures – especially for multimers. There was also essentially no follow up analysis performed to validate observations.

I believe the localisation and interaction data is relatively robust. However, additional methods to confirm interactions would be of benefit. I had endeavoured to perform proteomic analysis of Npl4 and Ufd1 interactomes, and I still believe this is worthwhile in the future. Other experimental methods such as yeast two hybrid and transient expression in Tobacco may also be useful. Utilizing the stably generated plant material further for interaction analysis is also perhaps wise.

Much work is still needed to assess the role of Npl4 and Ufd1 in CHLORAD. I believe this must start with the generation of complete mutant sets, and the phenotypic analysis to determine their operation in CHLORAD. I also think Ufd1C should be the focus of considerable future study. Domain analysis will I think form a highly useful experimental tool. Both elimination of domains and pulldowns with specific domains may reveal additional functional information. In

particular, the zinc finger of Ufd1 is a tantalizing domain offer a mechanism of action for Ufd1 proteins. It would be interesting to see if it is merely a Cdc48 interacting domain (as predicted by Alphafold) or has some other impact on ubiquitination or substrate processing.

Chapter VI

Materials and Methods

6.1 Bioinformatics

6.1.1 Predictive Tools and Databases

Predictions of protein structural homology were modelled using the intensive search option on the sever Phyre2 (Kelley *et al.*, 2015) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). The search output was used for domain analysis, by generating a consensus with annotations from Interpro and multiple sequence alignment with the MAFFT (multiple alignment using fast Fourier transform) program (Katoh & Standley, 2013) to search for specific motifs shared with yeast proteins.

ARAMEMNON (Schwacke *et al.*, 2003) (http://aramemnon.uni-koeln.de/) was used to investigate whether any protein contained transmembrane domains. SUBA4 (Hooper *et al.*, 2017) (http://suba.live/) provided non-experimental predictions of subcellular location. Expression profiles were collected from GENEINVESTIGATOR (https://genevestigator.com/gv/index.jsp), using data from the Arabidopsis gene atlas (Schmid *et al.*, 2005). Potential protein-protein interactions were then initially predicted using the BAR *Arabidopsis* interactions viewer database (Fucile *et al.*, 2011), followed by STRING database (Szklarczyk *et al.*, 2021). For the STRING database, the selection criteria was modified to >0.900 (highest confidence).

6.1.2 Phylogenetics

Sequence homologues were discovered through a BLASTP (Basic Local Alignment Search Tool Protein) search of the Viridiplantae protein database of Phytozome v.12.1 (https://phytozome.jgi.doe.gov/pz/portal.html#). The top BLASTP search results were downloaded and formatted into a fasta file (E values < 3×10^{-155}). Processing of sequences

selected only primary splice variants and excluded sequences with insufficient annotation (e.g. no attribution to NPL4 or UFD1). A single search was sufficient for NPL4 to identify basal embryophytes. UFD1 necessitated individual searches for each of the four homologues in *Arabidopsis* due to lower sequence identity between them. The top results were selected from each search. Two outgroup sequences were selected from *Homo sapiens* and *Saccharomyces cerevisiae.*

Multiple sequence alignments were produced using the program MAFFT on an EBI server (https://www.ebi.ac.uk/Tools/msa/mafft/). The output files were downloaded in the fasta format and inspected using the programme SnapGene software (www.snapgene.com) to search for anomalies.

Tree reconstruction was based on the principle of maximum likelihood. Computation was performed using the web server IQ-TREE (Trifinopoulos *et al.*, 2016) (http://iqtree.cibiv.univie.ac.at/), which offers comparable performance to RAxML and PhyML (Nguyen *et al.*, 2015). Model selection was performed using ModelFinder (Kalyaanamoorthy *et al.*, 2017), finding the best-fit model as JTTDCMUT+G4 for NPL4, and JTT+G4 for UFD1. For the model organism trees, the best fit model for Npl4 was LG+I+G4, while it was VT+F+G4 for Ufd1. Branch support was calculated with 1000 bootstrap replicates. The tree was imaged and managed using iTOL (https://itol.embl.de/).

6.1.3 Protein Structural Predictions

Protein data bank files were accessed from the Alphafold2 (Jumper *et al.*, 2021) database, for yeast and *Arabidopsis* Npl4, Ufd1 and Cdc48 sequences. Images for figures were rendered in ChimearaX (Pettersen *et al.*, 2021), using full shading and rainbow coloration of the sequence from N to C termini. The matchmaker function was used to align structures. For multimer and fusion proteins, the Google Colab Fold server (Mirdita *et al.*, 2022) was used. Amino acid sequences were entered for various combinations of multimers, with the

data sourced from Uniprot. The top-ranking model was taken as the final structure and visualised in ChimaeraX.

6.2 Plant Handling

6.2.1 Plant growth conditions

The model plant *Arabidopsis thaliana,* of the Columbia-0 (Col-0) ecotype was exclusively used in this study. Plants were generally grown in a glass house, under long day conditions (16 hrs light, 8 hrs dark); the glass house was air conditioned to maintain a temperature of approximately 20°C with a humidity of 50-70%. The light in the greenhouse was supplemented with an LED light source (Phytolux Attis-7 LEDs: λ= 460-660 nm, qp= 285 µmol s^{-1} m^{-2}), which remained on for 16 hrs each day. When sown on plates and not soil, plants were grown in a controlled environment growth chamber (Percival Scientific Inc.), also under long-day conditions (16 hours light, 8 hours dark). Light intensity was set at approximately 110 µE m^{-2} s^{-1}, and the temperature was held at a constant 20°C with a humidity of 70%.

6.2.2 Plants grown on media

For experiments involving the use of seedlings and selection for antibiotic or herbicide resistance, plants were grown on ½ Murashige and Skoog (MS) media. The MS media was composed of 4.3 g/l MS basal salt mix (Duchefa Biochemie), 0.05% (w/v) 2-(*N*-morpholino) ethanesulfonic acid (MES monohydrate; Melford), and 0.6% (w/v) phytoagar – phytoagar was omitted in instances where liquid MS was desired. For certain experiments, MS media was supplemented with sucrose 0.5% (w/v), in order to boost germination efficiency. The pH of MS media was adjusted to 5.7 with 1 M KOH, before the addition of phytoagar. MS media supplemented with mannitol or NaCl was also adjusted to a pH of 5.7 before the addition of those components. MS media was sterilised by autoclaving at 121 °C, 103.421 kPa for 15 minutes. The liquid MS media was allowed to cool to approximately 60°C, at which point chemical additives (such as antibiotics) would be added if required. The molten media was then poured into either 90 mm diameter circular petri dishes or 120 mm x 120 mm square petri

dishes, to a depth of 1 cm. Pouring of media took place under sterile conditions in a laminar flow hood. The media was then allowed to set for at least 30 minutes with the lids removed.

Seeds were routinely surface-sterilised using an ethanol-based sterilisation solution (70% (v/v) ethanol, 0.05 % (v/v) Triton X-100). For screening T1 seed, which presents additional contamination risk due to exposure to a solution containing bacteria and sugar, a sodium hypochlorite solution was preferred (5% (v/v) sodium hypochlorite, 0.05% (v/v) Tween-20). Following a 10-minute incubation with shaking (250 RPM on an orbital platform) with the first sterilising solution, the seeds were resuspended in 100% ethanol for a further 10 minutes of shaking. The sterilised seeds were then transferred to ethanol sterilised filter paper and left to dry in a laminar flow hood for at least 20 mins. After drying, seeds were sown directly onto plates and stratified at 4°C for 3 days. In most cases, plants were used in downstream experiments after 10-14 days of growth.

6.2.3 Plants grown on soil

Plants were either sown directly onto soil or transferred from MS plates. The soil medium contained a 3:1 ratio of Levington F2: vermiculite (Sinclair, fine particle size). The soil was then supplemented with 0.4 g/L of the insecticide Imidasect (Sinclair), to prevent the growth of sciarid flies (a pest). When sown directly onto soil, 14 pots were fitted per tray. For plants transferred from MS to soil, module trays containing 24 cells were preferred. Trays were bottom watered as required. Seeds were sown directly onto wetted soil by gently scattering from filter paper and stratified at 4°C for 3 days before transfer to a greenhouse. In some cases, seedlings were transferred from MS plates to soil after 2-3 weeks of growth.

6.2.4 Crossing plants

Plants were grown on soil for roughly 4-6 weeks, until primary and secondary bolts had developed. Crosses were attempted between plastid protein import mutants with pale phenotypes and various mutants of Npl4 and Ufd1. As the plastid protein import mutants

possessed a recessive chlorotic phenotype, those plants were chosen as the female acceptor – as successful crossing in the F1 generation could be easily determined by suppression of the chlorotic phenotype.

Siliques and opened flowers were discarded to prevent unwanted seed developing. Crossing was performed in the morning, so that the open flowers could be used as male donors. Under a light microscope, unopened flowers (to reduce the risk of prior fertilisation) were dissected to remove sepals, petals, and stamens. This was achieved by careful peeling using 'jewellers' forceps. The opened male flowers were then gently brushed on the carpels until pollen was visibly attached. This was performed at for at least three flowers per cross. Siliques were then harvested from crossed flowers within 14 days from development.

6.3 DNA and RNA work

6.3.1 DNA extraction from plant tissue

Tissue was harvested from mature rosette leaves for DNA isolation, either 1 leaf or at least 50 mg. The tissue was then immediately flash-frozen in liquid nitrogen in a 2 ml Eppendorf tube containing a sterile 5 mm diameter steel bead. The tissue was homogenised into a fine powder by the use of a TissueLyser (Qiagen) at 24 Hz for 1 minute, then immediately flash frozen again. Tissue was either stored at -80 °C or thawed on ice for immediate extraction. DNA extraction was performed using a CTAB based protocol. 500 µl of DNA extraction buffer (100 mM Tris(hydroxymethyl)aminomethane-HCl (Tris-HCl) pH 8.0, 1.4 mM NaCl, 20 mM (ethylenedinitrilo)tetraacetic acid (EDTA), 2.0% (w/v) Hexadecyltrimethylammonium bromide (CTAB), 1% β-mercaptoethanol (BME – added fresh)) was added to each sample. Each sample was then vortexed for 10 seconds and incubated at 60 °C for 30 minutes, with another brief vortexing half way through. 200 µl of chloroform:isoamyl alcohol 24:1 (v/v) were then added to each sample, followed by brief vortexing, then centrifugation at 12,000 RPM, 4 °C, for 10 mins. The upper aqueous phase was then removed and transferred to a new 1.5 ml Eppendorf tube containing 150 µl of chloroform:isoamyl alcohol 24:1 (v/v). The solution was

vortexed again, then centrifuged at 12,000 RPM, 4 °C, for 10 mins. The upper aqueous phase was removed again and transferred to another new 1.5 ml Eppendorf tube containing 300 µl of isopropanol, and mixed by inverting several times. DNA was precipitated by centrifugation at 12,000 RPM, 4 °C, for 30 mins. The supernatant was removed and the DNA pellet was washed with 180 µl of 70 % ethanol. The pellet was spun down by centrifugation (12,000 RPM, 4 °C, 10 mins), air dried, then resuspended in 40 µl of dH_2O. DNA samples were stored at -20 °C.

6.3.2 PCR

Amplification of DNA by polymerase chain reaction (PCR) was performed in each instance using a T3 Biometra Thermocycler. For general sequence amplification, a lab purified Taq polymerase was used, prepared according to a standard protocol (Pluthero, 1993). PCR solutions consisted of a standard concentration in most cases in a 20 µl reaction volume (10 mM Tris-HCL pH 8.3, 50 mM KCl, 1.5 mM MgCl, 200 µM mix of deoxyribonucleoside triphosphates (dNTPs), 500 nM of each primer, and 0.02 units (U) per µl of Taq DNA Polymerase). For sequencing or cloning applications, a proof-reading polymerase was used (Phusion Heat Shock II; ThermoFisher) with the supplied buffers and the addition of 3% (v/v) dimethyl sulfoxide (DMSO). The thermocycler programme varied according to primer annealing temperature (calculated using ThermoFisher's prediction tool, based on a standard formula (Allawi & SantaLucia, 1997)) and expected amplicon length (30s/1000 base pair (bp)); 30 amplification cycles were used in most instances.

6.3.3 PCR genotyping

T-DNA insertion mutants were genotyped by PCR using the SALK LBb1.3 primer, and a genomic primer set designed using the SIGNAL online toolkit. Gene specific primers were designed to produce amplicons of around 1200 bp, and an amplicon between 500-800 bp when used in combination with the T-DNA border primer. PCR reactions were run in series to check for the presence of T-DNA and genomic specific amplicons to evaluate the zygosity of

the knock-out mutants. 25 ng of genomic DNA were used as the template in each reaction. Bands were visualised by agarose gel electrophoresis. Homozygous T-DNA plants with only T-DNA border specific amplicon, and no genome specific amplicon, were propagated.

In the case of CRISPR lines, genomic DNA from antibiotic resistant plants was amplified from respective Npl4 and Ufd1 genes. PCR products were extracted from agarose gels using GenElute Gel Extraction Kit (Sigma). Purified PCR products were sent for Sanger sequencing (Source Bioscience) with gene specific primers. DNA edits were assessed by visual inspection of chromatograms, following sequence alignment to the reference CDS using Snapgene (Dotmatics), as well as output from TIDE software for DNA insertion and deletion detection (Brinkman *et al.*, 2014).

6.3.4 Agarose gel electrophoresis

Agarose gel electrophoresis (AGE) was used to resolve DNA band sizes and check the quality of DNA extractions from various sources. Gels were typically composed of a 1% (w/v) agarose solution, in 0.5× TBE buffer (40 mM Tris-HCl pH 8.3, 45 mM boric acid, 1 mM EDTA). The agarose was melted through heating the solution in a microwave, then allowed to cool to ~60°C. SYBR Safe DNA stain (Invitrogen) was then added at a ratio of 1:10,000. The solution was then poured into a gel casting mould with a comb to create wells of varying size and number, and allowed to set. DNA samples were loaded in 1× Orange G loading buffer (0.1% (w/v) Orange G, 50% (v/v) glycerol) and run in a tank (Mini-Sub Cell GT, Bio-Rad) containing 0.5× TBE buffer at 120 V for 30-40 minutes. 1kb+ DNA ladder was loaded into at least one well, to allow for a size reference of DNA fragments. Each gel was visualised under UV light, in a NuGenius (Syngene) gel imaging system.

6.3.5 RNA extraction from plant tissue

At least 50 mg of tissue was harvested from either from mature rosette leaves or 10-14-day old seedlings for RNA isolation. The tissue was homogenised in the same process as used

for DNA extraction. Total cellular RNA was extracted from the homogenised tissue according to the protocol of the Spectrum Plant Total RNA Kit (Sigma). To limit the contamination of DNA, an optional on-column DNA digestion step was included. To perform this, 1x DNase I reaction mixture (Sigma) was incubated on the column according to the manufacturer's instructions, and subsequently washed off. RNA was eluted with the supplied elution buffer in a volume of 50 µl. RNA concentrations were determined through the use of a NanoPhotometer P330 (Implen), which produced spectrophotometric data to calculate a concentration in ng/µl. RNA samples were then stored at -80 °C.

6.3.6 Reverse Transcription (RT)

Total RNA extracts from plants were used as a template for complementary DNA (cDNA) synthesis. The Superscript IV reverse transcription kit (Invitrogen) was used in accordance with the manufacturer's instructions. Based on recorded RNA concentrations, 1 µg of total RNA was added to each 20 µL reaction for all samples. An optional RNase inhibitor (RNasin Plus; Promega) was included in the reaction mixtures to limit RNA degradation. Oligo dT primers were used to amplify mRNA, and reaction was catalysed at median recommended temperature of 52 °C. The resultant cDNA reaction was then diluted in a 1:5 with ddH_2O, for immediate use in PCR reactions.

6.3.7 Semi-quantitative RT-PCR (sqRT-PCR)

Plant cDNAs, normalised with respect to the RNA concentration, were used for all semi-quantitative reverse transcription PCRs (sqRT-PCR). In the case of expression analysis of knock-out and overexpressing lines, RNA was isolated from 10-day old seedlings. Primers for sqRT-PCR were designed using the NCBI primer BLAST tool, to ensure non-specific annealing to additional templates. Primers were designed such that amplicons would be produced in the region of 500-800 bp (Table 6.1). cDNA were used as templates in a series of PCR reactions with gene specific, and control primers. The control primer chosen was

common standard ACT2, due to consistent expression in all tissues. Standard PCR conditions were used, with annealing temperatures tailored to each primer set. Cycle number was optimised to prevent signal saturation, and capture exponential amplification to allow for relative quantification. 27 cycles were determined as optimal based on eIF4E1 amplification and used for all sqRT-PCR analysis. DNA bands were resolved on a 1% agarose gel and captured as previously described. Band intensity was quantified using the “gels” analysis tool on 8 bit images in ImageJ – 3 technical replicates were quantified for statistical analysis.

