Characterising the functional domains of the candidate ALS protein, cyclin F

Kimberley Duncan

88257444

Australian School of Advanced Medicine, Macquarie University

A thesis submitted for the partial fulfillment of the requirements for the degree of Master of Research in Advanced Medicine

Supervisors

Associate Professor Ian Blair[1]

Dr Kelly Williams[1]

[1]Motor Neuron Disease Research Centre, Australian School of Advanced Medicine, Faculty of Human Sciences, Macquarie University, Sydney, NSW, Australia

Keywords: amyotrophic lateral sclerosis, motor neuron disease, cyclin F, *CCNF*, deletion cloning, homology modelling

Main text word count: 15653

Abstract word count: 200

Number of figures: 46

Declaration

I wish to acknowledge the following assistance in the research outlined in this project:

Primer sets for restriction cloning of *CCNF* were designed and optimised by Kelly Williams.

The pmCherry-C1_*CCNF*_WT was previously constructed by Alison Hogan.

Unless otherwise stated, all sequencing primers were designed previously by Alison Hogan.

DNA sequencing was carried out by Macrogen Sequencing Korea.

All other research described in this project is my own original work, and has not been submitted in whole or part for a degree in any university.

Acknowledgements

Firstly, I wish to thank my supervisor, Ian Blair, and his wonderful team, Kelly Williams, Shu Yang, Emily McCann, Katharine Zhang, Jennifer Fifita & Alison Hogan. Ian, you have to be the most supportive, positive and approachable supervisor anyone could ever hope for. I'm truly grateful for all the great advice that you've given me, and I've really enjoyed the opportunity to work in such a well-run and productive genetics lab. Thank you everyone for being so welcoming. Apart from all the assistance you've given me with the research and the write-up, I've really enjoyed our regular morning teas and lunchtime chats. This year has been about much more than just the work.

I particularly want to thank:

- Emily for getting me started in the lab. You're a natural teacher and always answer even my silliest questions with consideration and patience.
- Katharine for showing me everything there is to know about tissue culture and cloning while chatting about GOT in the downtime. I really couldn't have done it without you.
- Alison for helping me troubleshoot everything from cloning to sequencing to Latex problems.
- Jenn for always being there to help me troubleshoot. Jenn, you are always so calm and positive, and have an amazing ability for giving people confidence in themselves.
- Shu for guiding me through my first foray into cell biology. My science journey has gone in a direction I wasn't expecting and I've loved it. It's been great to get to know you, and I now count you as a friend.

- And of course, Kelly, who guided me along the journey from beginning to end. I always felt that I was in very capable hands. Thank you for all the help and advice you've given me this year.

I also wish to acknowledge Stephanie Raynor and Albert Lee, who gave me some great tips about proteins, and Emily Don, who was always there with sage cloning advice when I needed it.

Finally, thanks to my family for putting up with me for the last seven years of my science journey! Especially my husband Kimberley, and kids Sylvie and Xavier. Your cuddles always help when I'm feeling stressed. Thanks also to my parents, Jean and Wil, and sister, Michelle, for all your love and support, and for looking after the kids when I needed to get work done. I couldn't have completed this without you.

Abstract

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease characterised by the rapid degeneration of motor neurons leading to muscle paralysis and death. The only known causes of ALS are gene mutations, several of which are intrinsic to protein degradation pathways. Dysfunctional protein degradation could lead to abnormal protein accumulation in neurons, a hallmark of ALS pathology. Our laboratory recently identified ALS-linked mutations in *CCNF*, which encodes cyclin F, an E3 ubiquitin ligase that mediates the destruction of aberrant proteins by the ubiquitin proteasome system (UPS). The overall aim of this project was to characterise the function of cyclin F domains and predict the effect of ALS-linked mutations on protein function. First, bioinformatic modelling was performed to visualise the proximity of cyclin F mutations to active sites and predict their effect on protein function. Next, a series of *CCNF* deletion clones were generated for expression in neuronal cell lines to assess the effect of deleting functional domains on sub-cellular localisation of cyclin F as well as toxicity. 3D modelling of the mutations within the cyclin box show that they may affect substrate binding, while deletion cloning studies show that mutations within PEST could affect protein stability and cause mis-localisation, leading to protein aggregation. This could lead to the enhanced cytotoxicity that was found in ΔPEST cyclin F-transfected cells. These deletion constructs can be utilised in future *in vitro* studies to further examine the effects of individual domains on protein binding, stability and UPS function, to elucidate the potential mechanisms that cause neuronal death.

Conflict of Interest Statement

We declare that there are no conflicts of interest and this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Kimberley Duncan

Associate Professor Ian Blair

Dr Kelly Williams

Contents

List of Figures

List of Tables

1

Introduction

Amyotrophic Lateral Sclerosis (ALS) is an adult-onset, fatal neurodegenerative disease characterised by the degeneration of the upper and lower motor neurons. ALS is accompanied by cognitive impairment in around half of all cases including about 15% who are diagnosed with comorbid frontotemporal dementia (FTD) (**Al-Chalabi et al.**, 2012; **Robberecht and Philips**, 2013). Clinical features of ALS include muscle weakness and atrophy, fasciculations, and spasticity, which result in limb paralysis, and speech and swallowing difficulty. Disease progression is rapid, with most patients dying of respiratory failure caused by diaphragm paralysis within 3-5 years of diagnosis (**Robberecht and Philips**, 2013). The Australian incidence of ALS is 2.74 per 100,000 people (**Australian Institute of Health and Welfare**, 2011). Around 10% of cases show inheritance of disease and are classified as familial ALS (FALS); however the majority of cases occur sporadically (SALS) (**Al-Chalabi et al.**, 2012; **Robberecht and Philips**, 2013; **Renton et al.**, 2014). FALS and SALS are clinically indistinguishable and share the same hallmark ALS pathology, including ubiquitinated cytoplasmic inclusions of protein aggregates in motor neurons (**Neumann et al.**, 2006; **Arai et al.**, 2006).

The only known causes of ALS are gene mutations, which lead to motor neuron death (see table 1.1 for a summary of ALS genes mutated in familial ALS)(**Robberecht**

and Philips, 2013; Renton et al., 2014). The first ALS causative gene, superoxide dismutase (*SOD1*) (OMIM#147450) (Rosen, 1993), was identified in ALS families over twenty years ago. SOD1 mutations account for ~12% of FALS and ~1% of SALS respectively (Renton et al., 2014). The most frequently occurring gene mutation in ALS, a hexanucleotide repeat expansion in chromosome 9 open reading frame 72 (*C9ORF72*)(OMIM#614260) (DeJesus-Hernandez et al., 2011; Renton et al., 2011), was discovered in 2011 and accounts for ~40% of FALS and ~5% of SALS globally. Since 2008, many genes have been implicated as causative or associated with ALS, including genes producing proteins directly involved in overlapping homeostasis pathways, including RNA processing pathways and the protein degradation pathways (Ling et al., 2013). Dysfunction in RNA processing was first implicated as a potential pathogenic mechanism in ALS when it was found that most ALS patients have ubiquitinated neuronal cytoplasmic inclusions that are positive for TAR-DNA binding Protein 43 (TDP-43), a protein that plays a role in RNA processing (Arai et al., 2006; Neumann et al., 2006). Further support for the role of aberrant RNA processing was shown when ALS mutations were identified within genes that encode TDP-43 (OMIM# 605078) (Sreedharan et al., 2008), and another RNA binding protein, Fused in Sarcoma (*FUS*) (OMIM# 137070) (Vance et al., 2009; Kwiatkowski et al., 2009). Combined, these two genes account for around 5% of FALS cases (Renton et al., 2014). The cytoplasmic aggregation of TDP-43 is accompanied by a loss of TDP-43 from the nucleus, suggesting a combination of loss and gain-of-functions are at play (Lagier-Tourenne et al., 2012; Fecto and Siddique, 2011).

Dysfunction in protein degradation pathways could also explain ALS pathology, especially in ALS patients without mutations in genes intrinsic to RNA processing (Robberecht and Philips, 2013). Rare mutations have also been described in several genes involved in protein degradation, including ubiquilin 2 (*UBLQN2*) (OMIM#300264) (Williams et al., 2012; Deng et al., 2011), optineurin (*OPTN*) (OMIM#602432) (Maruyama et al., 2010) and valosin-containing protein (*VCP*) (OMIM#601023) (Johnson et al., 2010), with each accounting for less than 1% of FALS. The protein degradation pathways are responsible for the proper turnover of aggregation-prone proteins either via the ubiquitin-proteasome system (UPS) or autophagy (Bendotti et al., 2012). Small, short-lived aberrant proteins are tagged for destruction by the addition of poly-ubiquitin chains, which are recognised and bound by UPS chaperone proteins. These chaperones then unfold the attached proteins and transport them to the proteasome for enzymatic destruction

TABLE 1.1: **Known ALS gene mutations in families.**

	Gene Name	Gene locus	Mode of inheritance	Putative pathway involvement	Reference
Chromosome 9 open reading frame 72	*C9ORF72*	9p21.3p13.3	Dominant	RNA–binding and/or processing	(DeJesus-Hernandez et al., 2011; Renton et al., 2011)
V–Erb–B2 avian erythroblastic leukemia viral oncogene homolog 4	*ERBB4*	2q34	Dominant	RNA binding and/or processing	(Takahashi et al., 2013)
Ewing sarcoma breakpoint region 1	*EWSR1*	22q12.2	Unkown	RNA–binding and/or processing	(Couthouis et al., 2012)
Fused in sarcoma	*FUS*	16p11.2	Dominant	RNA–binding and/or processing	(Kwiatkowski et al., 2009; Vance et al., 2009)
Heterogeneous nuclear ribonucleoprotein A1	*HNRNPA1*	12q13.1	Dominant	RNA–binding and/or processing	(Kim et al., 2013)
Matrin 3	*MATR3*	5q31.2	Dominant	RNA–binding and/or processing	(Johnson et al., 2014)
Optineurin	*OPTN*	10p15p14	Dominant#	Proteostatic proteins	(Maruyama et al., 2010)
Profilin 1	*PFN1*	17p13.2	Dominant	Cytoskeleton/cellular transport deficits	(Wu et al., 2012)
Superoxide dismutase 1	*SOD1*	21q22.1	Dominant*	Enzyme	(Rosen, 1993)
Sequestosome	*SQSTM1*	5q35	Dominant	Proteostatic proteins	(Fecto et al., 2011)
Synovial Sarcoma Translocation Gene On Chromosome 18–Like	*SS18L1*	20q13.33	Unkown	RNA–binding and/or processing	(Chesi et al., 2013)
TATA–binding protein associated factor 15	*TAF15*	17q11.1q11.2	Unkown	RNA–binding and/or processing	(Couthouis et al., 2011)
TAR DNA–binding protein 43 (TDP–43)	*TARDBP*	1p36.2	Dominant	RNA–binding and/or processing	(Sreedharan et al., 2008)
Tubulin, alpha 4a	*TUBA4A*	2q35	Dominant	Cytoskeleton deficits	(Smith et al., 2014)
Ubiquilin 2	*UBQLN2*	Xp11	Dominant	Proteostatic proteins	(Deng et al., 2011; Williams et al., 2012)
Valosin–containing protein	*VCP*	9p13	Dominant	Proteostatic proteins	(Johnson et al., 2010)

* Recessive for p.D90A in Scandinavian populations

May be recessive in Japanese populations

(**Bendotti et al.**, 2012). Alternatively, larger cytoplasmic proteins, including neuronal cytoplasmic inclusions, are sequestered into vesicles enclosed within a double membrane, called an autophagosome, which subsequently fuses with a lysosome that contains protein degradative enzymes (**Bendotti et al.**, 2012). Several ALS-causing genes encode UPS or autophagy proteins, including *UBLQN2*, *OPTN* and *VCP* (see table 1.2)(**Robberecht and Philips**, 2013). Interestingly, most neurodegenerative diseases, including Parkinsons disease (PD) and Alzheimers disease (AD), are also characterised by the pathological aggregation of misfolded proteins (**Takalo et al.**, 2013).

Our laboratory recently identified a novel mutation in another UPS protein-coding gene, *CCNF*, segregating in an ALS family (unpublished data). International collaborative efforts have uncovered other mutations in *CCNF* in both FALS and SALS patients at a prevalence of 0.6 to 3.3% globally (see figure 1.1). These mutations are absent from 1831 Australian control individuals, as well as control SNP databases, including dbSNP134, the 1000 Genomes Project (<0.001 frequency, Nov 2010 release) and the NHLBI Exome Sequencing Project (ESP) exome variant server (6503 sequenced human exomes)(unpublished data) (table 1.2).

CCNF (accession # NM_001761) is a ubiquitously expressed 17-exon gene that encodes the 786 amino acid protein, cyclin F (NP_ 41002), with a molecular weight of 87kDa (**Bai et al.**, 1994). Cyclin F has two primary roles, based on its functional domains, as a cyclin protein and as an F-box protein. Most cyclins tightly regulate cyclin-dependent kinases (CDKs), which drive progression from one phase of the cell cycle to the next. Cyclin F differs from other cyclins in that it does not have an associated CDK so does not act in the cell cycle as a typical cyclin. F-box proteins are a family of proteins containing a highly-conserved 40 amino acid F-box motif (**Skaar et al.**, 2013). The F-box motif binds to a Skp1 adaptor protein, a central component of the Skp1-Cul1-F-box protein (SCF) complex (see figure 1.2). The SCF complex is an E3 ubiquitin ligase, which mediates the ubiquitination of proteins targeted for destruction by the ubiquitin-proteasome system (UPS) (see figure 1.3 for details of the ubiquitination reaction) (**D'Angiolella et al.**, 2013). To date, 69 F-box proteins have been identified in humans (**Skaar et al.**, 2013), and each one can bind multiple substrates, giving the SCF E3 ubiquitin ligases the facility to bind hundreds of proteins (**Winston et al.**, 1999). Many F-box proteins recognise a short degradation motif (degron) in substrates in a phosphorylation-dependent manner. However, cyclin F is unusual in that its recruitment of substrates depends only on

TABLE 1.2: **ALS-linked *CCNF* mutations.**

Amino acid change	Nucleotide change	ALS
S3G	c.7G>A	SALS
K97R	c.290A>G	FALS
S195R	c.585T>C	FALS
R392T	c.1175G>A	SALS
S509P	c.1525T>C	FALS
S621G	c.1861A>G	FALS
E624K	c.1870G>A	SALS

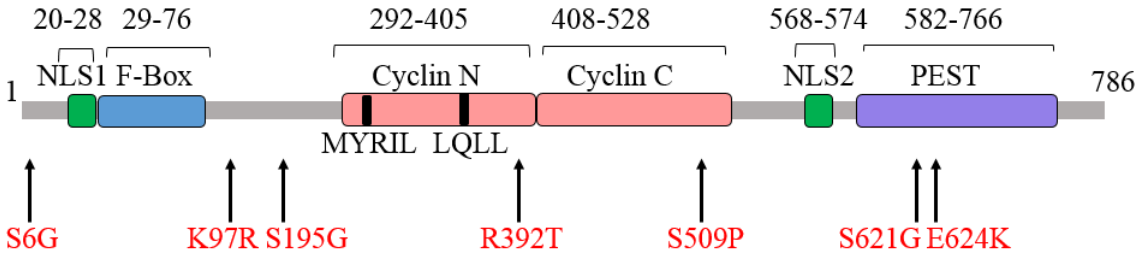


FIGURE 1.1: **A schematic of the ALS-linked mutations found in the cyclin F protein.** The hydrophobic patches responsible for binding substrates are shown in black with amino acid sequence underneath. NLS1: nuclear localisation signal 1; F-box: F-box motif; NLS2: nuclear localisation signal 2: PEST: an amino acid sequence rich in proline, glutamic acid, serine and threonine amino acids bordered by positively charged residues.

the presence of an RXL motif in the substrate (**D'Angiolella et al.**, 2012, 2010). It is speculated that the regulation of substrate binding is achieved via the control of cyclin F sub-cellular localisation and stability (**Skaar et al.**, 2013). Cyclin F is the only known F-box protein containing a cyclin box domain.