6.4 Molecular cloning and bacterial work

6.4.1 *E. coli* strains, growth conditions, and DNA isolation

Escherichia coli was generally cultured in Lysogeny broth (LB; 10% (w/v) tryptone, 5% (w/v) yeast extract, 10% (w/v) NaCl). For growth on plates, agar was added to the LB at a concentration of 1.5% (w/v); 90 mm diameter plates were used, with the solution at a 10 mm depth. All LB media was autoclaved to sterilise at 121 °C, 103.421 kPa for 15 minutes. Antibiotics were added after the solution had cooled to 60 °C. *E. coli* was grown at an optimal temperature of 37 °C, either with vigorous shaking to provide aeration for liquid cultures (250 RPM) or without shaking for plates.

For molecular cloning work, a proprietary variety of the DH5α strain called NEB 5-alpha was used from New England Biolabs (NEB). Heat shock transformation was utilised to introduce foreign plasmids, with slight modifications to the manufacturer’s protocol. A reduced volume, 25 µL aliquots, of competent cells were used per transformation and thawed on ice after removal from -80 °C. 1-5 µL of cloning reaction or 1 ng of plasmid DNA was added to each tube, and mixed by gentle flicking. The cells were then incubated on ice for 30 minutes. Heat shock transformation occurred by immersion in a 42 °C water bath for precisely 30 seconds. The cells were then placed on ice for 5 mins, after which 500 µL of LB was added. The cells were allowed to grow for 1 hour at 37 °C, with vigorous shaking of the tube positioned

horizontally. The entire volume was then plated on LB agar plates containing an antibiotic for selection of the plasmid for propagation at standard working concentrations (e.g. 100 µg/mL ampicillin, 50 µg/mL kanamycin, 50 µg/mL spectinomycin). Cells were spread around the plate until dry with a sterile L-shaped spreader, then placed in a 37 °C incubator overnight. Colony PCR was used to evaluate successful transformation of resistant colonies. Briefly, a single colony was picked with a 10 µl pipette tip and pricked into a PCR tube before retention in 50 µl of LB culture. Standard PCR and AGE protocols were then used to assess for the correct insert with sequence specific primers. Colonies displaying the desired amplicon were cultured for plasmid isolation and sent for Sanger sequencing (Source Bioscience) using gene specific primers. Sequence reads were aligned to a reference construct maps using Snapgene (Dotmatics) software, to confirm the complete sequence identity.

Glycerol stocks containing pre-existing sequenced plasmids were utilised to eliminate the burden of repeated cloning. *E. coli* plated from glycerol stocks (40% (v/v) glycerol and bacterial culture at the stationary growth phase), were streaked onto LB agar plates with the appropriate antibiotic using a sterile inoculation loop. A zig-zagging motion was used while spreading to dilute the glycerol stock to generate single colonies for culturing in liquid LB.

Plasmid DNA was isolated cultured *E. coli* through the use of either a miniprep kit (Qiagen Plasmid Mini Kit) or a midiprep kit (Qiagen Plasmid Midi Kit). 5 mL and 100 mL of culture (LB plus antibiotic) was used respectively for each miniprep and midiprep. Cultures were grown at 37 °C, 250 RPM for 12-16 hours before harvesting cells. The manufacturer's protocol was followed for the miniprep kit as written. Midiprep extractions required slight modifications as the centrifugation conditions were not able to be met. A Sorvall Legend RT centrifuge was used, fitted with buckets to accommodate 50 ml falcon tubes. The maximum centrifugation speed was 4600 RPM, which was used in place of the higher recommended conditions – longer centrifugation times were used to compensate for this. Midipreps were used to generate

high concentrations of plasmid DNA (>1 µg/µL) for transfection experiments. DNA concentrations were measured as previously described.

6.4.2 Agrobacterium growth conditions and transformation

Transgenic plants were generated by *Agrobacterium tumefaciens* mediated 'floral dip' transformation (Clough & Bent, 1998). Homemade competent cells of the GV3101 strain of *Agrobacterium* was used for transformation, generated by a standard protocol (Clough & Bent, 1998). *Arabidopsis* coding sequences (CDS) of Npl4 and Ufd1 (excluding Ufd1B) isoforms were introduced into pK7YWG2, (Karimi *et al.*, 2002), NTAPI and CTAPI vectors (Rohila *et al.*, 2004) via the Gateway cloning system. Restriction cloning was used to introduce a single guide RNA (gRNA) into the CRISPR (clustered regularly interspaced palindromic repeats)-Cas9 vector pKI1.1R – two gRNAs were designed to target the 5' region of each gene using CRISPR-P v2.0 (Lei *et al.*, 2014). The sequence verified constructs were used to transform *Agrobacterium* cultures via the freeze-thaw method (Hofgen & Willmitzer, 1988). In each transformation, 1 µg of plasmid DNA was added to a 1.5 mL Eppendorf tube with a 100 µL aliquot of competent cells. The contents were mixed by flicking then flash-frozen in liquid nitrogen. The tubes were then thawed in a water bath at 37°C for 5 minutes. 1 mL of LB medium was then added to each tube, before transferal to an incubator set at 28°C for 4 hours, 250 rpm. The transformed cells were pelleted by centrifugation (30 seconds, 5000 × *g*), and resuspended in the residual liquid. The cells were then plated onto LB agar plates with stringent antibiotic selection (spectinomycin, 50 µg/mL and gentamycin, 10 µg/mL; these antibiotics were used for selection in all LB media) and incubated at 28°C for two days. A starter culture of 5 ml of LB with antibiotics was then inoculated with a single transformed colony, and incubated overnight at 28 °C with vigorous shaking. 250 µl of the starter culture was then inoculated into 250 ml of LB media, and grown for 24 hrs. the OD600 measurement reached 0.8. Fresh sucrose (5%; w/v) was prepared and cells harvested by centrifugation (5000 RPM for 10 mins) were resuspended until the optical density (OD) reached 0.8 – as determined by spectrophotometric readings at 600 nm with a NanoPhotometer P330 (Implen).

Plants destined to be transformed were grown on soil for 4-5 weeks. After the emergence of the primary bolt and it had reached a height of approximately 10 cm, it was clipped to encourage the growth of secondary bolts, for use within the following week. Approximately, 4-6 plants were transformed per construct. Developing siliques were removed to reduce the carryover of non-transformed seeds.

The *Agrobacterium* culture suspended in sucrose solution was supplemented with a wetting agent to improve the access of the bacteria around the flowers: Silwet L-77 (Lehle Seeds) was added to a final concentration of 0.3% (v/v). Aerial parts of the flowers were suspended in the transforming solution for 30 seconds with gentle agitation. Plants were placed horizontally in trays with propagator lids, and placed in a greenhouse for one day, after which the lids were removed and the plants were grown normally until harvesting for seed. T1 seed was sown onto MS media supplemented with appropriate antibiotics given by the transforming construct between the border insertion sites (phosphrinocin 50 µg/mL (TAP vectors); hygromycin 25 µg/mL (pH2GW7, pKI1.1R)). Surviving plants were transferred to soil after 2-3 weeks and grown until maturation. T2 and T3 generation plants were then screened by western blotting for selection of expression of the epitope tagged transgene, to obtain and confirm homozygous lines.

6.4.3 Restriction enzyme cloning

Restriction cloning was used to generate a number of constructs for bimolecular fluorescence complementation (BiFC), co-immunoprecipitation (Co-IP), CRISPR knock-outs, and subcellular localisation experiments. *Arabidopsis* coding sequences (CDS) of Npl4 and Ufd1 isoforms were introduced into pSAT4A-nEYFP-N1 and pSAT4A-cEYFP-N1 (Tzfira *et al.*, 2005), pE3c (Dubin *et al.*, 2008) and pKI1.1R (Tsutsui & Higashiyama, 2017). The pSAT4A-nEYFP-N1 and pSAT4A-cEYFP-N1 vectors create a split-YFP interaction-based assay: each vector encodes either the N-terminus or C-terminus of the YFP fluorophore under the control

of the constitutive expression promoter CaMV 35s; if proteins interact the YFP fluorophore will form and produce fluorescence indicating protein-protein interaction and the localisation of that interaction (Kudla & Bock, 2016). The pE3c vector introduced a 6xMyc epitope tag at the C-terminus of proteins; it was used for co-immunoprecipitation assays. Additional constructs used in experiments (e.g. TOC proteins) were generated by former members of the lab and acquired from glycerol stocks.

Restriction sites were introduced into the 5' extensions at the end of primers, which in turn were designed to bind to the 3' and 5' regions of the respective isomer's CDS. Restriction sites were chosen that were compatible with the multiple cloning sites of the vectors, had single cut sites, and were not present in the CDS sequence. Primer sequences were designed to be roughly 20 bp with GC content of roughly 50%. Annealing temperatures of primers were predicted in addition, to ensure compatibility in PCR. CDSs were amplified by PCR from WT *Arabidopsis* cDNAs, using the proof-reading polymerase Phusion. Amplified bands were isolated after separation by AGE. The predicted band size was excised and DNA extracted through use of a GenElute Gel Extraction Kit (Sigma), according to the manufacture's protocol. Purified PCR products were then digested with appropriate combination of restriction enzymes (supplied by NEB), in a compatible buffer. Restriction digestion followed the manufacture's recommend protocol. In parallel, the vector plasmid was also digested by the same enzymes to produce compatible overlapping termini for ligation. Both digested products were again purified after separation by AGE, using a gel extraction kit. Ligated was performed using a T4 DNA ligase (Promega), in accordance with the manufacturer's protocol. The ligation mixture was then used to transform *E. coli* competent cells as previously described.

6.4.4 Gateway cloning

The Gateway cloning system was used to generate constructs for stable plant transformation (pH2GW7, NTAPI and CTAPI) and transient plasmid expression (pK7WYG2 and p2GW7 (Karimi *et al.*, 2002)). pH2GW7 is a binary vector which was used to generate stably

overexpressing lines (under the CaMV 35s promoter) with C-terminal YFP fusions for *in planta* localisation. The NTAPI and CTAPI are also binary vectors with CaMV 35s promoters, instead introducing N or C terminal TAP tags respectively, for TAP purification. pK7YWG2 and p2GW7 are transient overexpression vectors introducing a C-terminal YFP tag, and no tag respectively.

Gene specific primers to Npl4 and Ufd1 isoforms were designed with attB1 and attB2 recombination sites appended to 5' region of the primers, at the 5' and 3' ends respectively; the remainder of the sequence was designed with the same principles as the restriction cloning primers. The amplicon was generated using a two-step PCR and the Phusion polymerase. AGE and gel extraction were performed to generate purified products, which were then introduced into the Gateway entry vector pDONR201; the BP clonase (Invitrogen) reaction was performed to do this, according to the manufacturer's instructions. The clonase reaction was then used to transform and screen *E. coli* competent cells as previously described.

After the sequence identity of the entry clones was confirmed, plasmid DNA from the entry clone and destination vectors were recombined in the LR clonase reaction. In the case of Myc tagged constructs, sequences which were introduced into pE3c to create C-terminal translational fusions to 6xMyc by restriction cloning, then the LR reaction was used to introduce the sequence into the p2GW7 vector. The LR clonase reaction (Invitrogen) followed the manufacturer's instructions. The clonase reaction was then used to transform and screen *E. coli* competent cells as previously described.

6.4.5 CDC48^{E581Q}-mScarlet Cloning

To generate the CDC48 'trap' mutant with the mScarlet tag, a combination of restriction and gateway cloning was used. mScarlet was amplified by PCR with from a plasmid donated by Dr. Qihua Ling, an attB2 site on the 3' region binding primer, and a restriction site on the 5' region binding primer. Conversely, the CDC48^{E581Q} sequence was amplified from the

CDC48^{E581Q}-YFP construct, with an attB1 site at the 5' binding primer, and restriction enzyme site at the 3' binding primer. The primers were designed so that the linear PCR product would be an in-frame C-terminal translational fusion with a stop codon existing only at the end of the mScarlet sequence. Restriction digestion of both PCR fragments was used to create compatible ends for ligation, as previously described. The purified linear PCR product was then introduced into pDONR201, then p2GW7 by Gateway cloning, for transient overexpression in plant cells.

6.5 Protein analysis

6.5.1 Protein extraction

Protein was harvested either from rosette stage tissue (for screening transgene expression) or from 10/14-day old seedlings. 30-50 mg of tissue was taken for each sample. The tissue was then immediately flash-frozen in liquid nitrogen in a 2 ml Eppendorf tube containing a sterile 5 mm diameter steel bead. The tissue was homogenised into a fine powder by the use of a TissueLyser (Qiagen) at 24 Hz for 1 minute, then immediately flash frozen again. Tissue was either stored at -80 °C or thawed on ice for immediate extraction. Lysate was extracted from each sample on ice using 250-300 µL of freshly prepared ice-cold protein extraction buffer (100 mM Tris-HCl pH 6.8, 10% (v/v) glycerol, 1% (w/v) SDS, 0.1% (v/v) Triton X-100, 5 mM EDTA, 10 mM dithiothreitol (DTT), and 10 µL/mL protease inhibitor cocktail (PIC; Roche)). After the addition of extraction buffer, each sample was vortexed for at least 10 seconds and incubated on ice for 20 minutes. The lysate was clarified by centrifugation at 14,000 RPM, 4°C for 20 minutes, and the clarified lysate was then retained in a fresh 1.5 ml Eppendorf tube.

For routine work, protein concentration was not determined, as the starting mass of tissue was normalised. However, in some instances protein concentration was determined using the Bradford assay. A standard curve of protein absorbance was generated using a range of Bovine Serum Albumin (BSA; Sigma) dilutions (0.5, 1, 2, 4, and 8 µg/µL). Disposable cuvettes

were used to record absorbance at 595 nm using an Implen Nanophotometer P330. A total of 1 ml of solution was added to the cuvette in each case, comprised of 200 µL of protein assay dye (Bio-Rad), 1 µL of protein extraction buffer, 1 µL of protein sample or standard, and ddH_2O to the desired volume. The absorbance of the standards was plotted against their concentration on a scatter graph and the gradient was calculated. The absorbance values were then divided by the gradient (when the intercept approximated zero) to give a concentration of protein in µg/µl.

Laemmli buffer (Laemmli, 1970) (0.1 M Tris-HCl pH 6.8, 2% (w/v) SDS, 20% (v/v) glycerol, 0.02% bromophenol blue, 0.1 M Dithiothreitol (DTT) – DTT was always added fresh) was added to protein samples in a 1:1 ratio. Samples were then boiled to denature proteins at 95 °C for 5-10 minutes, and stored at -20 °C for short term use.

6.5.2 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were resolved by mass using SDS-PAGE (Laemmli, 1970). Polyacrylamide gels were prepared using glass plates fitted in a gel casting system (Mini Protean III, Bio-Rad). The resolving gel solution was made to a strength of 12% acrylamide, as a compromise for resolving proteins of a wide range of molecular masses. The resolving gel consisted of 40% (v/v) of 37.5:1 acrylamide:bisacrylamide (30% (v/v); Geneflow), resolving buffer (375 mM Tris-HCl pH 8.8, 0.1% (w/v) SDS), 0.5% (w/v) ammonium persulphate (APS), 0.05% (v/v) tetramethylethylenediamine (TEMED)). APS and TEMED were added last as they initiate polymerization of the gel matrix. The resolving gel solution was pipetted into prepared glass plates which produced a gel size of 500 mm x 850 mm x 0.75 mm (height x width x depth), and immediately overlaid with isopropanol, to ensure an even gel surface. The gel was allowed to set over 30 minutes at room temperature, after which the isopropanol was removed. Stacking gel solution contained 17% (v/v) of 37.5:1 acrylamide:bisacrylamide (30% (v/v); Geneflow) stacking buffer (0.125 M Tris-HCl pH 6.8, 0.1% (w/v) SDS), 0.75% (w/v) APS, and

0.125% (v/v) TEMED). A 15 well-forming comb was immediately inserted after the addition of the stacking gel solution, increasing the height of the gel by approximately 200 mm.

Gels were assembled into an electrophoresis chamber (Mini-Protean III, Bio-Rad) and the inner chamber of the tank was filled with Tris-Glycine-SDS running buffer (TGS; 25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS), and allowed to overflow to the point where the outer chamber was half full. The inner chamber was monitored to ensure the wells were fully immersed in running buffer and that there was no leakage. Protein samples were briefly centrifuged before loading into the well. Between 2-20 µL of sample was loaded per well; loading was typically adjusted to a mean signal intensity after an initial run and total protein staining to maintain equal loading. A protein ladder was loaded to flank the protein samples on each side, which was pre-stained (Precision Plus Protein Dual Color Standards, Bio-Rad) when gels were intended for western blotting or unstained when the gel was to be subjected to Coomassie staining (Precision Plus Unstained Standards, Bio-Rad). The gels were then run at 180 V at room temperature for 60 minutes in most cases, or at lower voltage (120 V) until completion in some cases (to give better resolution).