To date, only a handful of substrates have been identified for cyclin F, including proteins involved in centrosome duplication, CP110 (**D'Angiolella et al.**, 2010); microtubule assembly, NuSAP1 (**Emanuele et al.**, 2011); and dNTP production for DNA synthesis and repair, RRM2 (**D'Angiolella et al.**, 2012) (table 1.3). More recently, cyclin F has been implicated in G2/M checkpoint control through its interaction with the oncogene B-Myb although how cyclin F regulates this protein is uncertain (**Klein et al.**, 2015) (details in Table 1.3). While the exact role of cyclin F

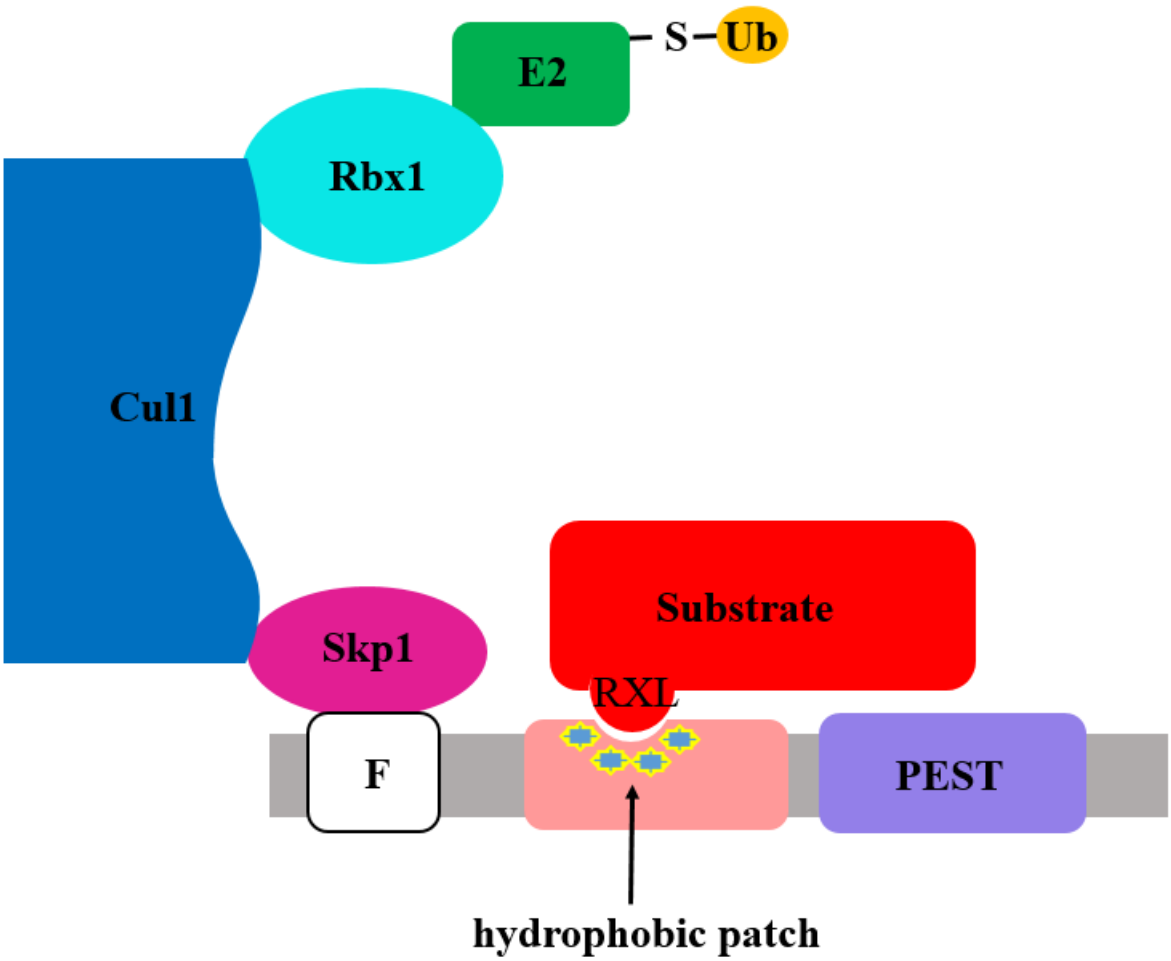


FIGURE 1.2: .

The Skp1-Cul1-F-box E3 ubiquitin ligase machinery. The SCF complex is comprised of a Cullin 1 (Cul1) protein scaffold, with a ring protein (Rbx1) attached to its C-terminus and a Skp1 adaptor/F-box complex attached to its N-terminus. The RBX1 protein recruits the E2 ubiquitin conjugating enzyme responsible for transferring the ubiquitin molecule to the substrate. The Skp1 protein binds to cyclin F through shared F-box motifs. Cyclin F binds to an RXL motif in the substrate via a hydrophobic patch in the cyclin box. F: F-box motif; Ub: ubiquitin molecule; Cul1: cullin 1 protein: Rbx1: ring box protein 1; Skp1: S-phase kinase-associated protein 1. Figure adapted from **Skaar et al.** (2013)

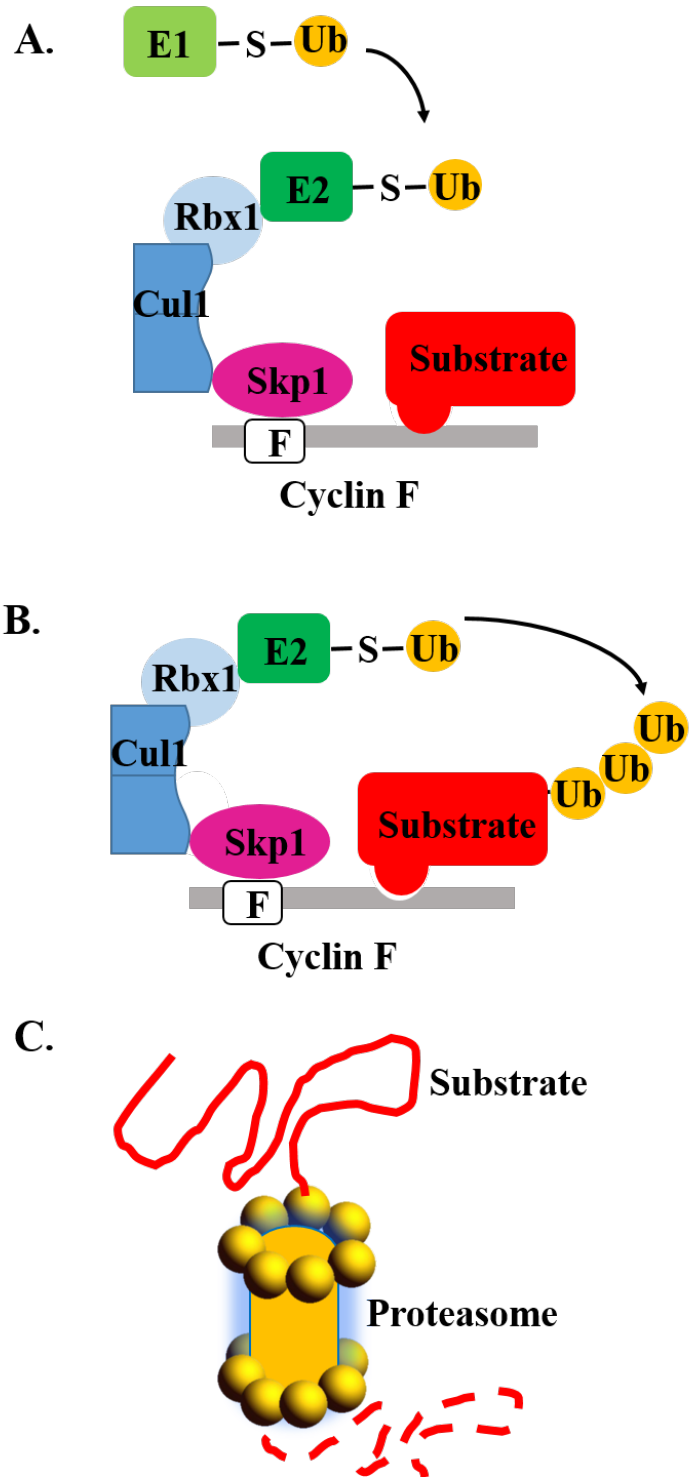


FIGURE 1.3: .

The role of cyclin F in the ubiquitination pathway. **(A.)**The E1 ubiquitin activating enzyme binds to a ubiquitin molecule via a thiolester bond. It transfers the ubiquitin molecule to E2 ubiquitin conjugating enzyme, which is attached to the E3 ubiquitin ligase. **(B.)** The Skp1-Cul 1-F-Box protein (SCF) E3 ubiquitin ligase mediates the transfer of the ubiquitin to a lysine residue in the substrate. **(C.)** Once the substrate is ubiquitinated, it is ferried to the ubiquitin proteasome for proteolytic destruction. Figure based on information from **Skaar et al.** (2013)

in the cell cycle is still to be fully described, E3-ubiquitin ligases are often employed in the rapid and irreversible degradation of cell cycle checkpoint regulators, such as cyclins. This ensures efficient one-way progression through the cell cycle (**Bai et al.**, 1996).

The mechanism by which mutated cyclin F causes ALS or FTD is currently unknown. Motor neurons are terminally differentiated and therefore do not undergo cell cycling; yet it has been shown that SCF family proteins are also expressed in post-mitotic cells (**Staropoli**, 2008). Furthermore, recent transcriptomic studies have demonstrated that cell cycle proteins are expressed in the brains of patients with sporadic ALS (**Aronica et al.**, 2015). Interestingly, another F-box E3 ligase, F-box protein 7 (FBX07) has been linked to juvenile Parkinsons disease (**Zhao et al.**, 2011; **Lohmann et al.**, 2015), and FBX07 has been shown to be highly expressed in the motor areas of the brain (**Zhao et al.**, 2011). Research conducted in our laboratory indicates that mutant cyclin F protein leads to dysregulation of the UPS, with increased cellular levels of ubiquitinated proteins (unpublished data). Further investigation of the function of cyclin F protein in motor neurons could help to elucidate its role in ALS.

Despite the gene discoveries in ALS over the last 20 years, disease mechanisms in ALS are poorly understood and there are limited treatment options (**Robberecht and Philips**, 2013). Therefore, characterising the effect of gene mutations on protein function is key to understanding the pathological mechanisms of the disease and to uncovering potential therapeutic targets. However, predicting the effect of disease-causing gene mutations on protein function is challenging, especially when little is known about the protein itself. Most proteins are comprised of domains, each with a unique function, and defining those functions is the first step in predicting the effect of specific mutations on protein behaviour. Three dimensional (3D) structural analysis of proteins using techniques such as x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have provided invaluable insight into protein structure and function relationships (**O'Donoghue et al.**, 2010). Amino acids that are far from active sites in 2D space can be very close in 3D space. Therefore, determining the 3D structure of a protein can aid in determining the effect of amino acid substitutions due to genetic mutations on protein function and stability. However, because 3D structural analysis techniques are technically difficult, highly specialised, and not suitable for all proteins, only a very small percentage of known protein sequences have an experimentally solved 3D structure (**Lander et al.**, 2012).

TABLE 1.3: **Interaction of cyclin F and substrates**

Substrate		**Interaction**		Function of cyclin F		
Substrate	**Function**	**Phase**	**Location**	**Mode of action**	**Result**	**Reference**
Cyclin B1	Regulation of G2-M transition	G2-M	Nucleus/Cytoplasm	Interacts with Cyclin B1 via CRM	Nuclear localisation	(**Kong et al.**, 2000)
CP110	Induces centrosome duplication	G2	Centrosomes	Ubiquitin-mediated proteolysis	Restricts replication to once/cycle	(**D'Angiolella et al.**, 2010)
NuSAP	Microtubule-binding protein	Unknown	Unknown	Unknown		(**Emanuele et al.**, 2011)
RRM2	DNA repair and replication	G2	Nucleus	Ubiquitin-mediated proteolysis	Controls degradation of RRM2 to prevent genomic instability	(**D'Angiolella et al.**, 2012)
B-Myb	Transcriptional activator G2-M transition	unknown	unknown	Possibly outcompetes Cyclin A to bind b-Myb to prevent activation via phosphorylation	Prevents damaged cells entering mitosis	(**Klein et al.**, 2015)

CRM: cytoplasmic retention motif
CP110: centriolar coiled-coil protein of 110 kDa
NUSAP1: nucleolar and spindle-associated protein 1
RRM2: ribonucleotide reductase M2
b-Myb: Myb-related protein B

More recently, bioinformatics tools have expanded our ability to determine 3D structure using *in silico* homology modelling and fold prediction software (**O'Donoghue et al.**, 2015; **Remmert et al.**, 2012; **O'Donoghue et al.**, 2010). These techniques can be divided into three categories: 1) Comparative modelling; 2) Threading; and 3) *Ab initio* modelling (**Roy et al.**, 2010). Comparative modelling involves matching a query sequence to an evolutionarily-conserved homologue. High-resolution models of the query sequence are then built by copying the backbone of the template structures, and adding loops and side-chains (**Arnold et al.**, 2006). Threading involves aligning a query sequence onto the 3D structures of solved proteins with similar folds. Unlike comparative modelling, these templates do not have to be related to the original query sequence. Factors such as solvent accessibility, secondary structures and interactions with neighbouring amino acids are taken into account when building the model (**Roy et al.**, 2010). *Ab initio* modelling is used for query sequences with no homologues, and predicts structure from scratch, but is only useful with short sequence lengths (<120 amino acids) (**Roy et al.**, 2010).

Although 3D structures have been solved for several cyclin proteins, including cyclin A2 (**Carbain et al.**, 2014) and cyclin B1 (**Brown et al.**, 2015), none have yet been constructed for cyclin F. Therefore, in the first part of the current study, I aim to use bioinformatic software to identify the 3D structure of the wild-type cyclin F and to determine the impact of ALS-linked mutations on this structure. I will use three different 3D modelling programs to find the best model of cyclin F: SWISS-MODEL (**Arnold et al.**, 2006), which uses comparative modelling; Aquaria (**O'Donoghue et al.**, 2015), which uses a threading approach; and I-TASSER, which uses a composite approach combining threading with *ab initio* techniques (**Roy et al.**, 2010). In addition, I will determine the changes in predicted phosphorylation patterns using NetPhos 2.0 software, as phosphorylation can play an important role in protein regulation, activation and localisation (**Blom et al.**, 1999).