6.5.3 Western blotting

Proteins separated by SDS-PAGE were then transferred to nitrocellulose membranes (Amersham Protran 0.2 NC nitrocellulose Western blotting membrane, GE Healthcare) in preparation for immunodetection. Western blotting followed either a ‘wet’ or ‘semi-wet’ transfer method. Resolved SDS-PAGE gels were overlaid with a 60 mm x 90 mm single layer of nitrocellulose membrane in ice cold transfer buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS, 20% (v/v) methanol). 2-3 layers of filter paper (70mm x 100 mm) were then assembled on either side of the gel and membrane. A roller was used to remove any bubbles that existed between the gel and membrane. For wet transfer, sponges were also added and the stack was assembled into a cassette, then placed inside a transfer tank (Mini Trans-Blot Module, Bio-Rad), filled with ice-cold transfer buffer. An ice pack was added to keep the temperature

cool during transfer, to reduce electrical resistance. Transfer occurred at 100 V for 60 minutes with stirring. During semi-wet transfer, the same stack was assembled without sponges. It was then loaded into a Transblot Turbo (Bio-Rad), and proteins were transferred at 1 mA, 25 V for 30 mins.

Following transfer, membranes were blocked with 5% (w/v) skimmed-milk in Tris-buffered-saline-Tween (TBST-T; 10 mM Tris-HCl pH 8, 150 mM NaCl, 0.05% (v/v) Tween 20)) – referred to as blocking solution. Blocking lasted for 1 hour at room temperature, with gentle shaking on an orbital platform (75 rpm).

Following blocking, molecular weight markers were used as guides to cut membranes into the minimum number of sections to incubate with desired antibodies corresponding to their predicted protein mass. Primary antibody dilutions were thawed on ice and incubated with membrane fractions overnight at 4°C, on a rocking platform set at 30 rpm.

Primary antibodies were diluted in 10-20 mL of blocking solution, supplemented with a small quantity (not measured) of thiomersal to inhibit bacterial and fungal contamination. Commercial antibodies were diluted at the manufacture's recommended ratio. Many native antibodies were homemade and the dilution ratios used were based on pre-determined signal:noise tests.

Following primary antibody incubation, membranes were washed three times with 15 ml of TBS-T (5 minutes, room temperature, 75 rpm on an orbital platform). The membranes were then incubated with a light chain specific (Peroxidase IgG Fraction Monoclonal Mouse Anti-Rabbit IgG; Jackson ImmunoResearch Europe Ltd.) or heavy chain specific (goat anti-rabbit IgG-HRP; Abcam) secondary antibody for 1 hour (room temperature, 75 rpm on an orbital platform). The secondary antibody was diluted in 15-20 mL Blotto. The membrane strips were then washed three times with 15 mL TBST-T (5 minutes, room temperature, 75 rpm on an orbital platform). Chemilumenescence was used to detect the secondaty antibody, and thus protein bands (EZ-Chemiluminescence Detection Kit for HRP, Geneflow), following the manufacter's instructions. A CCD camera equipped LAS-4000 imager (GE Healthcare) was

used to capture luminescence at a range of exposure times. Where quantified, gel image files were loaded into ImageJ software and converted into 8-bit files. The 'gels' function was then used to select lanes and intensity was quantified as an area.

6.5.4 Coomassie and ponceau staining

Following total protein extraction, the quality and quantity of the extracted proteins was assessed by total protein staining. Following SDS-PAGE, the gel was incubated in InstantBlue Ultrafast Protein Stain (Sigma) for 10-20 minutes on an orbital shaker (75 rpm) at room temperature. The Coomassie containing solution was then rinsed off with distilled water. In some cases, the quality of protein transfer to membrane was inspected by incubating the membrane with 1x ponceau-s (ThermoFisher) staining solution, under the same conditions for Coomassie staining. The proteins were then visualised with gradual washing of the membrane until bands became apparent. Both stained gels and membranes were photographed using either a scanner of the LAS 4000 (GE Healthcare).

6.6 Localisation and interaction experiments

6.6.1 Protoplast isolation

The protocol followed a modified version of the 'tape sandwich' isolation technique (Wu *et al.*, 2009). WT *Arabidopsis thaliana* Columbia were grown in long day (16 hrs light) conditions at 20 °C for 4-6 weeks (only pre-bolting plants were used). Rosette leaves were cut, and the adaxial surface affixed to a strip of PVC tape. A layer of Magic tape was then placed on top and gentle pressure applied over the leaves. The midrib was flattened by striking with forceps. Magic tape was removed - peeling away the lower epidermal surface. The PVC tape with affixed leaves was placed in a petri dish (with foil covered lid) containing enzyme solution (400 mM D-mannitol, 10 mM $CaCl_2$, 20 mM MES monohydrate, 20 mM KCl, 0.1% (w/v) BSA, 1 % (w/v) cellulase R-10 (Duchefa), 0.25% (w/v) macerozyme (Duchefa); pH 5.6 with KOH) for 2 hrs on an orbital shaker (Stuart), at <30 RPM. One 90 mm petri dish

was filled with strips of tape, which was sufficient to yield approximately 1.2 million isolated cells.

All centrifugation steps were carried out in a Sorvall Legend RT centrifuge fitted with 6441 swinging buckets; and at 4°C during protoplast isolation. The digested solution was centrifuged for 5 minutes at 60 x *g*. Care was taken to slowly remove the supernatant with a pipette. Protoplasts were washed twice with 25 ml prechilled W5 solution (154 mM NaCl, 125 mM $CaCl_2.6H_2O$, 5 mM KCl, 5 mM D-glucose, 1.5 mM MES monohydrate; pH 5.6 with KOH). Cells were only centrifuged after the first wash, at 40 x *g* for 5 minutes. Isolated cells were counted after the resuspension in the second volume of W5 solution and counted using a haemocytometer and phase contrast microscope (Nikon). The cells were then incubated on ice for 45 mins. W5 was then gently removed and the cells were resuspended in MMG solution (400 mM D-mannitol, 15 mM $MgCl_2.6H_2O$, 5 mM MES monohydrate; pH 5.6 with KOH).

6.6.2 Protoplast transfection

At all times of pipetting during transfection, pipette tips were cut to increase their aperture. The quantity of cells and plasmid required was dependant on the experiment: for microscopy, 5 µg of plasmid DNA and $1x10^5$ cells were used; whereas for protein detection experiments 30 µg of plasmid DNA and $6x10^5$ cells were typically used. Plasmid DNA concentrations were at a minimum of 1 µg/µl; the quantity of plasmid was often varied based on preliminary results to increase or reduce the level of signal. Plasmid and protoplast cells were gently mixed with a pipette, and briefly incubated on ice. A 1:1 volume of room temperature poly ethylene glycol (PEG) solution (400 mM D-mannitol, 200 mM $Ca(NO_3)_2$, 40% (w/v) PEG 4000 (average molecular mass)) was added to protoplasts as they were removed from ice, and the cells were gently mixed by inversion. Serial dilutions were then carried out with 1:1, 2:1 and 4:1 volumes of W5 relative to the starting volume of PEG and protoplasts – tubes were gently inverted and incubated for 5 mins at RT in between each addition. Protoplasts were then pelleted at 40 x *g* for 3 mins, then washed once with W5

solution. Protoplasts were resuspended in of W5 and placed in a tissue culture plate in the dark overnight at room temperature.

6.6.3 Localisation, BiFC experiments, and microscopy

The CDS of each Npl4 and Ufd1 isoform, including Ufd1B, was cloned into BiFC and YFP (transient and stably expressing vectors, as described previously. For BiFC, protoplasts were transfected with Npl4 and Ufd1 constructs encoding either the N-terminal or C-terminal of YFP were co-transfected with constructs expressing the complementary YFP terminus fused to potential interacting proteins and controls. For transient subcellular localisation, Npl4 and Ufd1 isoforms were introduced into overexpressing vectors which introduced a C-terminal translational fusion with full length YFP (specifically, eYFP: Ex λ 513, Em λ 527). For testing co-localisation with $CDC48^{E581Q}$-mScarlet (Ex λ 569, Em λ 594), protoplasts were co-transfected with Npl4 and Ufd1 constructs with YFP fusions.

Protoplast transfection followed as previously described with $1x10^6$ cells and 5 µg of each plasmid. Following protoplast isolation and transfection, protoplasts were concentrated by centrifugation at 100 x *g* for 2 mins. 90% of the supernatant was removed and cells were resuspended in the remaining solution (~50 µl). Approximately 20 µl was then pipetted onto a glass slide with a cover slip gently attached on top. A chamber was created by placing a two layers of PVC tape on the microscope slide, this helped ensure the viability of the sample.

Stably transformed plants were used for *in planta* localisation from at least the T2 generation. A small section of rosette leaf was abscised and placed on a glass cover slide. Perfluorodecalin was then pipetted on top of the sample, as it can efficiently enter air gaps in mesophyll tissue and improve the optical qualities of leaf tissue (Littlejohn *et al.*, 2010). A cover slide was then affixed on top in preparation for imaging.

A Leica TCS SP5 or Zeiss LSM 880 Airyscan confocal microscope was used for fluorescence imaging. Photomultiplier tubes were set to 3 independent detection channels for eYFP (518 nm – 540 nm), chlorophyll fluorescence (650 nm – 750 nm; as a marker for chloroplasts), and brightfield; mScarlet was detected on a fourth channel where necessary with a detection window of 575 nm – 600 nm. Gain was adjusted to the point where there was no signal saturation. Images were taken with a 63x or 40x water immersed objective as a single z slice. A clear fluorescent signal was interpreted to mean interaction or localisation of proteins. The images were analysed via use of the Leica LAS X or Zeiss Zen Blue software.

6.6.4 Co-Immunoprecipitation

The CDS of Npl4 and Ufd1 isoforms were cloned into C-terminally tagging YFP and Myc transient overexpression vectors (as detailed in previous sections). Constructs were transfected into protoplasts using the larger scale volumes (30 μg plasmid and $6x10^5$ cells), as previously described. Each test of interaction typically featured a co-transfection with a protein target of interest (e.g. HA-Cdc48). Quantities of plasmid were occasionally adjusted based on preliminary data from western blotting to attempt to increase or reduce the level of expression. In the case of Sp2-Myc, and the *in vivo* Ufd1C ubiquitination experiment, protoplasts were isolated from transgenic lines. A control transfection was typically used to test for non-specific interactions with the epitope tag used to immunoprecipitate proteins. In most cases, Npl4 and Ufd1 were the 'bait' proteins that were immunoprecipitated. Replicates would on occasion alter the orientation of the pulldown based on the intensity of the blotted protein (lower abundant proteins by immunoblotting were favoured for immunoprecipitation.

As many of the proteins involved in this study are substrates of the ubiquitin proteasome system, protoplasts were typically treated with bortezomib. Bortezomib is a 26S proteasome inhibitor which selectively binds preventing the degradation of ubiquitin tagged proteins (Curran & McKeage, 2009). 10 mM bortezomib (Selleckchem) was added to each protoplast

culture 3-4 hours before harvesting, to a final concentration of 5 μM (Ling *et al.*, 2019). Protoplasts were gently swayed to mix, then placed in the dark again.

Protoplast cells were pelleted in two rounds of centrifugation at 50 x *g* for 2 mins in a 2 ml Eppendorf tube. Supernatant was removed, and the cells were lysed on ice with 350 μL of freshly prepared lysis buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 1 mM EDTA, 1% (v/v) Triton X-100, 5 μM bortezomib, 0.5% (v/v) PIC) and rotated for 30 minutes at 4°C. The lysate was then clarified by centrifugation (14,000 rpm, 10 minutes, 4°C). A 30 μl sample was then removed and retained as a total lysate control.

Immunoconjugated beads (either magnetic-agarose or agarose) were used to immunoprecipitate proteins based on their epitope tag (GFP, Chromotek; Myc, Sigma; HA, Sigma; FLAG, Sigma). The beads were washed with lysis buffer several times before use, to equilibrate the buffer with that of the samples. At each stage of removal of supernatant from beads, either magnetic separation using a MagJET Separation Rack (Thermo Fisher) or brief centrifugation (8,200 x *g* for 30 seconds) was employed. Beads were kept at 4°C during all stages of washing. The clarified lysate was then added to the beads and they were allowed to rotate for 1-2 hours at 4°C.

A washing buffer was prepared fresh (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 1 mM EDTA, 0.5% (v/v) Triton X-100). The beads were separated from the solution and washed with the addition of 500 μl of washing buffer. The beads were then rotated at 4°C for 5 minutes. In total, the beads were washed between 3-5 times to remove non-specifically bound proteins. Proteins were then eluted from the immunoconjugated beads by incubation at 95 °C for 10 minutes in 60 μl of 2× Laemmli buffer with freshly added 0.1 M DTT. Total lysates were diluted with a 1:1 addition of × Laemmli buffer with freshly added 0.1 M DTT. Samples were then run through SDS-PAGE and western blotting to specifically probe for

proteins. The presence of proteins bands in the immunoprecipitated samples was taken to mean a physical interaction between a set of proteins.

6.7 Physiological and stress experiments

6.7.1 Chlorophyll Content Measurements

Measurements of chlorophyll content was taken using mature rosette leaves from 28-day old plants. A Konica-Minolta SPAD-502 meter was used to take measurements from at least 6 leaves from different plants of the same genotype; 3 readings were taken for each measured leaf and the average was recorded. The SPAD meter is a portable spectrophotometer, which records raw absorbance data. Chlorophyll concentrations (nmol/mg fresh weight) were generated by conversion using published equations (Huang *et al.*, 2011).

6.7.2 Chlorophyll Fluorescence and Senescence Analysis

Knock-out and overexpressing lines of Npl4 and Ufd1 lines were grown on soil until 28 days old. For induction of senescence, individual leaves from 6 different plants were covered with aluminium foil to prevent access to light. Leaves remained attached, so senescence would be induced only locally in the covered tissue, as previously reported (Schelbert *et al.*, 2009). Plants were then allowed to grow normally for 5 days.

Senescent induced leaves were then excised, uncovered and placed abaxial surface down on solid MS media. At the same time, uncovered leaves from the same senescent treated plants, were taken as controls for chlorophyll fluorescence; likewise, 6 leaves were placed on solid MS media. Leaves on plates were dark treated for at least 20 minutes prior to chlorophyll fluorescence analysis.

Plates containing the leaves were placed inside the imaging chamber of a PAM fluorimeter (MAXI IMAGING-PAM M Series; Walz). The focal point was adjusted to the correct plane, and background fluorescence subtracted with a blank sheet of paper. Measurement of chlorophyll fluorescent parameters occurred using the manufacture's default predefined light levels in a

light induction curve. Photosynthetically active radiation (PAR) pulses ranged from 0-700 μmol m^{-2} s^{-1}, at irregular increases over 240 seconds. The measures Fo (dark fluorescence yield), F (fluorescence yield), Fm (maximal fluorescence yield), and Fm' (maximal fluorescence yield, following photosystem saturation) were recorded by the device. The measures were used to calculate the parameters: Fv/Fm (maximum photosystem (PS) II quantum yield; Fv/Fm = (Fm - Fo)/Fm), Y(II) (effective PSII quantum yield; Y(II) = (Fm'-F)/Fm'), NPQ (non-photochemical quenching; NPQ = (Fm-Fm')/Fm'), and ETR (electron transport rate; ETR = 0.5 x Yield x PAR x 0.84 μequivalents m^{-2} s^{-1}). Data was saved as XPIM files and imported into the SPSS (IBM) statistical analysis software, for statistical testing and data presentation.

6.7.3 De-etiolation

De-etiolation was assessed based on a reported method (Ling *et al.*, 2012). Knock-out and overexpressing lines of Npl4 and Ufd1 lines were scatter sown on solid MS plates, without sucrose. After stratification at 4°C for 3 days, plates were exposed to light for two hours to initiate germination. The plates were then wrapped in two layers of aluminium foil, to create a dark environment, and placed in a Percival with standard environmental conditions for 5 days. Half of the genotypes on plates were then removed from the growth chamber and tissue was harvested for protein extraction, as previously described. The remaining plates were exposed to standard light levels for a continuous period of 24 hours. Tissue was then harvested from those plants, and protein similarly extracted. SDS-PAGE and western blotting was performed on the extracted proteins. Membranes were probed with TOC specific antibodies, and control antibodies (H3 or TIC). Band intensity of the blots was quantified using ImageJ and analysed statistically with SPSS.

6.7.4 Salt stress

Tolerance of salt stress was examined based on a previous report into the function of the CHLORAD system (Ling & Jarvis, 2015b). Knock-out and overexpressing lines of Npl4 and Ufd1 lines were scatter sown on solid MS plates, with 1% (w/v) sucrose – otherwise

germination would be too severely impacted (Ling & Jarvis, 2015b). NaCl was added to the media prior to autoclaving, at a concentration of 150 mM. Plants were allowed to grow for 14 days on the media, at which point they were photographed. Plants were counted for germination efficiency and for development; development was assessed by counting the proportion of plants which developed expanded cotyledons and were not pale.