The function of individual protein domains can also be validated using standard laboratory techniques including knockout models, site-directed mutagenesis and deletion cloning. Deletion cloning has been particularly informative for the study of neurodegenerative diseases, including ALS, FTD, Alzheimers disease and Parkinsons disease. It has been used to define the domain structure and associated functions in several disease-specific proteins. For example, deletion constructs of the ALS/FTD-linked TDP-43 have been generated and expressed in human and mouse cell lines (**Nonaka et al.**, 2009); primary neurons (**Yang et al.**, 2010);

Drosophila melanogaster (**Feiguin et al.**, 2009); *Caenorhabditis elegans* (**Ash et al.**, 2010); and murine models (**Caccamo et al.**, 2012). A range of functional assays, including immunofluorescence confocal microscopy, co-immunoprecipitation and immunoblotting were subsequently conducted to facilitate the discrimination of specific residues involved in sub-cellular localisation and aggregation propensity (**Nonaka et al.**, 2009; **Yang et al.**, 2010; **Ash et al.**, 2010); protein-protein interactions (**Nonaka et al.**, 2009; **Liu-Yesucevitz et al.**, 2010); and protein expression and cell viability (**Nonaka et al.**, 2009; **Ash et al.**, 2010), respectively.

Functional assays involving the *in vitro* expression of cyclin F with deletions or mutations (using expression constructs) have provided some insight into domain structure and function (**Klein et al.**, 2015; **D'Angiolella et al.**, 2010, 2012; **Bai et al.**, 1994; **Kong et al.**, 2000; **Fung et al.**, 2002) (figure 1.4). Cyclin F is composed of three distinct domains: an F-box domain, a cyclin domain, and a PEST sequence (a stretch of proline, glutamic acid, serine and threonine amino acids bordered by positively charged residues) (figure 1.4A). The F-box domain is responsible for binding to the E-3 ubiquitin ligase machinery via the Skp1 adaptor protein (figures 1.2 & 1.4A) (**Fung et al.**, 2002). The cyclin domain is divided into two segments, cyclin N and cyclin C (figure 1.4A). The cyclin N region contains a hydrophobic patch that enables binding to the substrate, similar to the way cyclin A and cyclin B bind their CDK partners (**D'Angiolella et al.**, 2010, 2012). The PEST sequence is responsible for destabilisation of the protein to ensure proper protein turnover during the cell cycle (**Fung et al.**, 2002) (figure 1.4A). Cyclin F also contains two nuclear localisation signals; one in the N terminal region (NLS1) and one between the cyclin box and the PEST sequence (NLS2) (**Kong et al.**, 2000). Our laboratory and collaborators have identified ALS-linked mutations that could potentially disrupt two of the three protein domains, the cyclin box and the PEST sequence (unpublished data) (figure 1.1).

The aforementioned functional studies have provided insight into the stability (table 1.4), subcellular localisation (table 1.5) and binding behaviour (table 1.6) of cyclin F. However, these studies used expression constructs that were designed to truncate either the C terminal or N terminal of the protein rather than the defined removal of whole domains (figure 1.4B-E). In addition, all of these assays were performed in non-neuronal cell lines, such as HeLa and Cos1 (see tables 1.4,1.5 & 1.6), which have very different characteristics to neuronal cells (**Wu et al.**, 2013a).

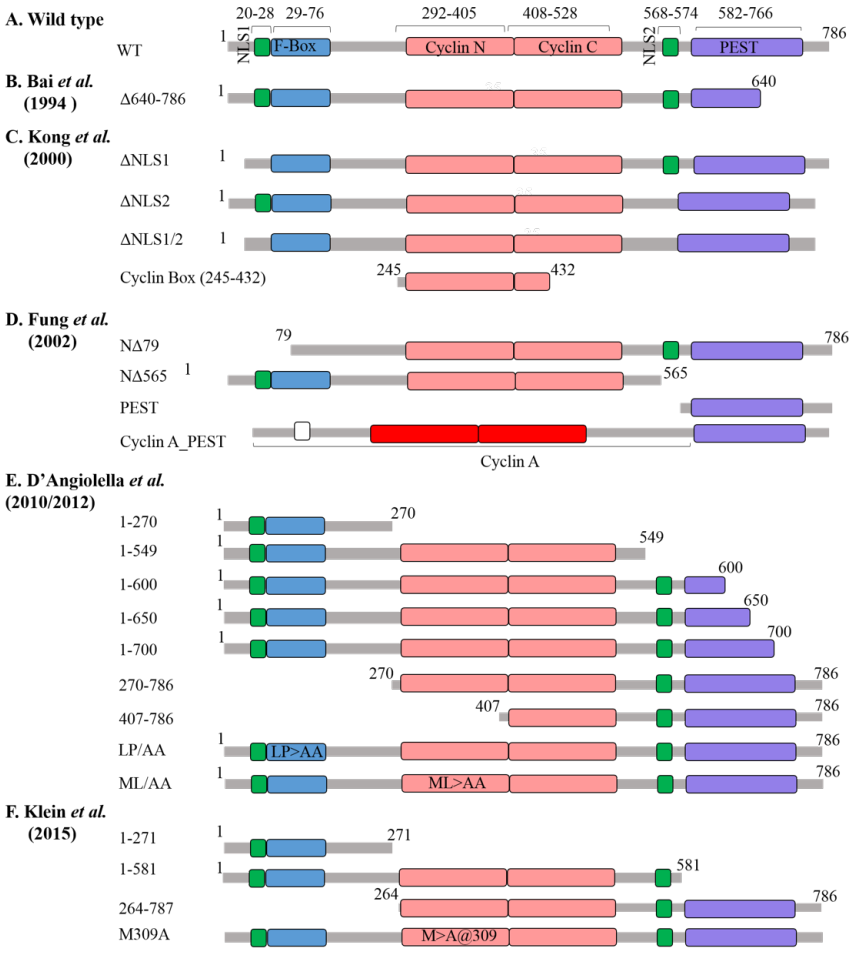


FIGURE 1.4: **Schematic of deletion and substitution mutants of cyclin F constructs used in previous studies** (Bai et al., 1994; Fung et al., 2002; Kong et al., 2000; D'Angiolella et al., 2010, 2012; Klein et al., 2015). Sequences are aligned by the cyclin domain. (A.) The functional domains of wild type cyclin F. (B.) Δ640-786: lacks part of PEST. (C.) ΔNLS1: lacks NLS1. ΔNL2: lacks NLS2. ΔNLS1/2 lacks both NLS regions. Cyclin box (245-432) was generated from mouse DNA. (D.) NΔ79 lacks the F box & NLS1; NΔ565 lacks the PEST & NLS2. PEST is the isolated PEST sequence. Cyclin A_PEST is a cyclin A gene fused to the isolated PEST sequence. (E.) 1-270 lacks the cyclin box & PEST; 1-549 lacks the NLS2 & PEST; 1-600, 1-650 & 1-700 lack partial PEST sequences. 270-786 lacks the NLS1 & F box; 407-786 lacks the NLS1, F box & cyclin N. Two constructs were generated with mutations rendering them dysfunctional: LP/AA: L37A & P38A in the F-box; & ML/AA: M309A & L313A in the hydrophobic patch. (F.) 1-271 lacks the F box; 1-581 lacks the PEST; 264-787 lacks the NLS1 & F box. One mutation was generated in the hydrophobic patch rendering it non-functional,p.M309A. NLS: Nuclear localisation signal.

TABLE 1.4: **Protein stability studies using deletion clones of Cyclin F**

Study	Construct	Vector	Cell line	Functional study	Result
(**Bai et al.**, 1994)	Δ640-786	CD20	COS-1	Flow cytometric analysis of DNA content in transfected cell lines	Significant increases in G2/M population compared to WT
			C33A cells	Microscopy of propidium iodide stained cells	No difference to WT cells. G2 population has increased.
(**Fung et al.**, 2002)	NΔ79	pET21d	HtTA1 cells	*in vitro* degradation assay reticulocyte lysate system	No change in stability
	NΔ565	pET21d			Increased stability
	PEST	pET21d			Stabilised endogenous Cyclin F
	Cyclin A_PEST	pET21d			Decreased stability

Constructs are defined in figure 1.4
HtTA1 cells: HeLa cells expressing tTA tetracycline repressor chimera

Table 1.5: **Sub-cellular localisation studies using deletion clones of Cyclin F**

Study	Construct	Vector	Cell line	Functional study	Result
(**Bai et al.**, 1994)	Δ640-786	CD20	C33A cells	Microscopy of propidium iodide stained cells	No difference to WT cells. G2 population has increased.
(**Kong et al.**, 2000)	ΔNLS1	pVP16 plasmid	COS-1	Immunofluorescence to detect subcellular localisation	50% more cytoplasmic localisation than WT
	ΔNLS2				50% more cytoplasmic localisation than WT
	ΔNLS1/NLS2				90% cytoplasmic localisation
(**D'Angiolella et al.**, 2010, 2012)	1-270	FLAG-tagged	U2OS	Fluorescent/Confocal microscopy	Centrosomal localisation
	1-549				Centrosomal localisation
	1-600,1-650,1-750 LP/AA mutant & ML/AA mutant				Centrosomal & nuclear localisation
	407-786				nuclear localisation

Constructs are defined in figure 1.4

To remove sequences of DNA from the internal regions of genes, as **Kong et al.** (2000) achieved with the ΔNLS1 and ΔNLS2 constructs, PCR-driven site-directed mutagenesis can be utilised (Figure 1.4). Most of the frequently used commercial PCR-driven mutagenesis kits are based on one of three methods: PCR-driven overlap extension (**Ho et al.**, 1989), inverse PCR (*i*PCR) (**Imai et al.**, 1991) or circular mutagenesis (**Hemsley et al.**, 1989), although overlap PCR has been largely overtaken by the latter two methods. The processes of circular mutagenesis and inverse PCR are summarised in Figures 1.5 and 1.6

Although circular mutagenesis has been used in the generation of deletion clones for several ALS-linked genes (**Sun et al.**, 2015; **Winton et al.**, 2008; **Ash et al.**, 2010), there are limitations. The most serious limitation is that overlapping mutagenic primers provide a better match to each other than the template, resulting in the formation of excessive primer dimers (**Edelheit et al.**, 2009; **Liu and Naismith**, 2008; **Zheng et al.**, 2004). This is especially problematic in the creation of large deletion fragments, as the primers must skip over large segments of the template DNA thereby hindering the rate of binding (see Figure 1.5). Although several strategies have been developed to overcome these limitations, deletions using the circular mutagenesis method are restricted to lengths of just 25 nucleotides before loss of efficiency (**Zheng et al.**, 2004; **Qi and Scholthof**, 2008). In contrast, the primers used in *i*PCR-driven deletion cloning do not overlap, as they are separated by the region targeted for deletion (Figure 1.6). The *i*PCR-driven deletion cloning strategy allows for the omission of much larger segments of DNA than in circular mutagenesis (**Imai et al.**, 1991). The mutant linear strands generated by the first round of PCR are used as templates for subsequent rounds of PCR, which increases amplification of the mutant strands (Figure 1.6). Therefore, *i*PCR is preferable to circular mutagenesis for the current study because of its ability to generate larger deletions and its increased amplification potential. It can also be designed to be specific to the desired target.

Table 1.6: **Binding studies using deletion and mutant clones of cyclin F**

Study	Construct	Vector	Cell line	Functional study	Result
(**Kong et al.**, 2000)	245-435(Cyclin)	pBTM116	Yeast	Yeast two-hybrid screen with cyclin B1 CRM derivatives	The cyclin box of cyclin F interacts with the CRM region of cyclin B1
(**Fung et al.**, 2002)	NΔ79	pET21d	HtTA1	Co-Immunoprecipitation with Skp1	Does not form a complex with Skp1. Therefore, Fbox responsible for forming SCF complex
(**D'Angiolella et al.**, 2010, 2012)	1-549 (ΔNLS2 & PEST)	FLAG-tagged	HEK293	Immunoprecitation with anti-FLAG resin. Immunocompexes probed with CP110 & RRM2 antibodies	Binds CP110 but not RRM2
	1-600; 1-650; 1-750 (partial PEST deletions)				Binds CP110 & RRM2
	1-270(ΔCyclin & PEST) 407-786 (ΔCyclinN, Fbox & NLS1) MA/LA (Cyclin mutant)				Binds neither CP110 nor RRM2
	LP/AA (Fbox mutant)				Increased binding CP110 & RRM2. reduced binding Skp1 & Cul1.
(**Klein et al.**, 2015)	1-271 (ΔCyclin)	pCMV Minkio HA-FLAG	HEK293T	Co-transfection with GFP-B-Myb Co-IP & immunoblotting with HA & GFP	Does not bind B-Myb
	1-581 (ΔNLS2 & PEST); 264-787 (ΔFbox & NLS1)				Binds B-Myb
	MR-AA mutant				Reduced binding of bMyb

Constructs are defined in figure 1.4

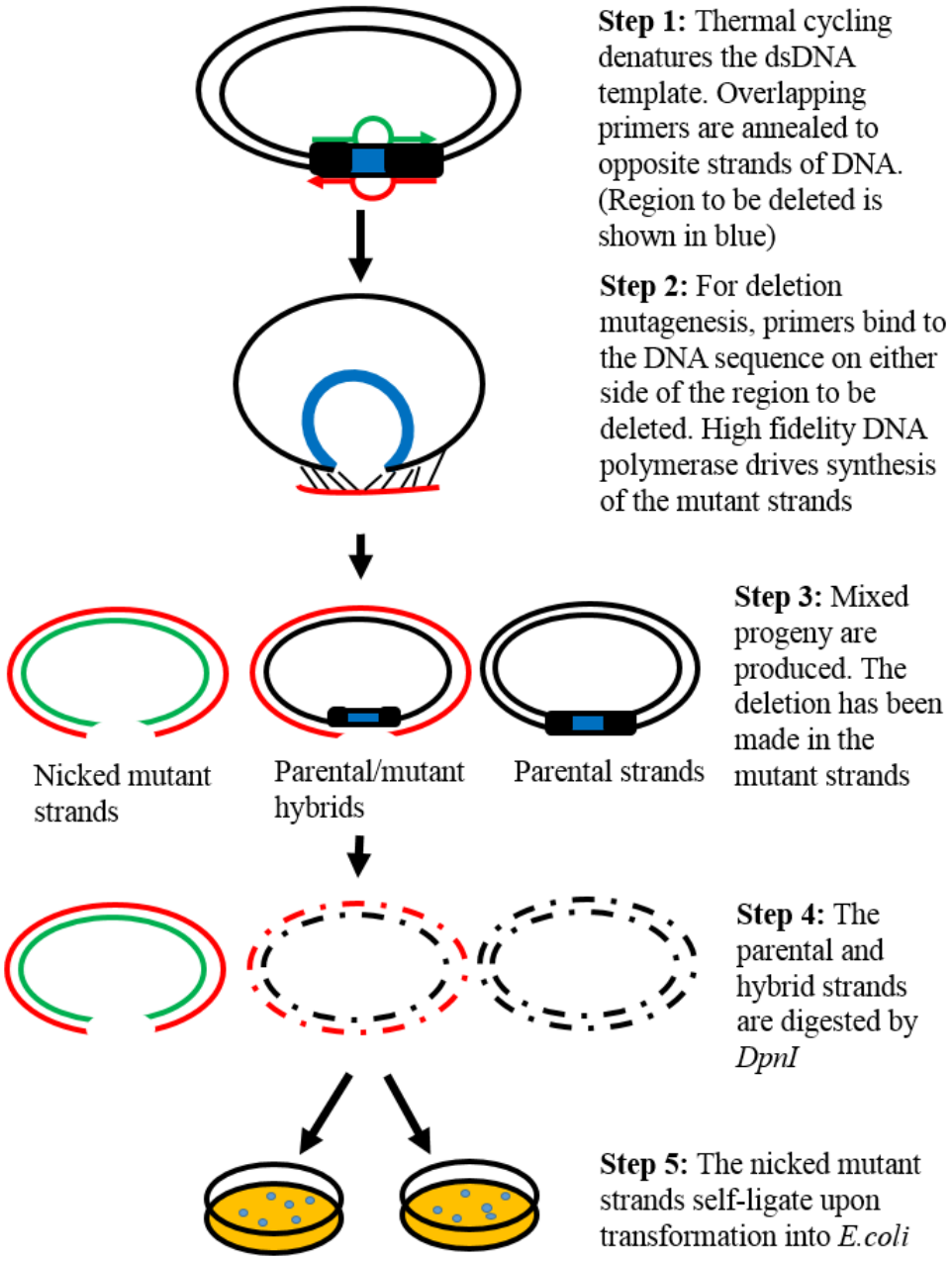
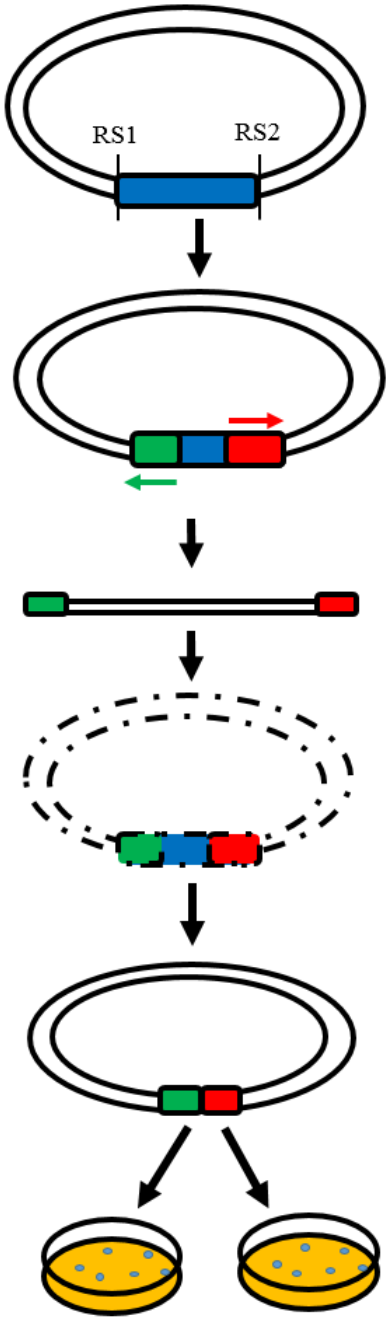


FIGURE 1.5: **Deletion cloning via circular PCR mutagenesis .**



Step 1: The cDNA of interest is cloned into a cloning vector using restriction enzymes

Step 2: Back-to-back primers bind to opposing target DNA strands on either side of the region to be deleted (shown in blue).

Step 3: The linearised mutant is amplified via PCR using a high fidelity polymerase

Step 4: *DpnI* digestion of the methylated parental plasmid DNA allows the selection of the unmethylated PCR products

Step 5: The PCR products self-ligate with the addition of T4 polynucleotide kinase and T4 DNA ligase

Step 6: The mutant strand are transformed into highly competent *E.coli* cells

FIGURE 1.6: **Deletion cloning via *i*PCR mutagenesis.**

Table 1.7: ***CCNF* deletion constructs.**

Construct design	**Amino acids deleted**	**Description**
ΔPEST	581-766	PEST region deleted
ΔCyclinN	292-405	Cyclin N region deleted
ΔCyclinC	408-528	Cyclin C region deleted
ΔCyclinBox	292-528	Full Cyclin Box deleted
Δ262-766	292-766	Cyclin box to PEST deleted
ΔNLS2	568-574	Removal of the NLS2 region

Therefore, in the second part of the current study, I aim to use *i*PCR mutagenesis for the precise deletion of internal protein domains in cyclin F (Table 1.7). Deletion clones will be designed to remove various combinations of the the two functional domains that are mutated in the ALS cohorts: cyclin N, cyclin C, the whole cyclin box, the PEST sequence or both the cyclin box and PEST sequence. An NLS2 knockout will also be created to be used as a reference for cytoplasmic staining as cyclin F without the NLS2 has been shown to localise to the cytoplasm 50% more frequently than wild-type cyclin F (**Kong et al.**, 2000). The deletion clones will be sub-cloned into a pmCherry-C1 vector, which carries an mCherry sequence at the 5′ end of the gene insert, facilitating visualisation via fluorescent microscopy and flow cytometry. The deletion clones will be expressed via transfection in neuronal cell lines, to assess cellular phenotypic changes and identify the neuronal function of those domains. Characterising the structure and function of cyclin F in neuronal cell lines will provide the underlying basis for elucidating the pathogenic mechanisms behind mutant *CCNF*-linked ALS and FTD. Others in our laboratory are currently working on *CCNF* point-mutation constructs. Results comparisons of the deletion constructs from the current study and these point-mutation constructs will likely lead to publication.

2
Materials

2.1 Materials

2.1.1 Reagents and buffers

1×TBE (Tris–Borate–EDTA) buffer
89mM Tris; 89mM boric acid; 2mM EDTA (pH 8.3).

1×TE(Tris-EDTA)buffer (Invitrogen, USA)
10 mM Tris-HCL (pH 8.0) & 0.1 mM EDTA.

SOC medium(Bioline, Australia)
2% tryptone; 0.5% yeast extract; 10mM NaCl; 2.5mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; & 20mM glucose.

Luria Broth (LB)
10g/L Bacto–tryptone; 5g/L Bacto–yeast extract; and 5g/L NaCl. LB was autoclaved before use.

LB agar

15g/L Davis Agar dissolved in LB.

T4 DNA ligase (Promega, USA)
T4 DNA ligase in buffer containing 10mM Tris–HCl (pH 7.4 at 25°C), 50mM KCl, 1mM DTT, 0.1mM EDTA and 50% glycerol.

2×Ligation buffer (Promega, USA)
60mM Tris–HCl (pH 7.8); 20mM $MgCl_2$; 20mM DTT; 2mM ATP; and 10% polyethylene glycerol.

10×NEBuffer 3.1 (NEB, USA)
1×components 100mM NaCl; 50mM Tris-HCl; 10mM $MgCl_2$; 100g/ml BSA. (pH7.9@25C).

IPTG (Isopropyl–1–thio–B–D–galactopyranoside)
1.2g IPTG dissolved in 50mL sterile water. Solution was filter–sterilized and stored at 4°C.

X–Gal (5–Bromo–5–chloro–3–indolyl–B–D–galactopyranoside)
100mg/mL 5–bromo–4–chloro–3–indolyl–B–D–galactoside in N,N'–dimethyl–formamide. X–Gal solution was stored in the dark at –20°C.

Ampicillin (Bioline, Australia)
100mg/mL in H_2O.

Kanamycin (Bioline, Australia)
100mg/mL in H_2O.

2.1.2 Growth media and reagents for tissue culture

1× Gibco PBS (Phosphate Buffered Saline)(Life Technologies, USA)
1.5mM KH_2PO_4, 155.2 mM NaCl and 2.7 mM Na_2HPO_4-$7H_2O$. (pH 7.2).

Trypsin-EDTA (Sigma-Aldrich, USA)
0.25%, 2.5g porcine trypsin and 0.2g EDTA.

Gibco Dulbeccos Modified Eagles Medium (DMEM) (Life Technologies, USA).

Gibco Dulbeccos Modified Eagles Medium: Nutrient Mixture F-12 (DMEM/F12) (Life Technologies, USA)
A 1:1 mixture of DMEM and Hams F-12s.

Penicillin-Streptomycin (Sigma-Aldrich, USA)
Solution stabilized, with 10,000 units penicillin and 10 mg streptomycin/mL, kept at -20C.

Growth medium for NSC-34 cell line
DMEM supplemented with 10% Foetal bovine serum (FBS) and 1% Penicillin-Streptomycin.

Growth medium for SH-SY5Y cell line
DMEM/F12 with 10% FBS and 1% Penicillin- Streptomycin.

2.2 Vectors

2.2.1 pGEM-T (Promega, USA)

(Figure 2.1)
The pGEM-T vector is a 3kb long linearised version of the pGEM-5Zf(+) Vector (GenBank Accession No. 65308. It has an EcoRV site added at base 51, which is lost upon ligation of the insert into the vector. It also has T-overhangs at both ends, which allow for the efficient ligation of PCR products with A-overhangs, while preventing re-circularisation of the plasmid vector. Two universal promoters (T7 and SP6) flank a multiple cloning site. When a DNA fragment is cloned into the multiple cloning site, it disrupts the α–peptide coding region of the enzyme β–galactosidase, which along with the presence of a LacZ gene facilitates blue/white selection of recombinants on LB Agar plates supplemented with IPTG and X-gal. The vector also contains an ampicillin resistance gene for antibiotic selection to prevent overproduction of undesired colonies.

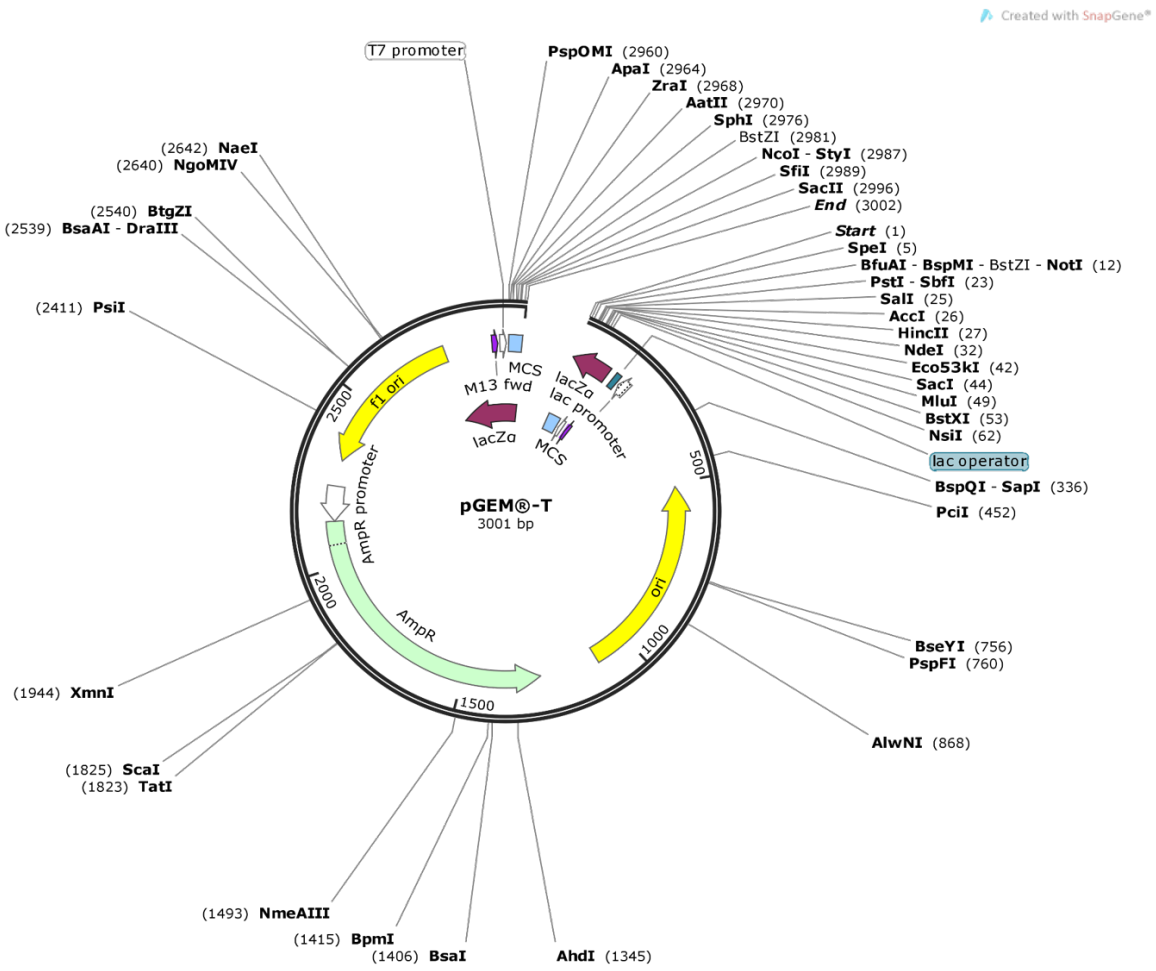


FIGURE 2.1: **Vector map of pGEM–T vector from Promega.**

2.2.2 pmCherry-C1 (Clontech, USA)

(Figure 2.2)
pmCherry-C1 is a 4.7kb mammalian expression vector carrying sequence encoding the mCherry fluorescent protein upstream of the multiple cloning site. Upon cloning the gene of interest, the resultant fusion protein includes an N terminal mCherry tag that enables protein expression to be monitored in flow cytometry assays and fluorescence microscopy. mCherry is excited at a maximum of 587 nm, and emits fluorescence at 610 nm. A bacterial promoter (PKanr) upstream of a Neomycin-resistance cassette confers kanamycin resistance in *E.coli*. Expression of the gene insert is controlled with the CMV promoter located directly upstream of the mCherry sequence. A SV40 polyadenylation signal located downstream of the multiple cloning site is responsible for the processing of the 3 prime end of the mCherry/insert fusion mRNA.