6.7.5 Mannitol Stress

Tolerance of mannitol (osmotic) stress was examined based on a previous report into the function of the CHLORAD system (Ling & Jarvis, 2015b). Knock-out and overexpressing lines of Npl4 and Ufd1 lines were scatter sown on solid MS plates fitted with nylon mesh. At 10 days old, the mesh embedded with plants was gently transferred (in a sterile environment) to new solid MS plates, either with or without 400 mM D-mannitol. Plates were then placed back into a Percival under normal growth conditions for a further 3 days. Tissue was then harvested for protein extraction as previously described. SDS-PAGE and western blotting was performed to examine the abundance of TOC proteins, alongside controls. Blots were then quantified using ImageJ and statistical analysis was performed using SPSS.

Primer Name	Sequence 5' to 3'	Purpose
UFD1-B-BglII-F	AAAGATCTATGGATGGAGATGGTTCTAGC	Restriction cloning
UFD1-BSalI-nsR	AAGTCGACAACCATTGAGTGAGTATTTCTTCCC	Restriction cloning
UFD1-C-BglII-F	AAAGATCTATGGATTTCGAGCTTAGATCAGC	Restriction cloning
UFD1-C-SalI-nsR	AAGTCGACAGCTACTCTTGCCATGAACAGC	Restriction cloning
UFD1-D-BglII-F	AAAGATCTATGTTTTACGATGGATACGCTTATC	Restriction cloning
UFD1-D-SalI-nsR	AAGTCGACAACCCCTCAATGAATATTTTTTACCAC	Restriction cloning
NPL4-BglII-F	AAAGATCTATGATGATGCTCAGAATCCGAAG	Restriction cloning
NPL4-SalI-nsR	AAGTCGACAACAAGTGTTGGCCATTGATTC	Restriction cloning
NPL4L-EcoRI-F	AAGAATTCATGACGATGCTCAGAGTCCG	Restriction cloning
NPL4L-A-SalI-R	AAGTCGACAAGAAGTATTGGCCATGGAGTCG	Restriction cloning
NPL4_R_attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCGTAACAAGTGTTGGCCATTGATTC	Gateway cloning
NPL4_F_attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGATGATGCTCAGAATCCGAAG	Gateway cloning
NPL4L_R_attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCGTAAGAAGTATTGGCCATGGAGTCG	Gateway cloning
NPL4L_F_attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGACGATGCTCAGAGTCCGAA	Gateway cloning
UFD1-2_R_attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCGTAACCATTGAGTGAGTATTTCTTCCC	Gateway cloning
UFD1-2_F_attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGATGGAGATGGTTCTAGC	Gateway cloning
UFD1-3_R_attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCGCAGCTACTCTTGCCATGAACAGC	Gateway cloning
UFD1-3_F_attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGATTTCGAGCTTAGATCAGCGA	Gateway cloning
UFD1-4_R_attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCGCAACCCCTCAATGAATATTTTTTACCAC	Gateway cloning
UFD1-4_F_attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTTTTACGATGGATACGCTTATC	Gateway cloning
UFD1A.1_SalI_F	AAAGATCTATGTTTTTCGATGGATACCATTATC	Restriction cloning
UFD1A.1_BglII_R	AAGTCGACAACCCCTCAATGAATACTTC	Restriction cloning
UFD1A.1attB1_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTTTTTCGATGGATACCATTATC	Gateway
UFD1A.1attB2_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCACAACCCCTCAATGAATACTTC	Gateway
UFD1A_BamHI_F	aaggatccATGTTTTTCGATGGATACCATTATC	Restriction cloning
UFD1A_NotI_R	aagcggccgcgtaaCAACCCCTCAATGAATACTTC	Restriction cloning
UFD1B_BamHI_F	aaggatccATGGATGGAGATGGTTCTAGC	Restriction cloning

UFD1B_NotI_R	aagcggccgcgtagTAACCATTGAGTGAGTATTTCTTCCC	Restriction cloning
UFD1C_BamHI_F	aaggatccATGGATTTCGAGCTTAGATCAGCGA	Restriction cloning
UFD1C_NotI_R	aagcggccgcgtagCAGCTACTCTTGCCATGAACAGC	Restriction cloning
UFD1D_BamHI_F	aaggatccATGTTTTACGATGGATACGCTTATC	Restriction cloning
UFD1D_NotI_R	aagcggccgcgtagCAACCCCTCAATGAATATTTTTTACCAC	Restriction cloning
NPL4_NotI_R	aagcggccgcgTAACAAGTGTTGGCCATTGATTC	Restriction cloning
NPL4L_BamHI_F	aaggatccATGACGATGCTCAGAGTCCGAA	Restriction cloning
NPL4L_NotI_R	aagcggccgcgtagTAAGAAGTATTGGCCATGGAGTCG	Restriction cloning
NPL4_SalI_F	aagtcgacATGATGATGCTCAGAATCCGAAG	Restriction cloning
npl4-1 LP	GATGATGCTCAGAATCCGAAG	T-DNA line genotyping
npl4-1 RP/ npl4-2 LP	GGAAAGCAAACACTGACAAGC	T-DNA line genotyping
npl4l-1 LP	TCAACCTGACCATCCTCAGAG	T-DNA line genotyping
npl4l-2 LP	GAGATTCGGAGGTTAGGATCG	T-DNA line genotyping
npl4l-3 LP	TAGGATCAGAACATTGGCGAC	T-DNA line genotyping
ufd1A-1 LP	GAGGGCTCAAATACAAGAGGG	T-DNA line genotyping
ufd1C-2 LP	ATGGTGCAGTTCTTGAACCAC	T-DNA line genotyping
ufd1D-2 LP	GTTCCACACATTGGTTGGTTC	T-DNA line genotyping
npl4-2 RP	ATCAGCTTCAGATCCCTCTCC	T-DNA line genotyping
npl4l-1 RP/npl4l-2 RP	AAGTGGCCAGAGAGGAGAGAC	T-DNA line genotyping
npl4l-3 RP	ctgtccgctcttaatttgtcg	T-DNA line genotyping
ufd1A-1 RP	TGCTGAGATTCGAAATTGAGG	T-DNA line genotyping
ufd1C-2 RP	AATTCGGTATATCCGATTGCC	T-DNA line genotyping
ufd1D-2 RP	GAGTCCTTGAGTTCATCGCAG	T-DNA line genotyping
N4_gRNA1_	attgGGTGAAGGCGAGTAGATCAG	CRISPR gRNA
N4_gRNA2_	attgGCTGAAGGAGTGGGTCACGG	CRISPR gRNA
N4L_gRNA1_	attgGCTGGGTTTTTGAACGAACG	CRISPR gRNA
N4L_gRNA2_	attgGATTCTGATGAGAGACTCTG	CRISPR gRNA
U1A_gRNA1_	attgGGGATTTTGCAAGTTGTAGA	CRISPR gRNA
U1A_gRNA2_	attgGAGCATACCTAGACGATCAA	CRISPR gRNA
U1C_gRNA1_	attgGATTCCTCGTAAAATTGTGT	CRISPR gRNA
U1C_gRNA2_	attgGAGTGGTCTTCACACGACGT	CRISPR gRNA
U1D_gRNA1_	attgGTACTCGTTTAGGAATCGAA	CRISPR gRNA
U1D_gRNA2_	attgGTGGGGTTGCAGTTTCACGT	CRISPR gRNA

Npl4_sqRT_F	AAACCTCCCCCTCTCCTCTC	sqRTPCR
Npl4_sqRT_R	GACAGAGGCCCCTGGTGAT	sqRTPCR
Npl4L_sqRT_F	TCCTCGCTAAGAGCCCTTCT	sqRTPCR
Npl4L_sqRT_R	AAGGTACACGACAACGGTCC	sqRTPCR
Ufd1A_sqRT_F	GCCCTTGATCGTCTAGCCTC	sqRTPCR
Ufd1A_sqRT_R	CTTTTGGAGCACGGTTTCCG	sqRTPCR
Ufd1C_sqRT_F	TGCGATTAAGCAACGCGAAG	sqRTPCR
Ufd1C_sqRT_R	GTCGTGTGAAGACCACTCGT	sqRTPCR
Ufd1D_sqRT_F	CTCACTGCGGAGTCCTTGAG	sqRTPCR
Ufd1D_sqRT_R	CGTAGCCTTCTCTGACCCAC	sqRTPCR
ACT2_F	TGAGAGATTCAGATGCCCAGAA	sqRTPCR
ACT2_R	TGGATTCCAGCAGCTTCCAT	sqRTPCR

Table 6.1 Primer sequences.

Construct	Vector	Purpose
Npl4-YFP	p2GWY7: Karimi *et al.* 2002	Transient expression of protein fusion with C-terminal YFP tag, for protoplast transfection with p35s promoter. For use in Co-IPs and subcellular localisation experiments.
Npl4L-YFP		
Ufd1A-YFP		
Ufd1C-YFP		
Ufd1D-YFP		
Npl4-YFP	pK7YWG2: Karimi *et al.* 2002	Binary vector for overexpression of a protein in fusion with C-terminal YFP tag, for stable transgenic line generation with p35s promoter. For use subcellular localisation experiments and experiments in Chapter IV.
Npl4L-YFP		
Ufd1A-YFP		
Ufd1C-YFP		
Ufd1D-YFP		
Npl4-Myc	p2GW7: Karimi *et al*. 2002	Transient expression of protein fusion with C-terminal 6xMyc tag, for protoplast transfection with p35s promoter. For use in Co-IPs.
Npl4L-Myc		
Ufd1A-Myc		
Ufd1C-Myc		
Ufd1D-Myc		
Npl4-cYFP	pSAT4A-cEYFP-N1: (Tzfira *et al.*, 2005	Transient expression of protein fusion with C-terminal 6xMyc tag, for protoplast transfection with p35s promoter. For use in BiFC experiments.
Npl4L-cYFP		
Ufd1A-cYFP		
Ufd1C-cYFP		
Ufd1D-cYFP		
Npl4-CTAP	CTAPI: Rohila *et al.* 2004	Binary vector for overexpression of a protein in fusion with C-terminal TAP tag, for stable transgenic line generation with p35s promoter. For use in experiments in Chapter IV.
Npl4L-CTAP		
Ufd1A-CTAP		

Ufd1C-CTAP		
Ufd1D-CTAP		
NTAP-Npl4	NTAPI: Rohila *et al*. 2004	Binary vector for overexpression of a protein in fusion with N-terminal TAP tag, for stable transgenic line generation with p35s promoter. Not used in experiments.
NTAP-Npl4L		
NTAP-Ufd1A		
NTAP-Ufd1C		
NTAP-Ufd1D		
Npl4 gRNA1	pKI1.1R: Tsutsui & Higashiyama, 2017	Binary vector for the expression of the CRISPR-Cas9 cassette for genome editing, expressing a single gRNA in each construct. Not used in experiments.
Npl4 gRNA2		
Npl4L gRNA1		
Npl4L gRNA2		
Ufd1A gRNA1		
Ufd1A gRNA2		
Ufd1C gRNA1		
Ufd1C gRNA2		
Ufd1D gRNA1		
Ufd1D gRNA2		

Table 6.1 List of constructs generated in this thesis. The construct's name and the final vector it was cloned into are detailed, as well as the final experimental purpose of each. Omitted are intermediate cloning vectors, including the Gateway entry vector pDONR201 and pE3C (to introduce a 6xMyc tag).

Bibliography

Agne, B., Andrès, C., Montandon, C., Christ, B., Ertan, A., Jung, F., Infanger, S., Bischof, S., Baginsky, S., & Kessler, F. (2010). The acidic A-domain of *Arabidopsis* TOC159 occurs as a hyperphosphorylated protein. *Plant Physiol.*, *153*(3), 1016-1030. https://doi.org/10.1104/pp.110.158048

Agne, B., Infanger, S., Wang, F., Hofstetter, V., Rahim, G., Martin, M., Lee, D. W., Hwang, I., Schnell, D., & Kessler, F. (2009). A *toc159* import receptor mutant, defective in hydrolysis of GTP, supports preprotein import into chloroplasts. *J. Biol. Chem.*, *284*(13), 8670-8679. https://doi.org/10.1074/jbc.M804235200

Agne, B., Köhler, D., & Baginsky, S. (2017). Protein import-independent functions of Tic56, a component of the 1-MDa translocase at the inner chloroplast envelope membrane. *Plant Signal. Behav.*, *12*(3), e1284726. https://doi.org/10.1080/15592324.2017.1284726

Akita, M., Nielsen, E., & Keegstra, K. (1997). Identification of protein transport complexes in the chloroplastic envelope membranes via chemical cross-linking. *J. Cell Biol.*, *136*(5), 983-994. https://doi.org/10.1083/jcb.136.5.983

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91-93. https://doi.org/10.1016/j.tjem.2018.08.001

Allawi, H. T., & SantaLucia, J., Jr. (1997). Thermodynamics and NMR of internal G.T mismatches in DNA. *Biochemistry*, *36*(34), 10581-10594. https://doi.org/10.1021/bi962590c

Allen, J. F. (2015). Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proc Natl Acad Sci U S A*, *112*(33), 10231-10238. https://doi.org/10.1073/pnas.1500012112

Allen, M. D., Buchberger, A., & Bycroft, M. (2006). The PUB domain functions as a p97 binding module in human peptide N-glycanase. *J Biol Chem*, *281*(35), 25502-25508. https://doi.org/10.1074/jbc.M601173200

Andrès, C., Agne, B., & Kessler, F. (2010). The TOC complex: preprotein gateway to the chloroplast. *Biochim. Biophys. Acta*, *1803*(6), 715-723. https://doi.org/10.1016/j.bbamcr.2010.03.004

Aronsson, H., Boij, P., Patel, R., Wardle, A., Topel, M., & Jarvis, P. (2007). Toc64/OEP64 is not essential for the efficient import of proteins into chloroplasts in *Arabidopsis thaliana*. *Plant J.*, *52*(1), 53-68. https://doi.org/10.1111/j.1365-313X.2007.03207.x

Aronsson, H., Combe, J., Patel, R., Agne, B., Martin, M., Kessler, F., & Jarvis, P. (2010). Nucleotide binding and dimerization at the chloroplast pre-protein import receptor, atToc33, are not essential in vivo but do increase import efficiency. *Plant J.*, *63*(2), 297-311. https://doi.org/10.1111/j.1365-313X.2010.04242.x

Barbrook, A. C., Howe, C. J., & Purton, S. (2006). Why are plastid genomes retained in non-photosynthetic organisms? *Trends Plant Sci*, *11*(2), 101-108. https://doi.org/10.1016/j.tplants.2005.12.004

Barthelme, D., & Sauer, R. T. (2016). Origin and Functional Evolution of the Cdc48/p97/VCP AAA+ Protein Unfolding and Remodeling Machine. *J Mol Biol*, *428*(9 Pt B), 1861-1869. https://doi.org/10.1016/j.jmb.2015.11.015

Baslam, M., Oikawa, K., Kitajima-Koga, A., Kaneko, K., & Mitsui, T. (2016). Golgi-to-plastid trafficking of proteins through secretory pathway: Insights into vesicle-mediated import toward the plastids. *Plant Signal. Behav.*, *11*(9), e1221558. https://doi.org/10.1080/15592324.2016.1221558

Bedard, J., Kubis, S., Bimanadham, S., & Jarvis, P. (2007). Functional similarity between the chloroplast translocon component, Tic40, and the human co-chaperone, Hsp70-interacting protein (Hip). *J Biol Chem*, *282*(29), 21404-21414. https://doi.org/10.1074/jbc.M611545200

Bègue, H., Jeandroz, S., Blanchard, C., Wendehenne, D., & Rosnoblet, C. (2017). Structure and functions of the chaperone-like p97/CDC48 in plants. *Biochim. Biophys. Acta*, *1861*(1 Pt A), 3053-3060. https://doi.org/10.1016/j.bbagen.2016.10.001

Bischof, S., Baerenfaller, K., Wildhaber, T., Troesch, R., Vidi, P. A., Roschitzki, B., Hirsch-Hoffmann, M., Hennig, L., Kessler, F., Gruissem, W., & Baginsky, S. (2011). Plastid proteome assembly without Toc159: photosynthetic protein import and accumulation of N-acetylated plastid precursor proteins. *Plant Cell*, *23*(11), 3911-3928. https://doi.org/10.1105/tpc.111.092882

Bodnar, N., & Rapoport, T. (2017). Toward an understanding of the Cdc48/p97 ATPase. *F1000Res*, *6*, 1318. https://doi.org/10.12688/f1000research.11683.1

Bodnar, N. O., Kim, K. H., Ji, Z., Wales, T. E., Svetlov, V., Nudler, E., Engen, J. R., Walz, T., & Rapoport, T. A. (2018). Structure of the Cdc48 ATPase with its ubiquitin-binding cofactor Ufd1-Npl4. *Nat Struct Mol Biol*, *25*(7), 616-622. https://doi.org/10.1038/s41594-018-0085-x