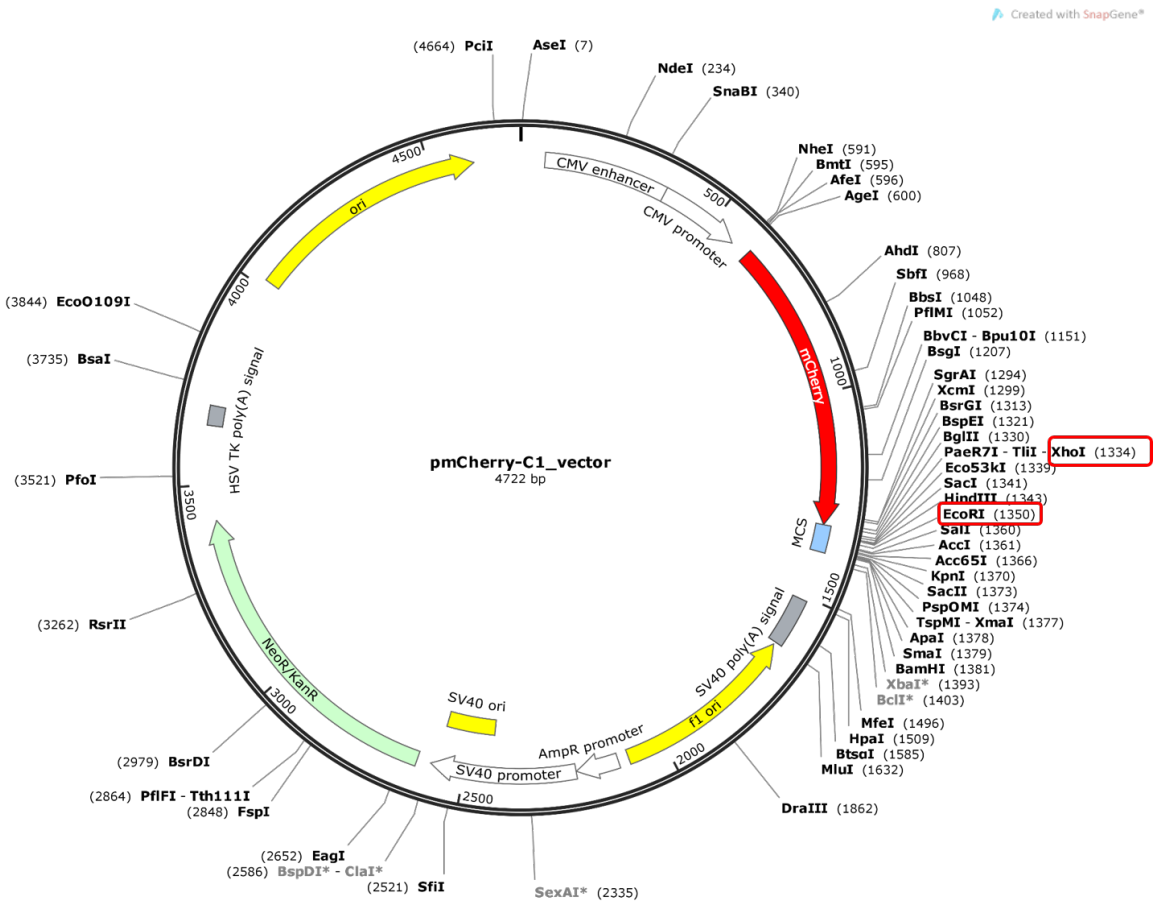


FIGURE 2.2: **Vector map of pmCherry-C1 from Clontech.**

3

Methods

3.1 3D structural analysis of cyclin F

Human cyclin F (*CCNF*) cDNA (NM_001761) and amino acid (NP_41002) sequences were obtained from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/). Searches for cyclin F homologue template sequences were conducted using the Position-Specific Iterated Basic Alignment Search Tool (PSI-BLAST) (**Altschul et al.**, 1997) against homologous Protein Data Bank (PDB) templates in the Protein Data Bank database (http://www.rcsb.org/) (**Berman et al.**, 2000) as an independent validation of the different 3D modelling programs to follow. The Protein Data Bank database is the repository of all experimentally determined 3D structures of large molecules, including proteins and nucleic acids (**Berman et al.**, 2000). Each 3D structure is given a unique identifying PDB file number.

3.1.1 Comparative modelling of cyclin F using SWISS-MODEL

Searches for cyclin F homologues were conducted using the inbuilt homology modelling algorithms in the 3D modelling program in SWISS-MODEL (**Arnold et al.**, 2006; **Bordoli et al.**, 2009). The FASTA sequence of the full-length human cyclin F was uploaded to the SWISS-MODEL server, which utilises both PSI-BLAST and HHBlits for sequence alignment with templates in the SWISS-MODEL template library (SMTL). The quality of the resulting structural models are assessed using the Qualitative Model Energy Analysis (QMEAN4) scoring function (**Arnold et al.**, 2006). A QMEAN4 value is derived using analyses of various structural elements such as torsion angles over three consecutive amino acids; long-range interactions; and solvation energy (**Benkert et al.**, 2008). The two top-ranked proteins structures were chosen for modelling.

3.1.2 Threading alignment of cyclin F using Aquaria

The *CCNF* gene was also searched for in the web-faced interface of Aquaria (**O'Donoghue et al.**, 2015). In Aquaria, the target sequences are searched for in SWISS-PROT (**Bairoch and Apweiler**, 1996) and Protein Data Bank databases (**Berman et al.**, 2000) using an inbuilt alignment tool based on HHBlits, which utilises iterative comparisons of hidden Markov models (HMMs) (**O'Donoghue et al.**, 2015; **Remmert et al.**, 2012). Homologues are ranked in order of sequence identity, and quality estimations are indicated by an HHBlits E-value, a measure of the number of times that the same degree of sequence similarity would occur randomly (**O'Donoghue et al.**, 2015). The top ranked template was modelled in Aquaria.

3.1.3 3D modelling of wild type and mutant cyclin F using the Iterative Threading ASSEmbly Refinement (I-TASSER) method

Unlike SWISS-MODEL and Aquaria, I-TASSER uses a composite approach, combining threading and *ab initio* techniques for 3D modelling (**Roy et al.**, 2010) (See figure 3.1 for an overview). Aligned and unaligned segments of the target protein can be assembled into a full-length sequence, and refined to remove steric clashes and find the lowest energy conformation (**Roy et al.**, 2010). The estimated quality of the

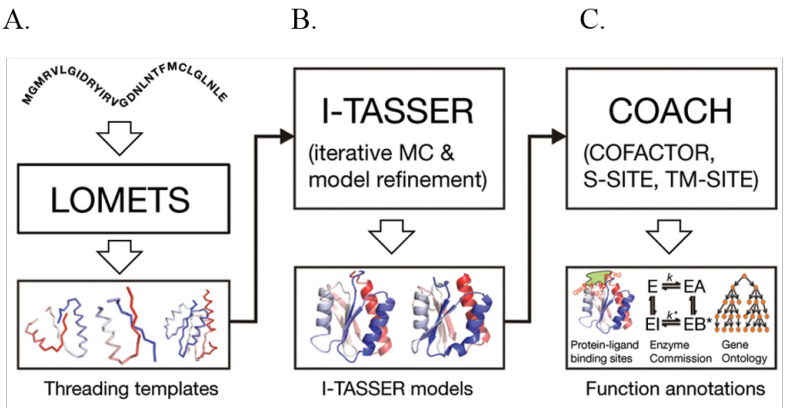


FIGURE 3.1: **Protein structure prediction pipeline used by I-TASSER. (A.)** Threading: The query sequence is aligned to homologues in a database containing 3D protein structures. The threading method, LOMETS, uses eight different fold recognition programs to find template sequences. The template sequences are ranked, and the best sequences are chosen for further consideration. **(B.)** Sequences are fragmented. Aligned and unaligned segments of the target protein are then re-assembled into a full length sequence. The unaligned segments are constructed using *ab initio* methods. The structure fold and reassembly are refined using the Monte Carlo (MC) and other refinement methods to remove steric clashes and find the lowest energy conformation (**Roy et al.**, 2010). **(C.)** Three different algorithms (COFACTOR, TM-SITE & S-SITE) are used to carry out functional analysis of the structure models with the highest confidence scores. Functional analysis includes exploration of the ligand-binding site (LBS), Enzyme Commission (EC) and Gene Ontology (GO) information. (Figure adapted from **Zhang et al.** (2015))

predicted 3D structure is assessed by a confidence score (C-score) and a threading alignment (TM-Score) that estimates the level of similarity between the query and template sequences. A C-score of $>$ -1.5 and TM score $>$ 0.50 are indicative of a structure of correct topology (**Roy et al.**, 2010). As I-TASSER was created for modelling single-domain proteins, it is suggested that each domain of a multi-domain protein be modelled individually to gain more confident results (**Roy et al.**, 2010). Therefore, both the full length cyclin F amino acid sequence (NP_41002), and separate batches of the F-box, cyclin box and PEST domain sequences were submitted to I-TASSER (**Roy et al.**, 2010) for the prediction of high resolution 3D structures. Domain boundaries were determined using PFAM (**Finn et al.**, 2014) and UniProt (**Bateman et al.**, 2015). The PDB files of the best predicted models were imported into PyMOL Molecular Graphics System (Version 1.7.4 Schrdinger, LLC.Pymol) for 3D modelling and analysis.

TABLE 3.1: **ALS-linked *CCNF* mutations in the cyclin box or PEST domains.**

Amino acid change	Nucleotide change	ALS	Domain
R392T	c.1175G>A	SALS	cyclin box
S509P	c.1525T>C	FALS	cyclin box
S621G	c.1861A>G	FALS	PEST
E624K	c.1870G>A	SALS	PEST

3.2 Predictive analysis of ALS-linked mutations in cyclin F

The predicted impact of ALS-linked mutations in the cyclin box or PEST domains of *CCNF* (Table 3.1) on cyclin F protein function were assessed using the following online programs: NetPhos 2.0 (**Blom et al.**, 1999), SIFT (**Kumar et al.**, 2009), PolyPhen 2 (**Adzhubei et al.**, 2010) and Mutation Taster (**Schwarz et al.**, 2014). NetPhos 2.0 is a program that utilises artificial neural networks to predict the phosphorylation sites of serine, threonine and tyrosine residues in a given sequence (**Blom et al.**, 1999). The FASTA sequence of wild-type and mutant cyclin F were uploaded to the online server, and comparisons made to predict changes in phosphorylation. FASTA sequences of the mutant cyclin F were uploaded to the SIFT and PolyPhen 2 online servers. The mutant cDNA sequences were uploaded into Mutation Taster.

3.3 Developing *CCNF* deletion clones for functional studies

3.3.1 Overview

Six different *CCNF* deletion constructs (Table 1.7) were designed and created for use in functional studies utilising neuronal cell lines. The deletion cloning process consisted of three parts: 1) the cloning of *CCNF* into an entry vector (Figure 3.2); 2) The introduction of the required deletions via *i*PCR mutagenesis (Figure 3.3); and 3) The sub-cloning of deletion cDNAs into the expression vector (Figure 3.4).

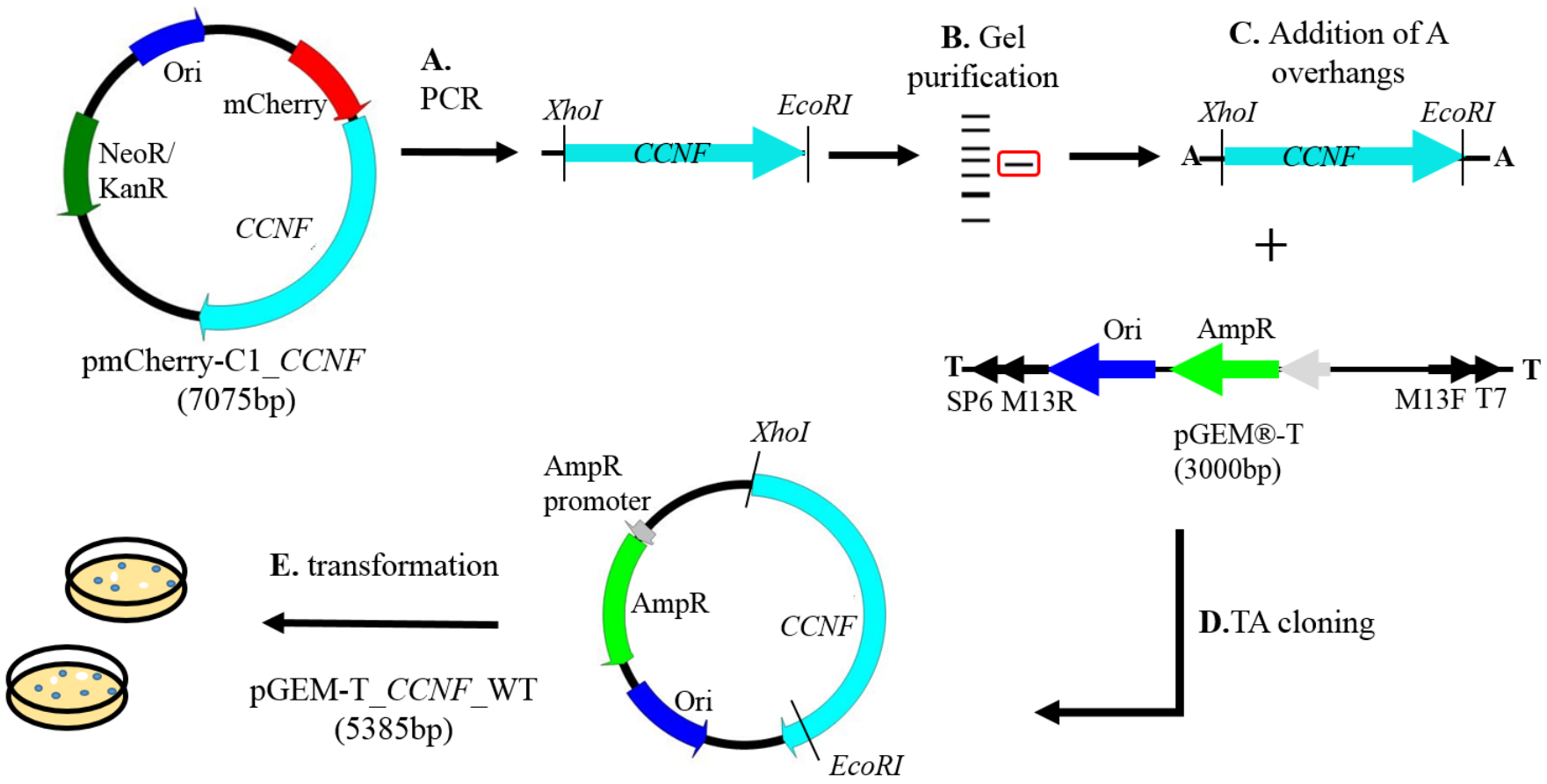


FIGURE 3.2: **Cloning *CCNF* into the pGEM-T entry vector. (A.)** The *CCNF* gene was PCR amplified from an existing vector, using primers to introduce an *XhoI* restriction site at the 5′ end and an *EcoRI* restriction site at the 3′ end (Section 3.3.3). **(B.)** The PCR products were gel-purified (Section 3.3.4). **(C.)** Adenine tails were added (Section 3.3.5). **(D.)** The *CCNF* PCR product was ligated into the pGEM-T entry vector via TA cloning (Section 3.3.6). **(E.)** Ligation products were transformed into *E.coli* (Section 3.3.7). The colonies were propagated, and purified (Section 3.3.9), and sequenced to confirm correct synthesis of required pGEM-T_*CCNF*_WT constructs(Section 3.3.10).