Bodnar, N. O., & Rapoport, T. A. (2017). Molecular mechanism of substrate processing by the Cdc48 ATPase complex. *Cell*, *169*(4), 722-735. https://doi.org/10.1016/j.cell.2017.04.020

Bolter, B., May, T., & Soll, J. (1998). A protein import receptor in pea chloroplasts, Toc86, is only a proteolytic fragment of a larger polypeptide. *FEBS Lett*, *441*(1), 59-62. https://doi.org/10.1016/s0014-5793(98)01525-7

Bölter, B., & Soll, J. (2017). Ycf1/Tic214 is not essential for the accumulation of plastid proteins. *Mol. Plant*, *10*(1), 219-221. https://doi.org/10.1016/j.molp.2016.10.012

Borner, T., Aleynikova, A. Y., Zubo, Y. O., & Kusnetsov, V. V. (2015). Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochim Biophys Acta*, *1847*(9), 761-769. https://doi.org/10.1016/j.bbabio.2015.02.004

Bouchnak, I., Brugiere, S., Moyet, L., Le Gall, S., Salvi, D., Kuntz, M., Tardif, M., & Rolland, N. (2019). Unraveling Hidden Components of the Chloroplast Envelope Proteome: Opportunities and Limits of Better MS Sensitivity. *Mol Cell Proteomics*, *18*(7), 1285-1306. https://doi.org/10.1074/mcp.RA118.000988

Braun, S., Matuschewski, K., Rape, M., Thoms, S., & Jentsch, S. (2002). Role of the ubiquitin-selective CDC48(UFD1/NPL4)chaperone (segregase) in ERAD of OLE1 and other substrates. *EMBO J*, *21*(4), 615-621. https://doi.org/10.1093/emboj/21.4.615

Bräutigam, K., Dietzel, L., Kleine, T., Stroher, E., Wormuth, D., Dietz, K. J., Radke, D., Wirtz, M., Hell, R., Dormann, P., Nunes-Nesi, A., Schauer, N., Fernie, A. R., Oliver, S. N., Geigenberger, P., Leister, D., & Pfannschmidt, T. (2009). Dynamic plastid redox signals integrate gene expression and metabolism to induce distinct metabolic states in photosynthetic acclimation in *Arabidopsis*. *Plant Cell*, *21*(9), 2715-2732. https://doi.org/10.1105/tpc.108.062018

Brinkman, E. K., Chen, T., Amendola, M., & van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*, *42*(22). https://doi.org/ARTN e168

10.1093/nar/gku936

Broad, W., Ling, Q., & Jarvis, P. (2016). New insights into roles of ubiquitin modification in regulating plastids and other endosymbiotic organelles. *Int. Rev. Cell Mol. Biol.*, *325*, 1-33. https://doi.org/10.1016/bs.ircmb.2016.02.007

Bruderer, R. M., Brasseur, C., & Meyer, H. H. (2004). The AAA ATPase p97/VCP interacts with its alternative co-factors, Ufd1-Npl4 and p47, through a common bipartite binding mechanism. *J Biol Chem*, *279*(48), 49609-49616. https://doi.org/10.1074/jbc.M408695200

Buchberger, A., Schindelin, H., & Hänzelmann, P. (2015). Control of p97 function by cofactor binding. *FEBS Lett.*, *589*(19 Pt A), 2578-2589. https://doi.org/10.1016/j.febslet.2015.08.028

Burki, F. (2014). The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb Perspect Biol*, *6*(5), a016147. https://doi.org/10.1101/cshperspect.a016147

Bustin, S. A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol*, *29*(1), 23-39. https://doi.org/10.1677/jme.0.0290023

Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P. E., Oberhauser, A. F., & Fernandez, J. M. (2003). The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol*, *10*(9), 738-743. https://doi.org/10.1038/nsb965

Celedon, J. M., & Cline, K. (2013). Intra-plastid protein trafficking: how plant cells adapted prokaryotic mechanisms to the eukaryotic condition. *Biochim. Biophys. Acta*, *1833*(2), 341-351. https://doi.org/10.1016/j.bbamcr.2012.06.028

Chen, Y. L., Chen, L. J., Chu, C. C., Huang, P. K., Wen, J. R., & Li, H. M. (2018). TIC236 links the outer and inner membrane translocons of the chloroplast. *Nature*, *564*(7734), 125-129. https://doi.org/10.1038/s41586-018-0713-y

Chou, M. L., Fitzpatrick, L. M., Tu, S. L., Budziszewski, G., Potter-Lewis, S., Akita, M., Levin, J. Z., Keegstra, K., & Li, H. M. (2003). Tic40, a membrane-anchored co-chaperone homolog in the chloroplast protein translocon. *EMBO J.*, *22*(12), 2970-2980. https://doi.org/10.1093/emboj/cdg281

Christians, M. J., Gingerich, D. J., Hansen, M., Binder, B. M., Kieber, J. J., & Vierstra, R. D. (2009). The BTB ubiquitin ligases ETO1, EOL1 and EOL2 act collectively to regulate ethylene biosynthesis in Arabidopsis by controlling type-2 ACC synthase levels. *Plant J.*, *57*(2), 332-345. https://doi.org/10.1111/j.1365-313X.2008.03693.x

Christianson, J. C., & Ye, Y. (2014). Cleaning up in the endoplasmic reticulum: ubiquitin in charge. *Nat Struct Mol Biol*, *21*(4), 325-335. https://doi.org/10.1038/nsmb.2793

Chu, C. C., & Li, H. M. (2018). Developmental regulation of protein import into plastids. *Photosynth. Res.*, *138*(3), 327-334. https://doi.org/10.1007/s11120-018-0546-4

Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J*, *16*(6), 735-743. https://doi.org/10.1046/j.1365-313x.1998.00343.x

Curran, M. P., & McKeage, K. (2009). Bortezomib A Review of its Use in Patients with Multiple Myeloma. *Drugs*, *69*(7), 859-888. https://doi.org/Doi 10.2165/00003495-200969070-00006

Dalal, S., Rosser, M. F., Cyr, D. M., & Hanson, P. I. (2004). Distinct roles for the AAA ATPases NSF and p97 in the secretory pathway. *Mol Biol Cell*, *15*(2), 637-648. https://doi.org/10.1091/mbc.e03-02-0097

Davies, J. M., Brunger, A. T., & Weis, W. I. (2008). Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. *Structure*, *16*(5), 715-726. https://doi.org/10.1016/j.str.2008.02.010

Day, P. M., Inoue, K., & Theg, S. M. (2019). Chloroplast outer membrane beta-barrel proteins use components of the general import apparatus. *Plant Cell*, *31*(8), 1845-1855. https://doi.org/10.1105/tpc.19.00001

Day, P. M., Potter, D., & Inoue, K. (2014). Evolution and targeting of Omp85 homologs in the chloroplast outer envelope membrane. *Front. Plant Sci.*, *5*, 535. https://doi.org/10.3389/fpls.2014.00535

de Michele, R., McFarlane, H. E., Parsons, H. T., Meents, M. J., Lao, J., Gonzalez Fernandez-Nino, S. M., Petzold, C. J., Frommer, W. B., Samuels, A. L., & Heazlewood, J. L. (2016). Free-Flow Electrophoresis of Plasma Membrane Vesicles Enriched by Two-Phase Partitioning Enhances the Quality of the Proteome from Arabidopsis Seedlings. *J Proteome Res*, *15*(3), 900-913. https://doi.org/10.1021/acs.jproteome.5b00876

Dowil, R. T., Lu, X., Saracco, S. A., Vierstra, R. D., & Downes, B. P. (2011). Arabidopsis membrane-anchored ubiquitin-fold (MUB) proteins localize a specific subset of ubiquitin-conjugating (E2) enzymes to the plasma membrane. *J Biol Chem*, *286*(17), 14913-14921. https://doi.org/10.1074/jbc.M110.158808

Dubin, M. J., Bowler, C., & Benvenuto, G. (2008). A modified Gateway cloning strategy for overexpressing tagged proteins in plants. *Plant Methods*, *4*, 3. https://doi.org/10.1186/1746-4811-4-3

Duy, D., Wanner, G., Meda, A. R., von Wiren, N., Soll, J., & Philippar, K. (2007). PIC1, an ancient permease in Arabidopsis chloroplasts, mediates iron transport. *Plant Cell*, *19*(3), 986-1006. https://doi.org/10.1105/tpc.106.047407

Emanuelsson, O., Brunak, S., von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, *2*(4), 953-971. https://doi.org/10.1038/nprot.2007.131

Esaki, M., & Ogura, T. (2012). Cdc48p/p97-mediated regulation of mitochondrial morphology is Vms1p-independent. *J Struct Biol*, *179*(2), 112-120. https://doi.org/10.1016/j.jsb.2012.04.017

Fang, C., Zhang, X., Zhang, L., Gao, X., Yang, P., & Lu, H. (2016). Identification of Palmitoylated Transitional Endoplasmic Reticulum ATPase by Proteomic Technique and Pan Antipalmitoylation Antibody. *J Proteome Res*, *15*(3), 956-962. https://doi.org/10.1021/acs.jproteome.5b00979

Fellerer, C., Schweiger, R., Schöngruber, K., Soll, J., & Schwenkert, S. (2011). Cytosolic HSP90 cochaperones HOP and FKBP interact with freshly synthesized chloroplast preproteins of *Arabidopsis*. *Mol. Plant*, *4*(6), 1133-1145. https://doi.org/10.1093/mp/ssr037

Flores-Pérez, U., Bédard, J., Tanabe, N., Lymperopoulos, P., Clarke, A. K., & Jarvis, P. (2016). Functional analysis of the Hsp93/ClpC chaperone at the chloroplast envelope. *Plant Physiol.*, *170*(1), 147-162. https://doi.org/10.1104/pp.15.01538

Flores-Pérez, U., & Jarvis, P. (2013). Molecular chaperone involvement in chloroplast protein import. *Biochim. Biophys. Acta*, *1833*(2), 332-340. https://doi.org/10.1016/j.bbamcr.2012.03.019

Franz, A., Ackermann, L., & Hoppe, T. (2016). Ring of Change: CDC48/p97 Drives Protein Dynamics at Chromatin. *Front Genet*, *7*, 73. https://doi.org/10.3389/fgene.2016.00073

Frydman, J., & Hohfeld, J. (1997). Chaperones get in touch: the Hip-Hop connection. *Trends Biochem Sci*, *22*(3), 87-92. https://doi.org/10.1016/s0968-0004(97)01005-0

Fu, Y., Li, X., Fan, B., Zhu, C., & Chen, Z. (2022). Chloroplasts Protein Quality Control and Turnover: A Multitude of Mechanisms. *Int J Mol Sci*, *23*(14). https://doi.org/10.3390/ijms23147760

Fucile, G., Di Biase, D., Nahal, H., La, G., Khodabandeh, S., Chen, Y., Easley, K., Christendat, D., Kelley, L., & Provart, N. J. (2011). ePlant and the 3D data display initiative: integrative systems biology on the world wide web. *PLoS One*, *6*(1), e15237. https://doi.org/10.1371/journal.pone.0015237

Galvão, R. M., Kota, U., Soderblom, E. J., Goshe, M. B., & Boss, W. F. (2008). Characterization of a new family of protein kinases from Arabidopsis containing phosphoinositide 3/4-kinase and ubiquitin-like domains. *Biochem J*, *409*(1), 117-127. https://doi.org/10.1042/BJ20070959

Ganesan, I., Shi, L. X., Labs, M., & Theg, S. M. (2018). Evaluating the functional pore size of chloroplast TOC and TIC protein translocons: import of folded proteins. *Plant Cell*, *30*(9), 2161-2173. https://doi.org/10.1105/tpc.18.00427

Ganesan, I., & Theg, S. M. (2019). Structural considerations of folded protein import through the chloroplast TOC/TIC translocons. *FEBS Lett.*, *593*(6), 565-572. https://doi.org/10.1002/1873-3468.13342

Gao, H., Li, F., Ji, Z., Shi, Z., Li, Y., & Yu, H. (2022). Cryo-EM structures of human p97 double hexamer capture potentiated ATPase-competent state. *Cell Discov*, *8*(1), 19. https://doi.org/10.1038/s41421-022-00379-1

Glaser, S., van Dooren, G. G., Agrawal, S., Brooks, C. F., McFadden, G. I., Striepen, B., & Higgins, M. K. (2012). Tic22 is an essential chaperone required for protein import into the apicoplast. *J. Biol. Chem.*, *287*(47), 39505-39512. https://doi.org/10.1074/jbc.M112.405100

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*, *40*(Database issue), D1178-1186. https://doi.org/10.1093/nar/gkr944

Goto, C., Hashizume, S., Fukao, Y., Hara-Nishimura, I., & Tamura, K. (2019). Comprehensive nuclear proteome of Arabidopsis obtained by sequential extraction. *Nucleus*, *10*(1), 81-92. https://doi.org/10.1080/19491034.2019.1603093

Gutensohn, M., Schulz, B., Nicolay, P., & Flügge, U. I. (2000). Functional analysis of the two *Arabidopsis* homologues of Toc34, a component of the chloroplast protein import apparatus. *Plant J.*, *23*(6), 771-783. https://doi.org/10.1046/j.1365-313x.2000.00849.x

Halperin, T., Ostersetzer, O., & Adam, Z. (2001). ATP-dependent association between subunits of Clp protease in pea chloroplasts. *Planta*, *213*(4), 614-619. https://doi.org/10.1007/s004250100527

Han, H., Fulcher, J. M., Dandey, V. P., Iwasa, J. H., Sundquist, W. I., Kay, M. S., Shen, P. S., & Hill, C. P. (2019). Structure of Vps4 with circular peptides and implications for translocation of two polypeptide chains by AAA+ ATPases. *Elife*, *8*. https://doi.org/10.7554/eLife.44071

Hänzelmann, P., & Schindelin, H. (2011). The structural and functional basis of the p97/valosin-containing protein (VCP)-interacting motif (VIM): mutually exclusive binding of cofactors to the N-terminal domain of p97. *J Biol Chem*, *286*(44), 38679-38690. https://doi.org/10.1074/jbc.M111.274506

Hänzelmann, P., & Schindelin, H. (2017). The interplay of cofactor interactions and post-translational modifications in the regulation of the AAA+ ATPase p97. *Front. Mol. Biosci.*, *4*, 21. https://doi.org/10.3389/fmolb.2017.00021

Hawes, C., Brandizzil, F., Batoko, H., & Moore, I. (2001). Organelle motility in plant cells: imaging golgi and ER dynamics with GFP. *Curr Protoc Cell Biol*, *Chapter 13*, Unit 13 13. https://doi.org/10.1002/0471143030.cb1303s09

Hedges, S. B. (2002). The origin and evolution of model organisms. *Nat Rev Genet*, *3*(11), 838-849. https://doi.org/10.1038/nrg929

Heo, J. M., Livnat-Levanon, N., Taylor, E. B., Jones, K. T., Dephoure, N., Ring, J., Xie, J., Brodsky, J. L., Madeo, F., Gygi, S. P., Ashrafi, K., Glickman, M. H., & Rutter, J. (2010). A stress-responsive system for mitochondrial protein degradation. *Mol Cell*, *40*(3), 465-480. https://doi.org/10.1016/j.molcel.2010.10.021

Herrmann, J. M. (2018). A force-generating machine in the plant's powerhouse: a pulling AAA-ATPase motor drives protein translocation into chloroplasts. *Plant Cell*, *30*(11), 2646-2647. https://doi.org/10.1105/tpc.18.00751

Hirosawa, Y., Ito-Inaba, Y., & Inaba, T. (2017). Ubiquitin-proteasome-dependent regulation of bidirectional communication between plastids and the nucleus. *Front. Plant Sci.*, *8*, 310. https://doi.org/10.3389/fpls.2017.00310

Hirsch, C., Gauss, R., Horn, S. C., Neuber, O., & Sommer, T. (2009). The ubiquitylation machinery of the endoplasmic reticulum. *Nature*, *458*(7237), 453-460. https://doi.org/10.1038/nature07962

Hjerpe, R., Aillet, F., Lopitz-Otsoa, F., Lang, V., England, P., & Rodriguez, M. S. (2009). Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. *EMBO Rep*, *10*(11), 1250-1258. https://doi.org/10.1038/embor.2009.192

Hofgen, R., & Willmitzer, L. (1988). Storage of competent cells for Agrobacterium transformation. *Nucleic Acids Res*, *16*(20), 9877. https://doi.org/10.1093/nar/16.20.9877

Hofmann, N. R., & Theg, S. M. (2005). Toc64 is not required for import of proteins into chloroplasts in the moss *Physcomitrella patens*. *Plant J.*, *43*(5), 675-687. https://doi.org/10.1111/j.1365-313X.2005.02483.x

Hooper, C. M., Castleden, I. R., Tanz, S. K., Aryamanesh, N., & Millar, A. H. (2017). SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Research*, *45*(D1), D1064-D1074. https://doi.org/10.1093/nar/gkw1041

Horn, S. C., Hanna, J., Hirsch, C., Volkwein, C., Schutz, A., Heinemann, U., Sommer, T., & Jarosch, E. (2009). Usa1 functions as a scaffold of the HRD-ubiquitin ligase. *Mol Cell*, *36*(5), 782-793. https://doi.org/10.1016/j.molcel.2009.10.015

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., & Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics*, *2008*, 420747. https://doi.org/10.1155/2008/420747

Hua, Z., & Vierstra, R. D. (2016). Ubiquitin goes green. *Trends Cell Biol.*, *26*(1), 3-5. https://doi.org/10.1016/j.tcb.2015.12.001

Huang, W., Ling, Q., Bedard, J., Lilley, K., & Jarvis, P. (2011). In vivo analyses of the roles of essential Omp85-related proteins in the chloroplast outer envelope membrane. *Plant Physiol*, *157*(1), 147-159. https://doi.org/10.1104/pp.111.181891

Inaba, T., Li, M., Alvarez-Huerta, M., Kessler, F., & Schnell, D. J. (2003). atTic110 functions as a scaffold for coordinating the stromal events of protein import into chloroplasts. *J. Biol. Chem.*, *278*(40), 38617-38627. https://doi.org/10.1074/jbc.M306367200

Infanger, S., Bischof, S., Hiltbrunner, A., Agne, B., Baginsky, S., & Kessler, F. (2011). The chloroplast import receptor Toc90 partially restores the accumulation of Toc159 client proteins in the Arabidopsis thaliana ppi2 mutant. *Mol Plant*, *4*(2), 252-263. https://doi.org/10.1093/mp/ssq071

Inoue, H., Li, M., & Schnell, D. J. (2013). An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.*, *110*(8), 3173-3178. https://doi.org/10.1073/pnas.1219229110

Inoue, H., Rounds, C., & Schnell, D. J. (2010). The molecular basis for distinct pathways for protein import into *Arabidopsis* chloroplasts. *Plant Cell*, *22*(6), 1947-1960. https://doi.org/10.1105/tpc.110.074328

Inoue, K., & Keegstra, K. (2003). A polyglycine stretch is necessary for proper targeting of the protein translocation channel precursor to the outer envelope membrane of chloroplasts. *Plant J.*, *34*(5), 661-669. https://doi.org/10.1046/j.1365-313x.2003.01755.x

Iqbal, H., Kenedy, M. R., Lybecker, M., & Akins, D. R. (2016). The TamB ortholog of Borrelia burgdorferi interacts with the beta-barrel assembly machine (BAM) complex protein BamA. *Mol Microbiol*, *102*(5), 757-774. https://doi.org/10.1111/mmi.13492

Irback, A., Mitternacht, S., & Mohanty, S. (2005). Dissecting the mechanical unfolding of ubiquitin. *Proc Natl Acad Sci U S A*, *102*(38), 13427-13432. https://doi.org/10.1073/pnas.0501581102

Ito, J., Batth, T. S., Petzold, C. J., Redding-Johanson, A. M., Mukhopadhyay, A., Verboom, R., Meyer, E. H., Millar, A. H., & Heazlewood, J. L. (2011). Analysis of the Arabidopsis cytosolic proteome highlights subcellular partitioning of central plant metabolism. *J Proteome Res*, *10*(4), 1571-1582. https://doi.org/10.1021/pr1009433

Izumi, M., Ishida, H., Nakamura, S., & Hidema, J. (2017). Entire Photodamaged Chloroplasts Are Transported to the Central Vacuole by Autophagy. *Plant Cell*, *29*(2), 377-394. https://doi.org/10.1105/tpc.16.00637

Jackson-Constan, D., & Keegstra, K. (2001). Arabidopsis genes encoding components of the chloroplastic protein import apparatus. *Plant Physiol*, *125*(4), 1567-1576. https://doi.org/10.1104/pp.125.4.1567

Jarvis, P. (2004). Organellar proteomics: chloroplasts in the spotlight. *Curr. Biol.*, *14*(8), R317-319. https://doi.org/10.1016/j.cub.2004.03.054

Jarvis, P. (2008). Targeting of nucleus-encoded proteins to chloroplasts in plants. *New Phytologist*, *179*(2), 257-285. https://doi.org/10.1111/j.1469-8137.2008.02452.x

Jarvis, P., Chen, L. J., Li, H., Peto, C. A., Fankhauser, C., & Chory, J. (1998). An Arabidopsis mutant defective in the plastid general protein import apparatus. *Science*, *282*(5386), 100-103. https://doi.org/10.1126/science.282.5386.100

Jarvis, P., & López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Rev. Mol. Cell. Biol.*, *14*(12), 787-802. https://doi.org/10.1038/nrm3702

Ji, Z., Li, H., Peterle, D., Paulo, J. A., Ficarro, S. B., Wales, T. E., Marto, J. A., Gygi, S. P., Engen, J. R., & Rapoport, T. A. (2022). Translocation of polyubiquitinated protein substrates by the hexameric Cdc48 ATPase. *Mol Cell*, *82*(3), 570-584 e578. https://doi.org/10.1016/j.molcel.2021.11.033

Jin, Z., Wan, L., Zhang, Y., Li, X., Cao, Y., Liu, H., Fan, S., Cao, D., Wang, Z., Li, X., Pan, J., Dong, M. Q., Wu, J., & Yan, Z. (2022). Structure of a TOC-TIC supercomplex spanning two chloroplast

envelope membranes. *Cell*, *185*(25), 4788-4800 e4713. https://doi.org/10.1016/j.cell.2022.10.030

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

Kakizaki, T., Matsumura, H., Nakayama, K., Che, F. S., Terauchi, R., & Inaba, T. (2009). Coordination of plastid protein import and nuclear gene expression by plastid-to-nucleus retrograde signaling. *Plant Physiol.*, *151*(3), 1339-1353. https://doi.org/10.1104/pp.109.145987

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587-+. https://doi.org/10.1038/Nmeth.4285

Karimi, M., Inze, D., & Depicker, A. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci*, *7*(5), 193-195. https://doi.org/10.1016/s1360-1385(02)02251-3

Kasmati, A. R., Topel, M., Khan, N. Z., Patel, R., Ling, Q., Karim, S., Aronsson, H., & Jarvis, P. (2013). Evolutionary, molecular and genetic analyses of Tic22 homologues in Arabidopsis thaliana chloroplasts. *PLoS One*, *8*(5), e63863. https://doi.org/10.1371/journal.pone.0063863

Kato, Y., & Sakamoto, W. (2009). Protein quality control in chloroplasts: a current model of D1 protein degradation in the photosystem II repair cycle. *J Biochem*, *146*(4), 463-469. https://doi.org/10.1093/jb/mvp073

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772-780. https://doi.org/10.1093/molbev/mst010

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845-858. https://doi.org/10.1038/nprot.2015.053

Kikuchi, S., Asakura, Y., Imai, M., Nakahira, Y., Kotani, Y., Hashiguchi, Y., Nakai, Y., Takafuji, K., Bédard, J., Hirabayashi-Ishioka, Y., Mori, H., Shiina, T., & Nakai, M. (2018). A Ycf2-FtsHi heteromeric AAA-ATPase complex Is required for chloroplast protein import. *Plant Cell*, *30*(11), 2677-2703. https://doi.org/10.1105/tpc.18.00357

Kikuchi, S., Bédard, J., Hirano, M., Hirabayashi, Y., Oishi, M., Imai, M., Takase, M., Ide, T., & Nakai, M. (2013). Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science*, *339*(6119), 571-574. https://doi.org/10.1126/science.1229262

Kikuchi, S., Hirohashi, T., & Nakai, M. (2006). Characterization of the preprotein translocon at the outer envelope membrane of chloroplasts by blue native PAGE. *Plant Cell Physiol.*, *47*(3), 363-371. https://doi.org/10.1093/pcp/pcj002

Kim, D. Y., Scalf, M., Smith, L. M., & Vierstra, R. D. (2013). Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. *Plant Cell*, *25*(5), 1523-1540. https://doi.org/10.1105/tpc.112.108613

Kim, J., Na, Y. J., Park, S. J., Baek, S. H., & Kim, D. H. (2019). Biogenesis of chloroplast outer envelope membrane proteins. *Plant Cell Rep.*, *38*(7), 783-792. https://doi.org/10.1007/s00299-019-02381-6

Klinger, A., Gosch, V., Bodensohn, U., Ladig, R., & Schleiff, E. (2019). The signal distinguishing between targeting of outer membrane beta-barrel protein to plastids and mitochondria in plants. *Biochim. Biophys. Acta Mol. Cell Res.*, *1866*(4), 663-672. https://doi.org/10.1016/j.bbamcr.2019.01.004

Kloppsteck, P., Ewens, C. A., Forster, A., Zhang, X., & Freemont, P. S. (2012). Regulation of p97 in the ubiquitin-proteasome system by the UBX protein-family. *Biochim Biophys Acta*, *1823*(1), 125-129. https://doi.org/10.1016/j.bbamcr.2011.09.006

Koenig, P., Mirus, O., Haarmann, R., Sommer, M. S., Sinning, I., Schleiff, E., & Tews, I. (2010). Conserved properties of polypeptide transport-associated (POTRA) domains derived from cyanobacterial Omp85. *J. Biol. Chem.*, *285*(23), 18016-18024. https://doi.org/10.1074/jbc.M110.112649

Köhler, D., Helm, S., Agne, B., & Baginsky, S. (2016). Importance of translocon subunit Tic56 for rRNA processing and chloroplast ribosome assembly. *Plant Physiol.*, *172*(4), 2429-2444. https://doi.org/10.1104/pp.16.01393

Köhler, D., Montandon, C., Hause, G., Majovsky, P., Kessler, F., Baginsky, S., & Agne, B. (2015). Characterization of chloroplast protein import without Tic56, a component of the 1-megadalton translocon at the inner envelope membrane of chloroplasts. *Plant Physiol.*, *167*(3), 972-990. https://doi.org/10.1104/pp.114.255562

Kouranov, A., Chen, X., Fuks, B., & Schnell, D. J. (1998). Tic20 and Tic22 are new components of the protein import apparatus at the chloroplast inner envelope membrane. *J. Cell Biol.*, *143*(4), 991-1002. https://doi.org/10.1083/jcb.143.4.991

Kouranov, A., & Schnell, D. J. (1997). Analysis of the interactions of preproteins with the import machinery over the course of protein import into chloroplasts. *J Cell Biol*, *139*(7), 1677-1685. https://doi.org/10.1083/jcb.139.7.1677

Kovács-Bogdan, E., Benz, J. P., Soll, J., & Bölter, B. (2011). Tic20 forms a channel independent of Tic110 in chloroplasts. *BMC Plant Biol.*, *11*, 133. https://doi.org/10.1186/1471-2229-11-133

Kraner, M. E., Muller, C., & Sonnewald, U. (2017). Comparative proteomic profiling of the choline transporter-like1 (CHER1) mutant provides insights into plasmodesmata composition of fully developed Arabidopsis thaliana leaves. *Plant J*, *92*(4), 696-709. https://doi.org/10.1111/tpj.13702

Kretzschmar, F. K., Mengel, L. A., Muller, A. O., Schmitt, K., Blersch, K. F., Valerius, O., Braus, G. H., & Ischebeck, T. (2018). PUX10 is a lipid droplet-localized scaffold protein that interacts with CELL DIVISION CYCLE48 and is involved in the degradation of lipid droplet proteins. *Plant Cell*, *30*(9), 2137-2160. https://doi.org/10.1105/tpc.18.00276

Kubis, S., Baldwin, A., Patel, R., Razzaq, A., Dupree, P., Lilley, K., Kurth, J., Leister, D., & Jarvis, P. (2003). The Arabidopsis ppi1 mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins. *Plant Cell*, *15*(8), 1859-1871. https://doi.org/10.1105/tpc.012955

Kubis, S., Patel, R., Combe, J., Bedard, J., Kovacheva, S., Lilley, K., Biehl, A., Leister, D., Rios, G., Koncz, C., & Jarvis, P. (2004). Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. *Plant Cell*, *16*(8), 2059-2077. https://doi.org/10.1105/tpc.104.023309

Kudla, J., & Bock, R. (2016). Lighting the Way to Protein-Protein Interactions: Recommendations on Best Practices for Bimolecular Fluorescence Complementation Analyses. *Plant Cell*, *28*(5), 1002-1008. https://doi.org/10.1105/tpc.16.00043

Laemmli, U. K. (1970). Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4. *Nature*, *227*(5259), 680-+. https://doi.org/DOI 10.1038/227680a0

Le, L. T., Kang, W., Kim, J. Y., Le, O. T., Lee, S. Y., & Yang, J. K. (2016). Structural Details of Ufd1 Binding to p97 and Their Functional Implications in ER-Associated Degradation. *PLoS One*, *11*(9), e0163394. https://doi.org/10.1371/journal.pone.0163394

Lee, D. W., & Hwang, I. (2018). Evolution and design principles of the diverse chloroplast transit peptides. *Mol. Cells*, *41*(3), 161-167. https://doi.org/10.14348/molcells.2018.0033

Lee, D. W., & Hwang, I. (2019). Protein import into chloroplasts via the Tic40-dependent and -independent pathways depends on the amino acid composition of the transit peptide. *Biochem Biophys Res Commun*, *518*(1), 66-71. https://doi.org/10.1016/j.bbrc.2019.08.009

Lee, D. W., Lee, J., & Hwang, I. (2017). Sorting of nuclear-encoded chloroplast membrane proteins. *Curr. Opin. Plant Biol.*, *40*, 1-7. https://doi.org/10.1016/j.pbi.2017.06.011

Lee, K. H., Kim, S. J., Lee, Y. J., Jin, J. B., & Hwang, I. (2003). The M domain of atToc159 plays an essential role in the import of proteins into chloroplasts and chloroplast biogenesis. *J Biol Chem*, *278*(38), 36794-36805. https://doi.org/10.1074/jbc.M304457200

Lee, S., Lee, D. W., Lee, Y., Mayer, U., Stierhof, Y. D., Lee, S., Jurgens, G., & Hwang, I. (2009). Heat shock protein cognate 70-4 and an E3 ubiquitin ligase, CHIP, mediate plastid-destined precursor degradation through the ubiquitin-26S proteasome system in Arabidopsis. *Plant Cell*, *21*(12), 3984-4001. https://doi.org/10.1105/tpc.109.071548

Lei, Y., Lu, L., Liu, H. Y., Li, S., Xing, F., & Chen, L. L. (2014). CRISPR-P: A Web Tool for Synthetic Single-Guide RNA Design of CRISPR-System in Plants. *Molecular Plant*, *7*(9), 1494-1496. https://doi.org/10.1093/mp/ssu044

Leipe, D. D., Wolf, Y. I., Koonin, E. V., & Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.*, *317*(1), 41-72. https://doi.org/10.1006/jmbi.2001.5378

Leister, D. (2003). Chloroplast research in the genomic age. *Trends Genet.*, *19*(1), 47-56. https://doi.org/10.1016/s0168-9525(02)00003-3

Li, H., & Chen, L. J. (1997). A novel chloroplastic outer membrane-targeting signal that functions at both termini of passenger polypeptides. *J Biol Chem*, *272*(16), 10968-10974. https://www.ncbi.nlm.nih.gov/pubmed/9099756

Li, H. M., & Teng, Y. S. (2013). Transit peptide design and plastid import regulation. *Trends Plant Sci.*, *18*(7), 360-366. https://doi.org/10.1016/j.tplants.2013.04.003

Li, J., Yuan, J., Li, Y., Sun, H., Ma, T., Huai, J., Yang, W., Zhang, W., & Lin, R. (2022). The CDC48 complex mediates ubiquitin-dependent degradation of intra-chloroplast proteins in plants. *Cell Rep*, *39*(2), 110664. https://doi.org/10.1016/j.celrep.2022.110664

Li, J., Zhang, F., Li, Y., Yang, W., & Lin, R. (2019). Chloroplast-Localized Protoporphyrinogen IX Oxidase1 Is Involved in the Mitotic Cell Cycle in Arabidopsis. *Plant Cell Physiol*, *60*(11), 2436-2448. https://doi.org/10.1093/pcp/pcz135

Li, K., Yu, R., Fan, L. M., Wei, N., Chen, H., & Deng, X. W. (2016). DELLA-mediated PIF degradation contributes to coordination of light and gibberellin signalling in Arabidopsis. *Nat Commun*, *7*, 11868. https://doi.org/10.1038/ncomms11868

Li, Y., Martin, J. R., Aldama, G. A., Fernandez, D. E., & Cline, K. (2017). Identification of putative substrates of SEC2, a chloroplast inner envelope translocase. *Plant Physiol.*, *173*(4), 2121-2137. https://doi.org/10.1104/pp.17.00012

Li, Y., Singhal, R., Taylor, I. W., McMinn, P. H., Chua, X. Y., Cline, K., & Fernandez, D. E. (2015). The Sec2 translocase of the chloroplast inner envelope contains a unique and dedicated SECE2 component. *Plant J.*, *84*(4), 647-658. https://doi.org/10.1111/tpj.13028