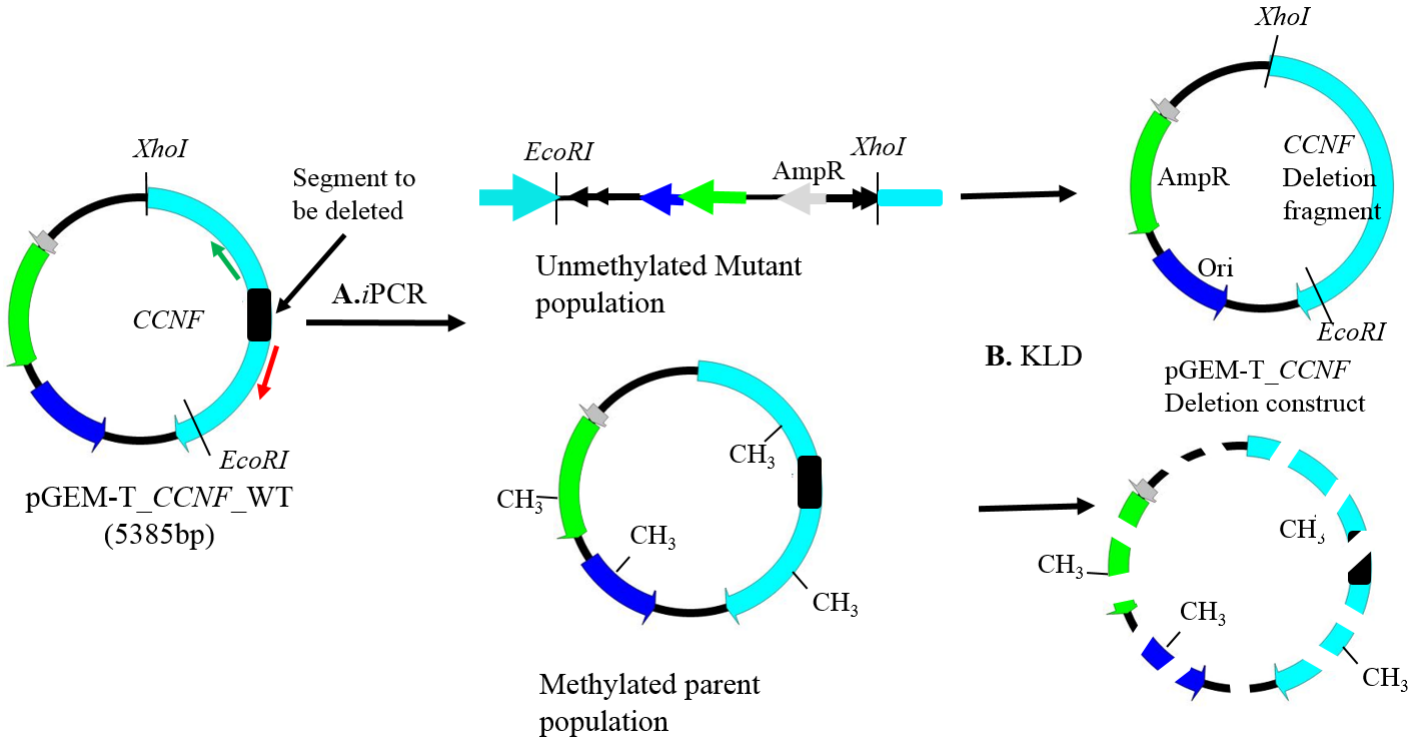


FIGURE 3.3: **Deletion cloning of *CCNF* via Q5 site directed mutagenesis. (A.)** Deletion clones were generated via Q5 Site directed mutagenesis utilising the generated pGEM-T_*CCNF*_WT construct as a template . Primers were designed using New England Biolabs (NEB) NEBaseChanger online design software website (http://nebasechanger.neb.com/) to bind to either side of the region to be deleted (Section 3.3). The mutant strands were synthesised by inverse PCR (*i*PCR)(Section 3.3.11). **(B.)** As amplified *i*PCR products are unmethylated, and parental DNA is methylated, the mutant strands were selected by *DpnI* digestion, which only digests methylated DNA. The mutant strands were simultaneously phosphorylated by a kinase and ligated together by DNA ligase(Section 3.3.11). The PCR products were then transformed, PCR amplified to check the correct construct had been obtained, and then propagated and purified before sequencing (Section 3.3.12)

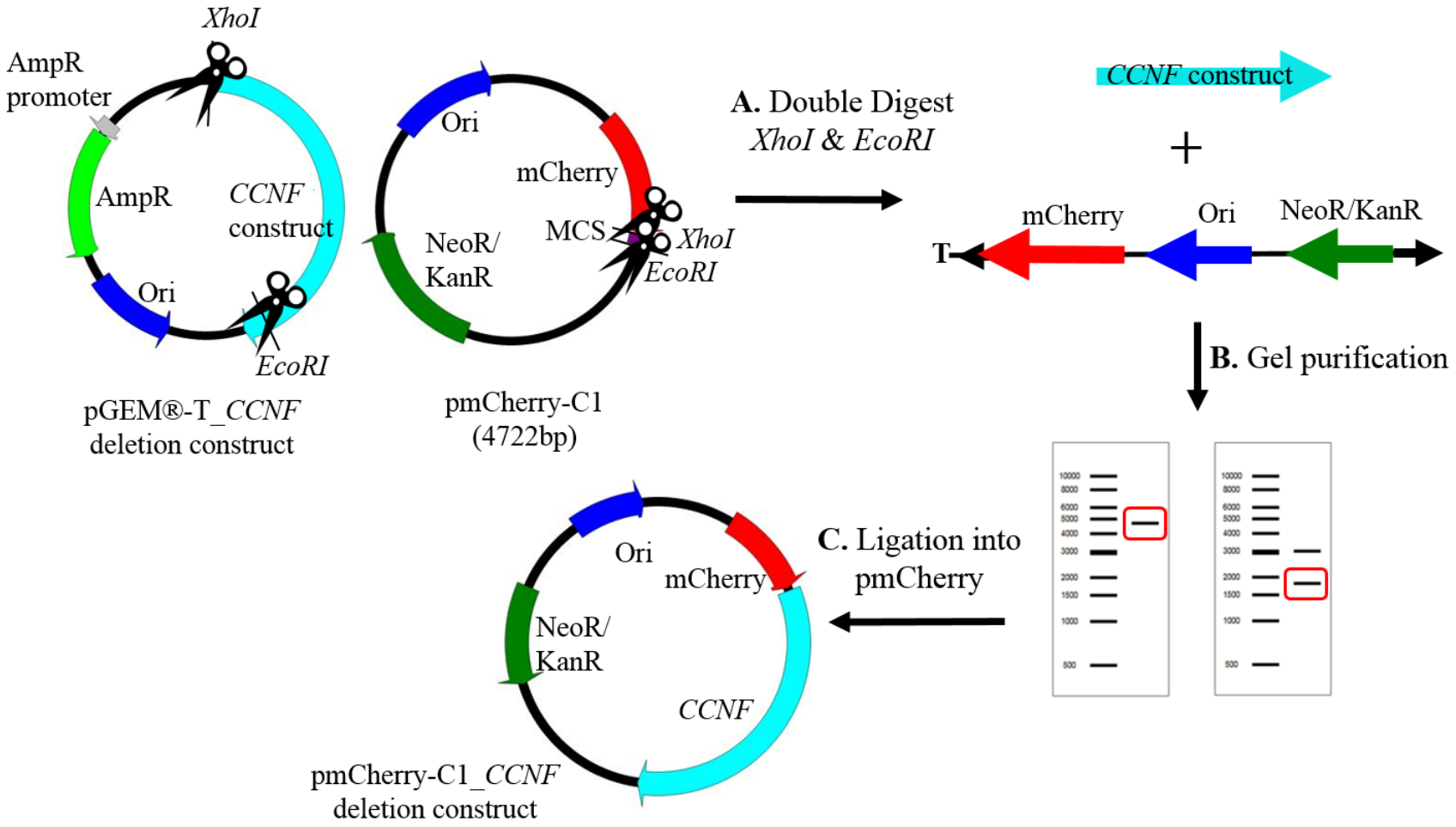


FIGURE 3.4: **Sub-cloning *CCNF* deletion fragments into the pmCherry-C1 vector.(A.)** The pGEM-T_*CCNF* deletion constructs and the pmCherry-C1 destination vector were digested with *XhoI* and *EcoRI* restriction enzymes (Section 3.3.13). **(B.)** The resulting restriction products were run on a gel, and the appropriate bands excised and gel purified (Section 3.3.4). **(C.)** The *CCNF* inserts and the open pmCherry-C1 vector were subsequently ligated together to form new pmCherry-C1_*CCNF* deletion constructs (Section 3.3.14).

3.3.2 Cloning primer design

Primers were designed as follows, and synthesised by Sigma Aldrich Pty Ltd (NSW, Australia).

Restriction enzyme-site cloning primers

Primers for restriction enzyme-site cloning had been previously designed. All primers were 23-25 nucleotides long, with ~11 nucleotides complementarity to the target sequence. Primers were designed to introduce an *XhoI* restriction site at the 5′ end of the PCR product and an *EcoRI* site at the 3′ end of the *CCNF* fragment for later use in restriction cloning. Primer sequences are shown in table 3.2.

Table 3.2: **Primers designed for the amplification of *CCNF* with the addition of restriction endonuclease sites.**

Oligo name	Sequence 5′ to 3′
CCNF_cDNA_XhoI_F	cagatctcgagctATGGGGAGCGG
CCNF_cDNA_EcoRI_R	ttgacgaattcTTACAGCCTCAC

Q5 Site-Directed Mutagenesis primers

For mutagenesis, primers were designed in the New England Biolabs (NEB) NEBaseChanger online design software website (http://nebasechanger.neb.com/). Primer sequences, and optimal annealing temperatures are shown in Table 3.3.

Table 3.3: **Primers designed to generate *CCNF* deletion mutants via Q5 site-directed mutagenesis.**

Oligo name	**Sequence 5' to 3'**	**T_A annealing temperature**
*CCNF*_ΔPEST_Q5_F	CGGATAAACCTATGCATAC	59°C
*CCNF*_ΔPEST_Q5_R	TTCCTGGAGGCTGTTCTC	59°C
*CCNF*_ΔCyclinN_Q5_F	CGAGTCCCCACTGTGGTG	64°C
*CCNF*_ΔCyclinN_Q5_R	CTGGGAAGCCTGAAATAGC	64°C
*CCNF*_ΔCyclinC_Q5_F	GAGGTGCTGAGCTACAGC	62°C
*CCNF*_ΔCyclinC_Q5_R	GACTCGAATCTTCCCTTCC	62°C
*CCNF*_ΔCyclinBox_Q5_F	GAGGTGCTGAGCTACAGC	64°C
*CCNF*_ΔCyclinBox_Q5_R	CTGGGAAGCCTGAAATAGC	64°C
*CCNF*_Δ262-766_Q5_F	CGGATAAACCTATGCATAC	59°C
*CCNF*_Δ262-766_Q5_R	CTGGGAAGCCTGAAATAG	59°C
*CCNF*_ΔNLS2_Q5_F	AGCCTCCAGGAAGACAGAG	66°C
*CCNF*_ΔNLS2_Q5_R	CCCCGAGGGAGAGCTGAG	66°C

3.3.3 PCR amplification of *CCNF* cDNA

CCNF was PCR amplified from an existing pmCherry-C1_*CCNF*_WT fusion construct template using Phusion High-Fidelity DNA polymerase (NEB) and primers from Table 3.2 (Figure 3.2A). (The pmCherry-C1_*CCNF*_WT fusion construct had previously been generated in our laboratory by PCR amplification of *CCNF* from a pmCherry_*CCNF* construct obtained from Addgene, followed by ligation into a pmCherry-C1 vector.) PCR conditions are shown in Table 3.2. The PCR products underwent agarose gel electrophoresis and gel purification (Section 3.3.4) (Figure 3.2B).

TABLE 3.4: **PCR conditions using Phusion High-Fidelity polymerase.**

Reagent	Volume(μl)
dH_2O	33.5
5X Phusion HF buffer	10
10mM dNTPs	1
20μM Forward primer	1.25
20μM Reverse primer	1.25
5ng/μl template DNA	1
DMSO	1.5
Phusion DNA polymerase	0.5
Total	50

Temperature	Time	Cycles
98°C	30 sec	1×
98°C	10 sec	
60°C	30 sec	25×
72°C	90 sec	
72°C	5 min	1×
15°C	hold	

3.3.4 Purification of PCR products via gel extraction

For gel purification of the PCR and restriction products, samples were run on a 0.8% (w/v) agarose gel in 1× TBE. The agarose gel was supplemented with 1× SYBR Safe DNA gel stain (Invitrogen, USA) for visualisation of PCR products under blue light. To maximise DNA yield, the full sample volume was loaded onto the gel, and electrophoresed at 100V for 45 to 75 minutes. The HyperLadder 100bp plus, HyperLadder 1 kb or HyperLadder 1 kb plus (Bioline, Australia) ladders were run to facilitate the determination of product size. All products were visualised and imaged on a Gel Doc EZ imager with Image Lab software (Bio-Rad, USA).

Bands of the correct size were excised from the gel using a scalpel, weighed, and then purified using the ISOLATE II PCR and Gel Kit (Bioline, Australia) as per the manufacturers instructions. The DNA was eluted into 15 μl of elution buffer supplied in the kit. Purified DNA was quantitated using the QIAxpert UV/VIS

spectrophotometer (QIAGEN, Netherlands).

3.3.5 Addition of adenine overhangs and TA cloning of *CCNF* into pGEM-T

To add adenine overhangs to the blunt-ended *CCNF* PCR fragment prior to ligation, a 10 μl reaction was set up as follows (Table 3.5), and incubated for 25 minutes at 72°C (Figure 3.3C).

Table 3.5: **Reagents for adding adenine overhangs in a 10 μl reaction.**

Reagents
~100ng gel-purified PCR product
0.2mM dATP nucleotides(NEB)
1× NEB Taq buffer
1U Thermopol Taq DNA polymerase (NEB)

3.3.6 TA Ligation of *CCNF* PCR product into pGEM-T entry vector

For ligation of the *CCNF* PCR product with A-overhangs into the pGEM-T vector, a 22 μl ligation reaction was set up as follows (Table 3.3.6) (Figure 3.3D).

Table 3.6: **pGEM-T ligation reaction.**

Reagents	Volumes
2× Rapid Ligation Buffer	10 μl
T4DNA ligase	1 μl
50ng Digested pGEM-T	x μl
100ng Insert DNA	y μl
Total	22 μl

A control ligation reaction was also conducted using the control insert provided in the pGEM-T vector kit. Ligation reactions were incubated overnight at 4°C. Products were then stored at -20°C.

3.3.7 Transformation protocol

Plasmid constructs were transformed into α-select Gold efficiency *E.coli* cells (Bioline, Australia) (Figure 3.3E). For each transformation reaction, 2 μl of ligation reaction was added to 25 μl of chilled *E.coli* cells. The cells were mixed gently with a pipette tip, then incubated on ice for 30 minutes. The cell suspensions were then heat shocked at 42°C for 45 seconds without shaking, and placed on ice for 2 minutes. The transformation reactions were diluted by the addition of 125 μl of SOC medium, and incubated at 37°C, with shaking at 200rpm for 60 minutes. The cells were then plated out under aseptic conditions at 10 μl and 100 μl volumes onto LB agar plates supplemented with 0.5 mM IPTG, 50 μg/mL XGal and 100 μg/ml ampicillin (Bioline, Australia). The plates were inverted and incubated overnight at 37°C.

3.3.8 Selection of and validation of recombinant constructs

Recombinant colonies of pGEM-T_*CCNF* constructs were identified by blue/white selection, with positive colonies appearing white due to disruption of the β-galactosidase gene situated in the multiple cloning site of the pGEM-T vector (Materials section 2.2.1; Figure 2.1). Eight colonies were picked from each plate for a colony PCR to ensure that wild-type *CCNF* had been inserted into the pGEM-T vector in the correct orientation. The plates were marked to allow colonies to be re-picked for purification and sequencing. The PCR was conducted using a forward primer situated within the vector backbone (T7 promoter_F) and a previously validated internal *CCNF* reverse primer (K97R_B_R) to PCR amplify a 416bp fragment (See Figure 4.16 & Table 3.7).

TABLE 3.7: **Primer sequences for validation of pGEM-T_*CCNF* recombinants.**

Primer Name	5'–3' Sequence	Product size
T7 Promoter_F	TAATACGACTCACTATAGGG	
K97R_B_R	GAAATCCCCTTTCAGCAGC	416bp

PCR was conducted under the following conditions (Table 3.8), and the PCR products underwent gel electrophoresis as previously described (Section 3.3.4) to verify the *CCNF* insert was present in the entry vector.

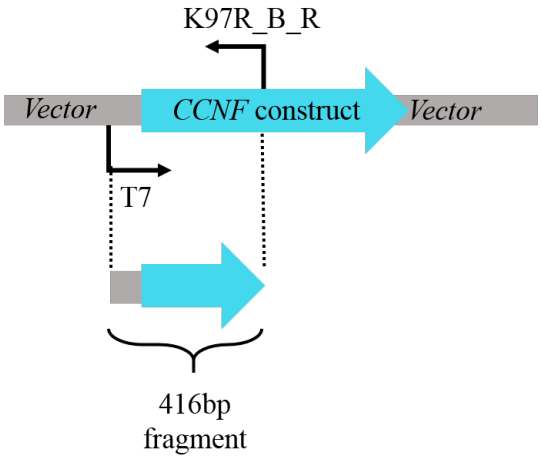


Figure 3.5: **Schematic of the check colony PCR primers showing the 416bp PCR product.** T7:T7 promoter_F

Table 3.8: **PCR conditions to check transformed *E.coli* colonies for target insert.**

Reagent	1× Volume (μl)
dH_2O	4.2
My*Taq* mix	5.0
20mM Forward Primer	0.4
20mM Reverse primer	0.4
Colony	+
Total	10.0

Temperature	Time	Cycles
95°C	3 min	1×
95°C	15 sec	
55°C	15 sec	30×
72°C	1 min	
15°C	Hold	

3.3.9 Purification of plasmid constructs

Prior to sequencing of the pGEM-T_*CCNF* constructs, all bacterial colonies confirmed to contain the desired plasmid construct were purified. Four colonies from each plate were re-picked and cultured in separate tubes of 5 mL LB supplemented with 5 μl ampicillin (100 mg/ml). The cultures were incubated overnight at 37°C at 200rpm. Glycerol stocks were made of each construct by adding 750 μl of culture to 250 μl of 100% glycerol (Sigma-Aldrich, USA), and stored at -80°C. The remaining plasmid DNA was purified using an ISOLATE II Plasmid Mini kit (Bioline, Australia) as per the manufacturers protocols. Purified plasmids were eluted in 25-50 μl of elution buffer supplied with the kit, and DNA concentrations were quantitated using a QIAxpert UV/VIS spectrophotometer (QIAGEN, Netherlands).

3.3.10 DNA sequencing

For each sample, 20 μg of plasmid DNA was Sanger sequenced using Big-Dye terminator sequencing on an ABI 3730XL sequencer (Macrogen, Korea). Primers used for sequencing were previously designed to amplify five overlapping segments of *CCNF* (Table 3.9) (Figure 3.6). Sequencing results were analysed using Sequencher v5.1 software (GeneCodes Corporation, USA). The sequences were trimmed of low quality bases, and subsequently compared to the reference sequence.

TABLE 3.9: **Primers for sequencing pGEM-T_*CCNF*_WT.**

Primer	Sequence
T7 promoter_F	TAATACGACTCACTATAGGG
*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
*CCNF*_NEWcDNApart_F	GGAGGTGAGAGCTTCCAGTG
*CCNF*_cDNApart_F	AGGACAAGCGCTATGGAGAA
*CCNF*_cDNAend_F	CCTGAACCCAGAACAGCATT

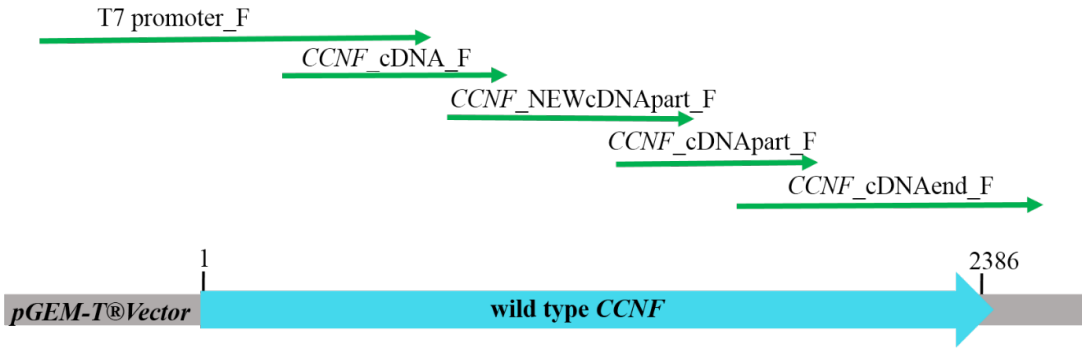


FIGURE 3.6: **Diagrammatic representation of the overlapping DNA fragments of the entire *CCNF* cDNA and flanking vector sequence generated by sequencing primers.**

3.3.11 Q5 site directed mutagenesis

Mutagenesis was performed on the pGEM-T_*CCNF*_WT construct using a Q5 site-directed mutagenesis kit (NEB, USA) as per manufacturers instructions (Figure

3.4A) to create six different *CCNF* deletion constructs (Table 1.7). Primer sequences and their optimal annealing temperatures for this process are shown in Table 3.3. PCR and thermocycling conditions for Q5 site-directed mutagenesis are shown in Table 3.10.

Table 3.10: **PCR conditions for Q5 site-directed Mutagenesis.**

Reagent	Volume(μl)
Q5 Hot Start High-Fidelity 2X Master Mix	12.5
10μM Forward Primer	1.25
10μM Reverse Primer	1.25
Template DNA (10ng/μl)	1
Nuclease free water	9
Total	25

Temperature	Time	No. of Cycles
98°C	30 sec	1 ×
98°C	10 sec	
50-72°C	30 sec	25 ×
72°C	2 min 45 sec	
72°C	2 min	1 ×

Q5 mutagenesis works on an inverse PCR (*i*PCR) principal, and involves the amplification of circular plasmid DNA (Figure 3.4A). Selection for deletion mutants and re-circularisation of the linear deletion vector is achieved in one step by a combined kinase, ligase and *DpnI* (KLD) treatment (NEB, USA) (Figure 3.4B) (Table 3.11). The kinase facilitates phosphorylation of the ends of the amplicon strands, enabling them to be ligated together with DNA ligase. The *DpnI* specifically binds to and digests methylated or hemi-methylated parental DNA, leaving the newly synthesised, un-methylated mutant strands untouched.

TABLE 3.11: **KLD reaction.**

Reagents	Volumes
PCR product	1μl
2$\times$ KLD reaction buffer	5μl
10$\times$ KLD enzyme	1μl
Nuclease free water	3μl
Total	10μl

The reaction was mixed and incubated for 5 minutes at room temperature.

The ligation products were transformed (Section 3.3.7), and cultured and purified (Section 3.3.9) as previously described.

3.3.12 Validation of pGEM-T_*CCNF* deletion constructs

Prior to sequencing, several colonies of each pGEM-T_*CCNF* deletion construct were PCR amplified using primers T7 promoter_F and SP6, and the PCR products underwent gel electrophoresis as previously described (Section 3.3.4) to ascertain if products of the correct size had been obtained. Predicted PCR product sizes are shown in Table 3.12.

TABLE 3.12: ***CCNF* deletion constructs in pGEM-T vector.**

Plasmid construct	Predicted size of PCR product
pGEM-T_*CCNF*_WT	2546
pGEM-T_*CCNF*_ΔPEST	1988
pGEM-T_*CCNF*_ΔCyclinN	2204
pGEM-T_*CCNF*_ΔCyclinC	2183
pGEM-T_*CCNF*_ΔCyclinBox	1835
pGEM-T_*CCNF*_Δ262-766	1121
pGEM-T_*CCNF*_ΔNLS2	2519

Sequencing was conducted as previously described (Section 3.3.10) using different combinations of sequencing primers to account for regions of *CCNF* that had been deleted (Table 3.13). As the CCNF_NEWcDNApart_F primer binds to a region that has been deleted in the ΔCyclinN construct, a new primer, CCNF_dCyclinN_F was designed using Serial Cloner software (version 2.6.1; SerialBascis). Although this

region is also deleted in ΔCyclinBox, the construct is small enough to require only three primers to sequence the full transcript.

Table 3.13: **Primers for sequencing pGEM-T_*CCNF* deletion constructs.**

Construct	Primer	Sequence
pGEM-T_CCNF_ΔPEST	T7 promoter_F	TAATACGACTCACTATAGGG
	*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
	SP6	ATTTAGGTGACACTATAG
pGEM-T_CCNF_ΔCyclinN	T7 promoter_F	TAATACGACTCACTATAGGG
	*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
	*CCNF*_dCyclinN_F	ATTACAAGGAGGTCCTGCTG
	SP6	ATTTAGGTGACACTATAG
pGEM-T_CCNF_ΔCyclinC	T7 promoter_F	TAATACGACTCACTATAGGG
	*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
	*CCNF*_NEWcDNApart_F	GGAGGTGAGAGCTTCCAGTG
	SP6	ATTTAGGTGACACTATAG
pGEM-T_CCNF_ΔCyclinBox	T7 promoter_F	TAATACGACTCACTATAGGG
	*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
	SP6	ATTTAGGTGACACTATAG
pGEM-T_CCNF_Δ292-766	T7 promoter_F	TAATACGACTCACTATAGGG
	SP6	ATTTAGGTGACACTATAG
pGEM-T_CCNF_ΔNLS2	T7 promoter_F	TAATACGACTCACTATAGGG
	*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
	*CCNF*_NEWcDNApart_F	GGAGGTGAGAGCTTCCAGTG
	SP6	ATTTAGGTGACACTATAG

T7 promoter_F & SP6 are universal primers
*CCNF*_cDNA_F & *CCNF*_NEWcDNApart_F were previously designed in our laboratory

3.3.13 Cloning facilitated by restriction enzyme digestion

The restriction enzymes, *EcoRI* and *XhoI* were used to digest the *CCNF* inserts out of the pGEM-T vector, as well as to digest the pmCherry-C1 destination vector (Figure 2.2). The pGEM-T_*CCNF* constructs and pmCherry-C1 vector were digested in 20 μl-30 μl reactions as shown in Table 3.14 (Figure 3.4A).

TABLE 3.14: **Double digest reactions.**

Reagents	Volumes
2 μg plasmid DNA	x μl
10$\times$ NEB 3.1 buffer	3 μl
XhoI (20,000 Units/mL)	0.5 μl
EcoRI (20,000 Units/mL)	1 μl
Nuclease free water	y μl
Total	up to 30 μl maximum

The reactions were incubated at 37°C for 2.5-16 hours and inactivated at 65°C for 20 minutes. After inactivation, 3 μl of ThermoSensitive Alkaline Phosphatase (TSAP; Promega, USA) was added to the digested pmCherry-C1 vector to remove 5′ phosphate groups, thereby preventing re-circularisation of the plasmid. The reaction was incubated at 37°C for 15 minutes, followed by heat inactivation at 80°C for 15 minutes.

The *CCNF* inserts and pmCherry-C1 vector were gel purified prior to ligation (See Section 3.3.4)(Figure 3.4B)

3.3.14 Ligation of restriction products into the pmCherry-C1 vector, and transformation in *E-coli*.

For ligation of *CCNF* deletion fragments into pmCherry-C1 vector (Figure 3.4C), a 10 μl ligation was set up as follows (Table 3.15), with the appropriate amount of insert DNA for a 1:1 molar ratio as determined using the following formula:

$$\frac{\text{kb of insert}}{\text{kb of vector}} \times \text{ng of vector}$$

TABLE 3.15: **Reagents for ligation of insert into pmCherry-C1 vector.**

Reagents	Volumes
2$\times$ Rapid Ligation Buffer	5 μl
T4DNA ligase	1 μl
~100ng digested pmCherry-C1 vector	x μl
Insert DNA (1:1 ratio)	y μl
Total	10 μl

Ligation reactions were incubated overnight at 4°C. Products were then stored at -20°C. Transformations were conducted as described previously (Section 3.3.7); however the cells were plated out onto LB agar plates supplemented with 2.5 μl of kanamycin.

3.3.15 Selection of and validation of recombinant constructs

To verify that the *CCNF* deletion constructs had successfully ligated into the pmCherry-C1 vector, several colonies from each construct were PCR amplified as previously described (Section 3.3.8) using a forward primer located in the pmCherry-C1 vector backbone and a reverse primer located within the *CCNF* gene (Table 3.16). If ligation is successful, these primers generate a 425bp fragment, which can be visualised on an agarose gel (Section 3.3.4). The recombinant colonies were re-picked, cultured and purified as previously described (Section 3.3.9).

Table 3.16: **Primer sequences for validation of pmCherry-C1_*hCCNF* recombinants.**

Primer Name	5'–3' Sequence	Product size
K97R_A_F	CATCAAGTTGGACATCACCTCC	
K97R_B_R	GAAATCCCCTTTCAGCAGC	425bp

Two purified plasmids were Sanger sequenced for each construct as previously described (Section 3.3.10) using the following sequencing primers (Table 3.17). Primers were designed previously in our laboratory to ensure the generation of overlapping segments that covered the mCherry sequence, the *CCNF* gene insert and a part of the vector backbone at the 3′ end of the insert. However, one primer, *CCNF*_cDNAend_F binds to regions that have either been deleted (ΔPEST), or are close to the deletion site (ΔCyclinC) in two constructs. Therefore, a new primer (*CCNF*_dCyclinC_F) was designed in Serial Cloner software (version 2.6.1; SerialBasics) to cover the 3′ region.