Ling, Q., Broad, W., Trösch, R., Topel, M., Demiral Sert, T., Lymperopoulos, P., Baldwin, A., & Jarvis, R. P. (2019). Ubiquitin-dependent chloroplast-associated protein degradation in plants. *Science*, *363*(6429). https://doi.org/10.1126/science.aav4467

Ling, Q., Huang, W., Baldwin, A., & Jarvis, P. (2012). Chloroplast biogenesis is regulated by direct action of the ubiquitin-proteasome system. *Science*, *338*(6107), 655-659. https://doi.org/10.1126/science.1225053

Ling, Q., & Jarvis, P. (2015a). Functions of plastid protein import and the ubiquitin-proteasome system in plastid development. *Biochim. Biophys. Acta*, *1847*(9), 939-948. https://doi.org/10.1016/j.bbabio.2015.02.017

Ling, Q., & Jarvis, P. (2015b). Regulation of chloroplast protein import by the ubiquitin E3 ligase SP1 is important for stress tolerance in plants. *Curr. Biol.*, *25*(19), 2527-2534. https://doi.org/10.1016/j.cub.2015.08.015

Ling, Q., Li, N., & Jarvis, P. (2017). Chloroplast ubiquitin E3 ligase SP1: does it really function in peroxisomes? *Plant Physiol.*, *175*(2), 586-588. https://doi.org/10.1104/pp.17.00948

Littlejohn, G. R., Gouveia, J. D., Edner, C., Smirnoff, N., & Love, J. (2010). Perfluorodecalin enhances in vivo confocal microscopy resolution of Arabidopsis thaliana mesophyll. *New Phytologist*, *186*(4), 1018-1025. https://doi.org/10.1111/j.1469-8137.2010.03244.x

Liu, H., Li, A., Rochaix, J. D., & Liu, Z. (2023). Architecture of chloroplast TOC-TIC translocon supercomplex. *Nature*, *615*(7951), 349-357. https://doi.org/10.1038/s41586-023-05744-y

Liu, L., McNeilage, R. T., Shi, L. X., & Theg, S. M. (2014). ATP requirement for chloroplast protein import is set by the Km for ATP hydrolysis of stromal Hsp70 in *Physcomitrella patens*. *Plant Cell*, *26*(3), 1246-1255. https://doi.org/10.1105/tpc.113.121822

Liu, S., Fu, Q. S., Zhao, J., & Hu, H. Y. (2013). Structural and mechanistic insights into the arginine/lysine-rich peptide motifs that interact with P97/VCP. *Biochim Biophys Acta*, *1834*(12), 2672-2678. https://doi.org/10.1016/j.bbapap.2013.09.021

Liu, Y., & Li, J. (2014). Endoplasmic reticulum-mediated protein quality control in *Arabidopsis*. *Front. Plant Sci.*, *5*, 162. https://doi.org/10.3389/fpls.2014.00162

Lopez-Juez, E., Jarvis, R. P., Takeuchi, A., Page, A. M., & Chory, J. (1998). New Arabidopsis cue mutants suggest a close connection between plastid- and phytochrome regulation of nuclear gene expression. *Plant Physiol*, *118*(3), 803-815. https://doi.org/10.1104/pp.118.3.803

Loudya, N., Maffei, D. P. F., J, B. D., Ali, S. M., Devlin, P. F., Jarvis, R. P., & E, L. P.-J. (2022). Mutations in the chloroplast inner envelope protein TIC100 impair and repair chloroplast protein import and impact retrograde signaling. *Plant Cell*, *34*(8), 3028-3046. https://doi.org/10.1093/plcell/koac153

Loudya, N., Okunola, T., He, J., Jarvis, P., & Lopez-Juez, E. (2020). Retrograde signalling in a virescent mutant triggers an anterograde delay of chloroplast biogenesis that requires GUN1 and is essential for survival. *Philos Trans R Soc Lond B Biol Sci*, *375*(1801), 20190400. https://doi.org/10.1098/rstb.2019.0400

Lumme, C., Altan-Martin, H., Dastvan, R., Sommer, M. S., Oreb, M., Schuetz, D., Hellenkamp, B., Mirus, O., Kretschmer, J., Lyubenova, S., Kugel, W., Medelnik, J. P., Dehmer, M., Michaelis, J., Prisner, T. F., Hugel, T., & Schleiff, E. (2014). Nucleotides and substrates trigger the dynamics of the Toc34 GTPase homodimer involved in chloroplast preprotein translocation. *Structure*, *22*(4), 526-538. https://doi.org/10.1016/j.str.2014.02.004

Ma, Y., Kouranov, A., LaSala, S. E., & Schnell, D. J. (1996). Two components of the chloroplast protein import apparatus, IAP86 and IAP75, interact with the transit sequence during the recognition and translocation of precursor proteins at the outer envelope. *J. Cell Biol.*, *134*(2), 315-327. https://doi.org/10.1083/jcb.134.2.315

May, T., & Soll, J. (2000). 14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants. *Plant Cell*, *12*(1), 53-64. https://doi.org/10.1105/tpc.12.1.53

McFadden, G. I. (2014). Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harb. Perspect. Biol.*, *6*(4), a016105. https://doi.org/10.1101/cshperspect.a016105

Mérai, Z., Chumak, N., Garcia-Aguilar, M., Hsieh, T. F., Nishimura, T., Schoft, V. K., Bindics, J., Slusarz, L., Arnoux, S., Opravil, S., Mechtler, K., Zilberman, D., Fischer, R. L., & Tamaru, H. (2014). The AAA-ATPase molecular chaperone Cdc48/p97 disassembles sumoylated centromeres, decondenses heterochromatin, and activates ribosomal RNA genes. *Proc Natl Acad Sci U S A*, *111*(45), 16166-16171. https://doi.org/10.1073/pnas.1418564111

Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., Samaras, P., Richter, S., Shikata, H., Messerer, M., Lang, D., Altmann, S., Cyprys, P., Zolg, D. P., Mathieson, T., Bantscheff, M., Hazarika, R. R., Schmidt, T., Dawid, C., . . . Kuster, B. (2020). Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature*, *579*(7799), 409-414. https://doi.org/10.1038/s41586-020-2094-2

Metzger, M. B., Liang, Y. H., Das, R., Mariano, J., Li, S., Li, J., Kostova, Z., Byrd, R. A., Ji, X., & Weissman, A. M. (2013). A structurally unique E2-binding domain activates ubiquitination by the ERAD E2, Ubc7p, through multiple mechanisms. *Mol Cell*, *50*(4), 516-527. https://doi.org/10.1016/j.molcel.2013.04.004

Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, *19*(6), 679-682. https://doi.org/10.1038/s41592-022-01488-1

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H. Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., . . . Finn, R. D. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*, *47*(D1), D351-D360. https://doi.org/10.1093/nar/gky1100

Moir, D., Stewart, S. E., Osmond, B. C., & Botstein, D. (1982). Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies. *Genetics*, *100*(4), 547-563. https://doi.org/10.1093/genetics/100.4.547

Movahedi, S., Van Bel, M., Heyndrickx, K. S., & Vandepoele, K. (2012). Comparative co-expression analysis in plant biology. *Plant Cell Environ*, *35*(10), 1787-1798. https://doi.org/10.1111/j.1365-3040.2012.02517.x

Na, H., & Song, G. (2016). Predicting the functional motions of p97 using symmetric normal modes. *Proteins*, *84*(12), 1823-1835. https://doi.org/10.1002/prot.25164

Nagahama, M., Hara, Y., Seki, A., Yamazoe, T., Kawate, Y., Shinohara, T., Hatsuzawa, K., Tani, K., & Tagaya, M. (2004). NVL2 is a nucleolar AAA-ATPase that interacts with ribosomal protein L5 through its nucleolar localization sequence. *Mol Biol Cell*, *15*(12), 5712-5723. https://doi.org/10.1091/mbc.e04-08-0692

Nakai, M. (2018). New perspectives on chloroplast protein import. *Plant Cell Physiol.*, *59*(6), 1111-1119. https://doi.org/10.1093/pcp/pcy083

Nakrieko, K. A., Mould, R. M., & Smith, A. G. (2004). Fidelity of targeting to chloroplasts is not affected by removal of the phosphorylation site from the transit peptide. *Eur. J. Biochem.*, *271*(3), 509-516. https://doi.org/10.1046/j.1432-1033.2003.03950.x

Neuber, O., Jarosch, E., Volkwein, C., Walter, J., & Sommer, T. (2005). Ubx2 links the Cdc48 complex to ER-associated protein degradation. *Nat Cell Biol*, *7*(10), 993-998. https://doi.org/10.1038/ncb1298

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268-274. https://doi.org/10.1093/molbev/msu300

Nguyen, T. Q., My Le, L. T., Kim, D. H., Ko, K. S., Lee, H. T., Kim Nguyen, Y. T., Kim, H. S., Han, B. W., Kang, W., & Yang, J. K. (2022). Structural basis for the interaction between human Npl4 and Npl4-binding motif of human Ufd1. *Structure*, *30*(11), 1530-1537 e1533. https://doi.org/10.1016/j.str.2022.08.005

Nielsen, E., Akita, M., Davila-Aponte, J., & Keegstra, K. (1997). Stable association of chloroplastic precursors with protein translocation complexes that contain proteins from both envelope membranes and a stromal Hsp100 molecular chaperone. *EMBO J.*, *16*(5), 935-946. https://doi.org/10.1093/emboj/16.5.935

Nishimura, K., Kato, Y., & Sakamoto, W. (2017). Essentials of proteolytic machineries in chloroplasts. *Mol. Plant*, *10*(1), 4-19. https://doi.org/10.1016/j.molp.2016.08.005

Niwa, H., Ewens, C. A., Tsang, C., Yeung, H. O., Zhang, X., & Freemont, P. S. (2012). The role of the N-domain in the ATPase activity of the mammalian AAA ATPase p97/VCP. *J Biol Chem*, *287*(11), 8561-8570. https://doi.org/10.1074/jbc.M111.302778

O'Neil, P. K., Richardson, L. G. L., Paila, Y. D., Piszczek, G., Chakravarthy, S., Noinaj, N., & Schnell, D. (2017). The POTRA domains of Toc75 exhibit chaperone-like function to facilitate import into chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.*, *114*(24), E4868-E4876. https://doi.org/10.1073/pnas.1621179114

Paila, Y. D., Richardson, L. G. L., & Schnell, D. J. (2015). New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. *J. Mol. Biol.*, *427*(5), 1038-1060. https://doi.org/10.1016/j.jmb.2014.08.016

Pan, M., Zheng, Q., Yu, Y., Ai, H., Xie, Y., Zeng, X., Wang, C., Liu, L., & Zhao, M. (2021). Seesaw conformations of Npl4 in the human p97 complex and the inhibitory mechanism of a disulfiram derivative. *Nat Commun*, *12*(1), 121. https://doi.org/10.1038/s41467-020-20359-x

Panchy, N., Lehti-Shiu, M., & Shiu, S. H. (2016). Evolution of Gene Duplication in Plants. *Plant Physiol*, *171*(4), 2294-2316. https://doi.org/10.1104/pp.16.00523

Panigrahi, R., Adina-Zada, A., Whelan, J., & Vrielink, A. (2013). Ligand recognition by the TPR domain of the import factor Toc64 from *Arabidopsis thaliana*. *PLoS One*, *8*(12), e83461. https://doi.org/10.1371/journal.pone.0083461

Park, S., Isaacson, R., Kim, H. T., Silver, P. A., & Wagner, G. (2005). Ufd1 exhibits the AAA-ATPase fold with two distinct ubiquitin interaction sites. *Structure*, *13*(7), 995-1005. https://doi.org/10.1016/j.str.2005.04.013

Pereira, J., Simpkin, A. J., Hartmann, M. D., Rigden, D. J., Keegan, R. M., & Lupas, A. N. (2021). High-accuracy protein structure prediction in CASP14. *Proteins*, *89*(12), 1687-1699. https://doi.org/10.1002/prot.26171

Perry, S. E., & Keegstra, K. (1994). Envelope membrane proteins that interact with chloroplastic precursor proteins. *Plant Cell*, *6*(1), 93-105. https://doi.org/10.1105/tpc.6.1.93

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*, *30*(1), 70-82. https://doi.org/10.1002/pro.3943

Pluthero, F. G. (1993). Rapid purification of high-activity Taq DNA polymerase. *Nucleic Acids Res*, *21*(20), 4850-4851. https://doi.org/10.1093/nar/21.20.4850

Ramundo, S., Asakura, Y., Salome, P. A., Strenkert, D., Boone, M., Mackinder, L. C. M., Takafuji, K., Dinc, E., Rahire, M., Crevecoeur, M., Magneschi, L., Schaad, O., Hippler, M., Jonikas, M. C., Merchant, S., Nakai, M., Rochaix, J. D., & Walter, P. (2020). Coexpressed subunits of dual genetic origin define a conserved supercomplex mediating essential protein import into chloroplasts. *Proc Natl Acad Sci U S A*, *117*(51), 32739-32749. https://doi.org/10.1073/pnas.2014294117

Reddick, L. E., Vaughn, M. D., Wright, S. J., Campbell, I. M., & Bruce, B. D. (2007). In vitro comparative kinetic analysis of the chloroplast Toc GTPases. *J Biol Chem*, *282*(15), 11410-11426. https://doi.org/10.1074/jbc.M609491200

Richardson, L. G. L., Small, E. L., Inoue, H., & Schnell, D. J. (2018). Molecular topology of the transit peptide during chloroplast protein import. *Plant Cell*, *30*(8), 1789-1806. https://doi.org/10.1105/tpc.18.00172

Rochaix, J. D., & Ramundo, S. (2018). Chloroplast signaling and quality control. *Essays Biochem.*, *62*(1), 13-20. https://doi.org/doi:10.1042/EBC20170048

Rohila, J. S., Chen, M., Cerny, R., & Fromm, M. E. (2004). Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. *Plant J*, *38*(1), 172-181. https://doi.org/10.1111/j.1365-313X.2004.02031.x

Rolland, N., Curien, G., Finazzi, G., Kuntz, M., Marechal, E., Matringe, M., Ravanel, S., & Seigneurin-Berny, D. (2012). The biosynthetic capacities of the plastids and integration between cytoplasmic and chloroplast processes. *Annu. Rev. Genet.*, *46*, 233-264. https://doi.org/10.1146/annurev-genet-110410-132544

Sadali, N. M., Sowden, R. G., Ling, Q., & Jarvis, R. P. (2019). Differentiation of chromoplasts and other plastids in plants. *Plant Cell Rep.*, *38*(7), 803-818. https://doi.org/10.1007/s00299-019-02420-2

Sako, K., Yanagawa, Y., Kanai, T., Sato, T., Seki, M., Fujiwara, M., Fukao, Y., & Yamaguchi, J. (2014). Proteomic analysis of the 26S proteasome reveals its direct interaction with transit peptides of plastid protein precursors for their degradation. *J. Proteome Res.*, *13*(7), 3223-3230. https://doi.org/10.1021/pr401245g

Sato, Y., Tsuchiya, H., Yamagata, A., Okatsu, K., Tanaka, K., Saeki, Y., & Fukai, S. (2019). Structural insights into ubiquitin recognition and Ufd1 interaction of Npl4. *Nat Commun*, *10*(1), 5708. https://doi.org/10.1038/s41467-019-13697-y

Schäfer, P., Helm, S., Kohler, D., Agne, B., & Baginsky, S. (2019). Consequences of impaired 1-MDa TIC complex assembly for the abundance and composition of chloroplast high-molecular mass protein complexes. *PLoS One*, *14*(3), e0213364. https://doi.org/10.1371/journal.pone.0213364

Schelbert, S., Aubry, S., Burla, B., Agne, B., Kessler, F., Krupinska, K., & Hortensteiner, S. (2009). Pheophytin Pheophorbide Hydrolase (Pheophytinase) Is Involved in Chlorophyll Breakdown during Leaf Senescence in Arabidopsis. *Plant Cell*, *21*(3), 767-785. https://doi.org/10.1105/tpc.108.064089

Schleiff, E., & Becker, T. (2011). Common ground for protein translocation: access control for mitochondria and chloroplasts. *Nat. Rev. Mol. Cell Biol.*, *12*(1), 48-59. https://doi.org/10.1038/nrm3027

Schleiff, E., Eichacker, L. A., Eckart, K., Becker, T., Mirus, O., Stahl, T., & Soll, J. (2003). Prediction of the plant beta-barrel proteome: a case study of the chloroplast outer envelope. *Protein Sci.*, *12*(4), 748-759. https://doi.org/10.1110/ps.0237503

Schleiff, E., Soll, J., Kuchler, M., Kuhlbrandt, W., & Harrer, R. (2003). Characterization of the translocon of the outer envelope of chloroplasts. *J. Cell Biol.*, *160*(4), 541-551. https://doi.org/10.1083/jcb.200210060

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., & Lohmann, J. U. (2005). A gene expression map of Arabidopsis thaliana development. *Nat Genet*, *37*(5), 501-506. https://doi.org/10.1038/ng1543

Schnell, D. J., Kessler, F., & Blobel, G. (1994). Isolation of components of the chloroplast protein import machinery. *Science*, *266*(5187), 1007-1012. https://doi.org/10.1126/science.7973649

Schuberth, C., & Buchberger, A. (2005). Membrane-bound Ubx2 recruits Cdc48 to ubiquitin ligases and their substrates to ensure efficient ER-associated protein degradation. *Nat. Cell Biol.*, *7*(10), 999-1006. https://doi.org/10.1038/ncb1299

Schünemann, D. (2007). Mechanisms of protein import into thylakoids of chloroplasts. *Biol. Chem.*, *388*(9), 907-915. https://doi.org/10.1515/BC.2007.111

Schwacke, R., Schneider, A., van der Graaff, E., Fischer, K., Catoni, E., Desimone, M., Frommer, W. B., Flugge, U. I., & Kunze, R. (2003). ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiology*, *131*(1), 16-26. https://doi.org/10.1104/pp.011577

Shanmugabalaji, V., Chahtane, H., Accossato, S., Rahire, M., Gouzerh, G., Lopez-Molina, L., & Kessler, F. (2018). Chloroplast biogenesis controlled by DELLA-TOC159 interaction in early plant development. *Curr. Biol.*, *28*(16), 2616-2623 e2615. https://doi.org/10.1016/j.cub.2018.06.006

Shen, G., Adam, Z., & Zhang, H. (2007). The E3 ligase AtCHIP ubiquitylates FtsH1, a component of the chloroplast FtsH protease, and affects protein degradation in chloroplasts. *Plant J.*, *52*(2), 309-321. https://doi.org/10.1111/j.1365-313X.2007.03239.x

Shen, G., Yan, J., Pasapula, V., Luo, J., He, C., Clarke, A. K., & Zhang, H. (2007). The chloroplast protease subunit ClpP4 is a substrate of the E3 ligase AtCHIP and plays an important role in chloroplast function. *Plant J.*, *49*(2), 228-237. https://doi.org/10.1111/j.1365-313X.2006.02963.x

Shi, L. X., & Theg, S. M. (2010). A stromal heat shock protein 70 system functions in protein import into chloroplasts in the moss *Physcomitrella patens*. *Plant Cell*, *22*(1), 205-220. https://doi.org/10.1105/tpc.109.071464

Shi, L. X., & Theg, S. M. (2013). Energetic cost of protein import across the envelope membranes of chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.*, *110*(3), 930-935. https://doi.org/10.1073/pnas.1115886110

Shih, P. M., & Matzke, N. J. (2013). Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc Natl Acad Sci U S A*, *110*(30), 12355-12360. https://doi.org/10.1073/pnas.1305813110

Sibbald, S. J., & Archibald, J. M. (2020). Genomic Insights into Plastid Evolution. *Genome Biol Evol*, *12*(7), 978-990. https://doi.org/10.1093/gbe/evaa096

Sjögren, L. L., Tanabe, N., Lymperopoulos, P., Khan, N. Z., Rodermel, S. R., Aronsson, H., & Clarke, A. K. (2014). Quantitative analysis of the chloroplast molecular chaperone ClpC/Hsp93 in *Arabidopsis* reveals new insights into its localization, interaction with the Clp proteolytic core, and functional importance. *J. Biol. Chem.*, *289*(16), 11318-11330. https://doi.org/10.1074/jbc.M113.534552

Sjuts, I., Soll, J., & Bölter, B. (2017). Import of soluble proteins into chloroplasts and potential regulatory Mechanisms. *Front. Plant Sci.*, *8*, 168. https://doi.org/10.3389/fpls.2017.00168

Stapf, C., Cartwright, E., Bycroft, M., Hofmann, K., & Buchberger, A. (2011). The general definition of the p97/valosin-containing protein (VCP)-interacting motif (VIM) delineates a new family of p97 cofactors. *J Biol Chem*, *286*(44), 38670-38678. https://doi.org/10.1074/jbc.M111.274472

Stein, A., Ruggiano, A., Carvalho, P., & Rapoport, T. A. (2014). Key steps in ERAD of luminal ER proteins reconstituted with purified components. *Cell*, *158*(6), 1375-1388. https://doi.org/10.1016/j.cell.2014.07.050

Su, P. H., & Li, H. M. (2010). Stromal Hsp70 is important for protein translocation into pea and *Arabidopsis* chloroplasts. *Plant Cell*, *22*(5), 1516-1531. https://doi.org/10.1105/tpc.109.071415

Sun, Y., Yao, Z., Ye, Y., Fang, J., Chen, H., Lyu, Y., Broad, W., Fournier, M., Chen, G., Hu, Y., Mohammed, S., Ling, Q., & Jarvis, R. P. (2022). Ubiquitin-based pathway acts inside chloroplasts to regulate photosynthesis. *Sci Adv*, *8*(46), eabq7352. https://doi.org/10.1126/sciadv.abq7352

Suzuki, M., Takahashi, S., Kondo, T., Dohra, H., Ito, Y., Kiriiwa, Y., Hayashi, M., Kamiya, S., Kato, M., Fujiwara, M., Fukao, Y., Kobayashi, M., Nagata, N., & Motohashi, R. (2015). Plastid proteomic analysis in tomato fruit development. *PLoS One*, *10*(9), e0137266. https://doi.org/10.1371/journal.pone.0137266

Svozil, J., Hirsch-Hoffmann, M., Dudler, R., Gruissem, W., & Baerenfaller, K. (2014). Protein abundance changes and ubiquitylation targets identified after inhibition of the proteasome with syringolin A. *Mol. Cell Proteomics*, *13*(6), 1523-1536. https://doi.org/10.1074/mcp.M113.036269

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*, *49*(D1), D605-D612. https://doi.org/10.1093/nar/gkaa1074

Tarnowski, B. I., Spinale, F. G., & Nicholson, J. H. (1991). DAPI as a useful stain for nuclear quantitation. *Biotech Histochem*, *66*(6), 297-302. https://www.ncbi.nlm.nih.gov/pubmed/1725854

Teng, Y. S., Su, Y. S., Chen, L. J., Lee, Y. J., Hwang, I., & Li, H. M. (2006). Tic21 is an essential translocon component for protein translocation across the chloroplast inner envelope membrane. *Plant Cell*, *18*(9), 2247-2257. https://doi.org/10.1105/tpc.106.044305

Thomson, S. M., Pulido, P., & Jarvis, R. P. (2020). Protein import into chloroplasts and its regulation by the ubiquitin-proteasome system. *Biochem Soc Trans*, *48*(1), 71-82. https://doi.org/10.1042/BST20190274

Timmis, J. N., Ayliffe, M. A., Huang, C. Y., & Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.*, *5*(2), 123-135. https://doi.org/10.1038/nrg1271

Tokumaru, M., Adachi, F., Toda, M., Ito-Inaba, Y., Yazu, F., Hirosawa, Y., Sakakibara, Y., Suiko, M., Kakizaki, T., & Inaba, T. (2017). Ubiquitin-proteasome dependent regulation of the GOLDEN2-

LIKE 1 transcription factor in response to plastid signals. *Plant Physiol.*, *173*(1), 524-535. https://doi.org/10.1104/pp.16.01546

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, *44*(W1), W232-W235. https://doi.org/10.1093/nar/gkw256

Tripp, J., Hahn, A., Koenig, P., Flinner, N., Bublak, D., Brouwer, E. M., Ertel, F., Mirus, O., Sinning, I., Tews, I., & Schleiff, E. (2012). Structure and conservation of the periplasmic targeting factor Tic22 protein from plants and cyanobacteria. *J. Biol. Chem.*, *287*(29), 24164-24173. https://doi.org/10.1074/jbc.M112.341644

Trösch, R., & Jarvis, P. (2011). The stromal processing peptidase of chloroplasts is essential in Arabidopsis, with knockout mutations causing embryo arrest after the 16-cell stage. *PLoS One*, *6*(8), e23039. https://doi.org/10.1371/journal.pone.0023039

Tsai, J. Y., Chu, C. C., Yeh, Y. H., Chen, L. J., Li, H. M., & Hsiao, C. D. (2013). Structural characterizations of the chloroplast translocon protein Tic110. *Plant J*, *75*(5), 847-857. https://doi.org/10.1111/tpj.12249

Tsutsui, H., & Higashiyama, T. (2017). pKAMA-ITACHI Vectors for Highly Efficient CRISPR/Cas9-Mediated Gene Knockout in Arabidopsis thaliana. *Plant and Cell Physiology*, *58*(1), 46-56. https://doi.org/10.1093/pcp/pcw191

Twomey, E. C., Ji, Z., Wales, T. E., Bodnar, N. O., Ficarro, S. B., Marto, J. A., Engen, J. R., & Rapoport, T. A. (2019). Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding. *Science*, *365*(6452). https://doi.org/10.1126/science.aax1033

Tzfira, T., Tian, G. W., Lacroix, B., Vyas, S., Li, J., Leitner-Dagan, Y., Krichevsky, A., Taylor, T., Vainstein, A., & Citovsky, V. (2005). pSAT vectors: a modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. *Plant Mol Biol*, *57*(4), 503-516. https://doi.org/10.1007/s11103-005-0340-5

van Wijk, K. J., & Baginsky, S. (2011). Plastid proteomics in higher plants: current state and future goals. *Plant Physiol*, *155*(4), 1578-1588. https://doi.org/10.1104/pp.111.172932

Vaz, B., Halder, S., & Ramadan, K. (2013). Role of p97/VCP (Cdc48) in genome stability. *Front Genet*, *4*, 60. https://doi.org/10.3389/fgene.2013.00060

Venne, A. S., Kollipara, L., & Zahedi, R. P. (2014). The next level of complexity: crosstalk of posttranslational modifications. *Proteomics*, *14*(4-5), 513-524. https://doi.org/10.1002/pmic.201300344

Viana, A. A., Li, M., & Schnell, D. J. (2010). Determinants for stop-transfer and post-import pathways for protein targeting to the chloroplast inner envelope membrane. *J. Biol. Chem.*, *285*(17), 12948-12960. https://doi.org/10.1074/jbc.M110.109744

Vierstra, R. D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.*, *10*(6), 385-397. https://doi.org/10.1038/nrm2688

Wallas, T. R., Smith, M. D., Sanchez-Nieto, S., & Schnell, D. J. (2003). The roles of toc34 and toc75 in targeting the toc159 preprotein receptor to chloroplasts. *J Biol Chem*, *278*(45), 44289-44297. https://doi.org/10.1074/jbc.M307873200

Wang, J., Yu, Q., Xiong, H., Wang, J., Chen, S., Yang, Z., & Dai, S. (2016). Proteomic insight into the response of *Arabidopsis* chloroplasts to darkness. *PLoS One*, *11*(5), e0154235. https://doi.org/10.1371/journal.pone.0154235

Watson, S. J., Sowden, R. G., & Jarvis, P. (2018). Abiotic stress-induced chloroplast proteome remodelling: a mechanistic overview. *J. Exp. Bot.*, *69*(11), 2773-2781. https://doi.org/10.1093/jxb/ery053

Weibel, P., Hiltbrunner, A., Brand, L., & Kessler, F. (2003). Dimerization of Toc-GTPases at the chloroplast protein import machinery. *J Biol Chem*, *278*(39), 37321-37329. https://doi.org/10.1074/jbc.M305946200

Wiesemann, K., Simm, S., Mirus, O., Ladig, R., & Schleiff, E. (2019). Regulation of two GTPases Toc159 and Toc34 in the translocon of the outer envelope of chloroplasts. *Biochim. Biophys. Acta Proteins Proteom.*, *1867*(6), 627-636. https://doi.org/10.1016/j.bbapap.2019.01.002

Witwicka, H., Jia, H., Kutikov, A., Reyes-Gutierrez, P., Li, X., & Odgren, P. R. (2015). TRAFD1 (FLN29) Interacts with Plekhm1 and Regulates Osteoclast Acidification and Resorption. *PLoS One*, *10*(5), e0127537. https://doi.org/10.1371/journal.pone.0127537

Wolf, D. H., & Stolz, A. (2012). The Cdc48 machine in endoplasmic reticulum associated protein degradation. *Biochim Biophys Acta*, *1823*(1), 117-124. https://doi.org/10.1016/j.bbamcr.2011.09.002

Woodson, J. D., Joens, M. S., Sinson, A. B., Gilkerson, J., Salome, P. A., Weigel, D., Fitzpatrick, J. A., & Chory, J. (2015). Ubiquitin facilitates a quality-control pathway that removes damaged chloroplasts. *Science*, *350*(6259), 450-454. https://doi.org/10.1126/science.aac7444

Woodson, J. D., Perez-Ruiz, J. M., Schmitz, R. J., Ecker, J. R., & Chory, J. (2013). Sigma factor-mediated plastid retrograde signals control nuclear gene expression. *Plant J.*, *73*(1), 1-13. https://doi.org/10.1111/tpj.12011

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

10.1186/1746-4811-5-16

Wu, W., Cheng, Z., Liu, M., Yang, X., & Qiu, D. (2014). C3HC4-type RING finger protein NbZFP1 is involved in growth and fruit development in Nicotiana benthamiana. *PLoS One*, *9*(6), e99352. https://doi.org/10.1371/journal.pone.0099352

Wu, X., Li, L., & Jiang, H. (2016). Doa1 targets ubiquitinated substrates for mitochondria-associated degradation. *J. Cell Biol.*, *213*(1), 49-63. https://doi.org/10.1083/jcb.201510098

Wu, X., & Rapoport, T. A. (2018). Mechanistic insights into ER-associated protein degradation. *Curr Opin Cell Biol*, *53*, 22-28. https://doi.org/10.1016/j.ceb.2018.04.004

Xie, H., Wang, Y., Ding, Y., Qiu, C., Sun, L., Gai, Z., Gu, H., & Ding, Z. (2019). Global ubiquitome profiling revealed the roles of ubiquitinated proteins in metabolic pathways of tea leaves in responding to drought stress. *Sci. Rep.*, *9*(1), 4286. https://doi.org/10.1038/s41598-019-41041-3

Yang, X., Li, Y., Qi, M., Liu, Y., & Li, T. (2019). Targeted control of chloroplast quality to improve plant acclimation: from protein import to degradation. *Front. Plant Sci.*, *10*, 958. https://doi.org/10.3389/fpls.2019.00958

Yang, X. F., Wang, Y. T., Chen, S. T., Li, J. K., Shen, H. T., & Guo, F. Q. (2016). PBR1 selectively controls biogenesis of photosynthetic complexes by modulating translation of the large chloroplast gene Ycf1 in *Arabidopsis*. *Cell Discov.*, *2*, 16003. https://doi.org/10.1038/celldisc.2016.3

Ye, Y., Tang, W. K., Zhang, T., & Xia, D. (2017). A mighty "protein extractor" of the cell: structure and function of the p97/CDC48 ATPase. *Front. Mol. Biosci.*, *4*, 39. https://doi.org/10.3389/fmolb.2017.00039

Zang, Y., Gong, Y., Wang, Q., Guo, H., & Xiao, W. (2020). Arabidopsis OTU1, a linkage-specific deubiquitinase, is required for endoplasmic reticulum-associated protein degradation. *Plant J*, *101*(1), 141-155. https://doi.org/10.1111/tpj.14524

Zhang, J., Wu, S., Boehlein, S. K., McCarty, D. R., Song, G., Walley, J. W., Myers, A., & Settles, A. M. (2019). Maize *defective kernel5* is a bacterial TamB homologue required for chloroplast envelope biogenesis. *J. Cell Biol.*, *218*(8), 2638-2658. https://doi.org/10.1083/jcb.201807166

Zhang, N., Xu, J., Liu, X., Liang, W., Xin, M., Du, J., Hu, Z., Peng, H., Guo, W., Ni, Z., Sun, Q., & Yao, Y. (2019). Identification of HSP90C as a substrate of E3 ligase TaSAP5 through ubiquitylome profiling. *Plant Sci.*, *287*, 110170. https://doi.org/10.1016/j.plantsci.2019.110170

Zhao, G., Zhou, X., Wang, L., Li, G., Schindelin, H., & Lennarz, W. J. (2007). Studies on peptide:N-glycanase-p97 interaction suggest that p97 phosphorylation modulates endoplasmic reticulum-associated degradation. *Proc Natl Acad Sci U S A*, *104*(21), 8785-8790. https://doi.org/10.1073/pnas.0702966104

Zufferey, M., Montandon, C., Douet, V., Demarsy, E., Agne, B., Baginsky, S., & Kessler, F. (2017). The novel chloroplast outer membrane kinase KOC1 is a required component of the plastid protein import machinery. *J. Biol. Chem.*, *292*(17), 6952-6964. https://doi.org/10.1074/jbc.M117.776468