TABLE 3.17: **Primers for sequencing pmCherry-C1_*CCNF* deletion constructs.**

Construct	Primer	Sequence
pmCherry-C1_*CCNF*_ΔPEST	*CCNF*_cDNA_start_R	GCTTTGGCTAAAGACAGCTG
	*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
	*CCNF*_NEWcDNApart_F	GGAGGTGAGAGCTTCCAGTG
	*CCNF*_dCyclinC_F	TGAAGGACTTCACAAGCCTG
pmCherry-C1_*CCNF*_ΔCyclinN	*CCNF*_cDNA_start_R	GCTTTGGCTAAAGACAGCTG
	*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
	*CCNF*_NEWcDNApart_F	GGAGGTGAGAGCTTCCAGTG
	*CCNF*_dCyclinN_F	ATTACAAGGAGGTCCTGCTG
	*CCNF*_cDNAend_F	CCTGAACCCAGAACAGCATT
pmCherry-C1_*CCNF*_ΔCyclinC	*CCNF*_cDNA_start_R	GCTTTGGCTAAAGACAGCTG
	*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
	*CCNF*_NEWcDNApart_F	GGAGGTGAGAGCTTCCAGTG
	*CCNF*_dCyclinC_F	TGAAGGACTTCACAAGCCTG
	*CCNF*_cDNAend_F	CCTGAACCCAGAACAGCATT
pmCherry-C1_*CCNF*_ΔCyclinBox	*CCNF*_cDNA_start_R	GCTTTGGCTAAAGACAGCTG
	*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
	*CCNF*_cDNAend_F	CCTGAACCCAGAACAGCATT
pmCherry-C1_*CCNF*_Δ292-766	*CCNF*_cDNA_start_R	GCTTTGGCTAAAGACAGCTG
	*CCNF*_cDNA_F	CAACCACGCCAGTGTGT
	*CCNF*_NEWcDNApart_F	GGAGGTGAGAGCTTCCAGTG
	*CCNF*_cDNAend_F	CCTGAACCCAGAACAGCATT

All primers except for *CCNF*_dCyclinN_F & *CCNF*_dCyclinC_F were previously designed in our laboratory

3.3.16 DNA purification for *in vitro* functional studies

For *in vitro* experiments, large amounts of endotoxin-free DNA are required. Therefore, the pmCherry-C1*CCNF* plasmids are required to be purified using an endotoxin-free purification method. Glycerol stocks of the pmCherry-C1_*CCNF* plasmids (Table 3.18) were grown up in 40 mL of LB supplemented with 20 μl of Kanamycin (100mg/ml) overnight at 37°C with shaking at 200rpm. The plasmids were then purified using Hi-Pure Plasmid Filter Midi-Prep kit (Invitrogen, USA) as per manufacturers instructions. The purified DNA pellet was resuspended in 100 μl of TE buffer, and quantitated (Section 3.3.4).

TABLE 3.18: **pmCherry-C1_*CCNF* deletion constructs used in transfections.**

Construct	Description
pmCherry-C1_*CCNF*_WT*	Wild-type CCNF
pmCherry-C1_*CCNF*_ΔPEST	PEST region deleted
pmCherry-C1_*CCNF*_ΔCyclinN	Cyclin N deleted
pmCherry-C1_*CCNF*_ΔCyclinC	Cyclin C deleted
pmCherry-C1_*CCNF*_ΔCyclinBox	Whole Cyclin box deleted
pmCherry-C1_*CCNF*_Δ262-766	Cyclin box to PEST deleted

* Previously made in our laboratory

3.4 *In vitro* studies in neuronal cell lines

3.4.1 Overview

Six generated pmCherry-C1_*CCNF* constructs (Table 3.18) were used for *in vitro* studies in two neuronal cell lines: the murine NSC-34 (neuroblastoma/motor neuron-enriched primary spinal cord hybrid) cell line (**Cashman et al.**, 1992) and the human neuroblastoma, SH-SY5Y cell lines obtained from the ATCC repository (ATCC product nos. CRL-131 and CRL-2266). The constructs were transfected into the cell lines, and cyclin F expression and aggregation formation was visualised using fluorescent/confocal microscopies. The toxicity of cyclin F-transfected NSC-34 cells were examined using a cell impermeable SYTOX-blue dead cell stain (Molecular Probes, Thermo-Fisher, USA) followed by flow cytometry analysis.

3.4.2 NSC34 murine and SH-SY5Y cell culture

NSC-34 and SH-SY5Y cells were maintained in cell specific growth media (see Materials 2.1.2) in 75 cm^2 sterile Falcon tissue culture flasks (Corning, USA). The cells were incubated at 37°C with 5% CO_2 in a HERACELL 150i CO_2 humidified incubator (Thermo Fisher, USA).

3.4.3 Cell passaging

When the cells reached ~90% confluence (every 3-4 days), they were passaged into new flasks. The media was aspirated, and the flasks rinsed with 5 mL of 1× PBS. For the SH-SY5Y cells, the aspirated media and PBS washes were retained and added to the resuspended cells prior to centrifugation as SH-SY5Y cell cultures contain both adherent and floating cells (**Kovalevich and Langford**, 2013). The cells were

dislodged from the flask by the addition of 1 mL of pre-warmed 0.25% Trypsin-EDTA (Sigma-Aldrich, USA) and subsequent incubation at 37C for <5 minutes. The flask was tapped gently to dislodge cells. The cells were resuspended in 5 mL of growth media to neutralise the trypsin. The cell suspension was centrifuged for 5 minutes at room temperature at 1,105 rcf. The supernatant was removed, and the cell pellet was resuspended in 5 mL of growth media by repeated pipetting. The cells were split in a 1:10 ratio into a new flask containing 10 mL of pre-warmed growth media and incubated at 37°C and 5% CO_2 in a humidified incubator.

3.4.4 Cell seeding for transfection

For visualising gene expression and aggregation formation, NSC-34 and SH-SY5Y cell lines were grown on (#1.5) 12mm glass coverslips (Menzel-Gläser, Germany) in Falcon 24 well tissue culture plates at a total density of 1×10^5 cells per well in 1 mL of the cell-appropriate growth medium (see Materials 2.1.2). Cells were cultured for 24 hours in a humidified incubator at 37°C at 5% CO_2 before transfection. For the toxicity study, NSC-34 cells were grown in sterile Falcon 6-well tissue culture plates, at a density of 5×10^5 cells/ well in 5mL of growth medium, and incubated for 24 hours at 37°C with 2% CO_2.

3.4.5 Transfections

Prior to transfections, each pmCherry-C1_*CCNF* construct (Table 1.7) was diluted to a concentration of 500 ng/μl in TE buffer. For each well in 24 well plates, 1 μg of pmCherry-C1_*CCNF* construct DNA was diluted in 48 μl Gibco $1\times$ Opti-MEM (Life Technologies, USA); and 4 μl of Lipofectamine 2000 (Invitrogen, USA) was added to 46 μl Opti-MEM. For each well in 6 well plates, 4 μg of DNA was diluted in 242 μl of Opti-Mem; and 16 μl of Lipofectamine 2000 was diluted in 234 μl of Opti-MEM. After 5 minutes incubation, the diluted DNA was combined with the diluted Lipofectamine for each reaction and incubated at room temperature for 20 minutes. The medium in plates was aspirated, the plates were washed once with PBS, and Opti-MEM was added (500 μl of Opti-MEM for 24 well plates, or 2 mL of Opti-MEM for 6 well plates). One hundred microlitres of the DNA/Lipofectamine complexes were added to each well of the 24 well plates, and 500 μl of DNA/Lipofectamine complexes were added to each well of the 6 well plates. The cells and media were mixed by gentle rocking.

The cells were incubated at 37°C with 5% CO_2 for 4 hours in a humidified incubator. After this time, growth media with 20% FBS was added to each well (600 μl for 24 well plates and 3 mL for 6 well plates), and the cells were further incubated at 37°C with 5% CO_2 overnight. The media was removed from plates after 24 hours, and replaced with the appropriate standard growth media.

3.4.6 Fixation

After 48 hours, cells grown on coverslips were fixed with 500 μl of 4% paraformaldehyde (PFA) (Merck) in PBS, and permeabilised in 0.2% Triton X-100 in PBS. The cells were incubated for 15 minutes at room temperature with rocking, then washed twice with PBS. The glass coverslips growing NSC-34 cells were immediately removed, air-dried and then mounted on glass slides with ProLong Gold Anti-fade with DAPI mounting medium (Molecular Probes, Thermo-Fischer, USA).

3.4.7 Confocal microscopy

Cells were visualised on a Leica DM6000 upright laser scanning confocal microscope using the Leica Application Suite Advanced Fluorescence software (Leica, Germany). Images were taken using a 63$\times$ (1.4 numerical aperture) oil immersion objective using Leica immersion oil type F (Leica, Germany). Images were merged and cells were counted using the cell counting tool in ImageJ (V1.49) software (**Schneider et al.**, 2012).

3.4.8 Cell counting

For each construct, at least eight images were randomly acquired from two experimental repeats. To quantify the localisation of each construct, cells with cyclin F expression in the nucleus, cytoplasm, or both nucleus and cytoplasm were counted from each image. Cells with either nuclear or cytoplasmic aggregates were also counted from each image.

3.4.9 SYTOX blue dead cell assay using flow cytometry

For the cell death assay, the NSC-34 cells were cultured in 6 well plates, and transfected with pmCherry-C1_*CCNF* constructs in triplicate. The Lipofectamine 2000 with no DNA was used as a vehicle control. After 48 hours, the media from the transfection plates was removed and placed in Falcon tubes. The plates were washed with PBS and the solution transferred to the relevant tubes. Cells were dislodged by the addition of 250 μl trypsin and incubated for 1-2 minutes at 37°C. Seven hundred and fifty microlitres of growth medium was added to each well, and then the media/trypsin solution was transferred to the relevant tubes. The tubes were centrifuged at 1,105 rcf for 5 minutes. The supernatant was discarded and cells were resuspended in 1 mL of PBS. The resuspended cells were transferred to flow cytometry tubes, and 1 μl of Sytox Blue Dead cell stain (Molecular Probes) stain was added to each tube and incubated for 5-30 minutes in the dark. SYTOX Blue (excitation and emission maxima: 444 nm and 480 nm, respectively) is a high-affinity nucleic acid stain used in conjunction with violet laser flow cytometry to quantify the proportion of dead cells in a sample. It is a cell impermeable dye and can only penetrate the compromised membranes of dead cells. This allows discrimination between live and dead cells. To eliminate the effect of difference in transfection efficiency, cell death was quantified in transfected cells that expressed mCherry fluorescence (excitation and emission maxima: 587 nm and 610 nm, respectively).The tubes were run on a FACSCanto II flow cytometer (BD Biosciences, country), and 10,000 mCherry positive events were captured. Flow cytometry data were analysed using FlowJo software (FlowJo, LLC, USA) (version 10.0.8).

To select a flow cytometry sub-population for further analysis and eliminate unwanted particles, a gating process was applied to the negative control (Lipofectamine 2000, (Invitrogen, USA) with no DNA). The first step in the gating process involved selecting a population based on their forward scattered light (FSC) and scattered light (SSC) measurements. As shown in Figure 3.7, FSC is proportional to the size of the cell while SSC is proportional to the internal cellular complexity, including size of nucleus, cell membrane and granularity. Debris and other unwanted particles, such as clumps and doublets, and can be roughly excluded by drawing a gate around the population of interest (denoted as P1 in Figure 3.8A). The P1 population is further gated for mCherry expression (mCherry pos) (Figure 3.8B). A gate P2 is drawn to exclude most of the negative control population (Figure 3.8B). This allows examination of transfected cells (Figure 3.8B). The mCherry positive population is then further gated by a gate P3 to define SYTOX Blue positive population (Figure

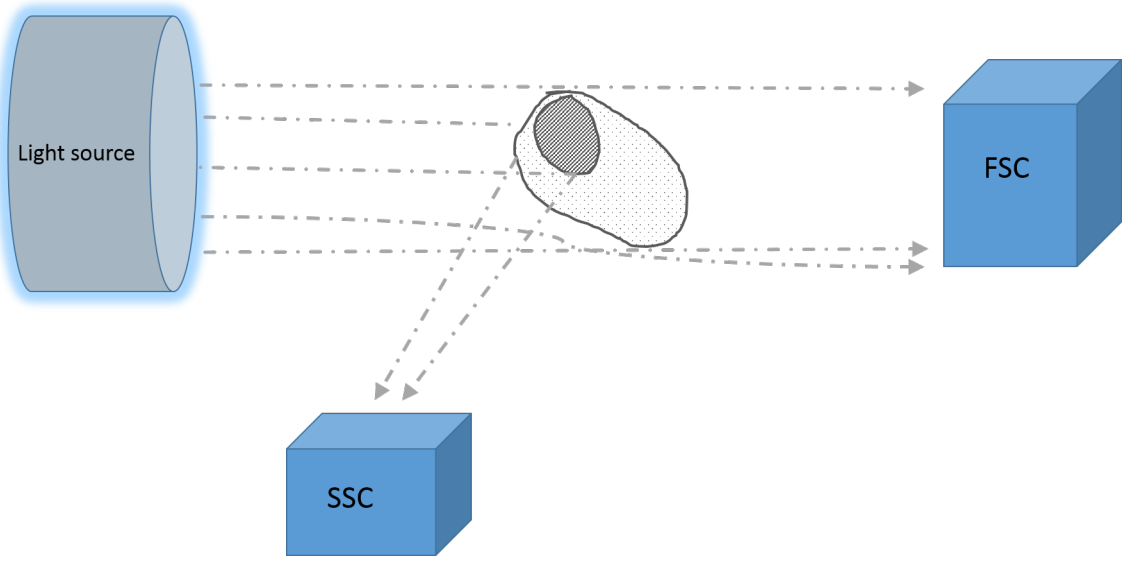


FIGURE 3.7: **Forward scatter light (FSC) versus side scatter light (SSC) as detected by a flow cytometer.** The forward scattered light is projected forward onto the forward scatter detector, and is proportional to the size of the cell. The side scattered light is deflected onto the side scatter detector when it hits internal organelles or the cell membrane. It is therefore directly proportional to the intracellular complexity of the cell (e.g. size of organelles, level of granularity, thickness of cell membrane). Cells with similar FSC and SSC measurements will be selected as the initial gating. FSC: Forward scattered channel. SSC: Side scattered channel.

3.8C). The same gating was applied to all other samples.

3.4.10 Statistics

Statistical analyses on the cell counting and flow cytometry data were performed using the GraphPad Prism (version 6). One-Way Anova tests were performed to assess if there were any significant differences between the different constructs. Significance of the differences between wild type and truncated cyclin F constructs were determined by a student *t*- test for sub-cellular localisation, mCherry expression and cell toxicity. All values in the figures are shown as means ± standard error of the mean (SEM). (* $p<0.05$; ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$.)

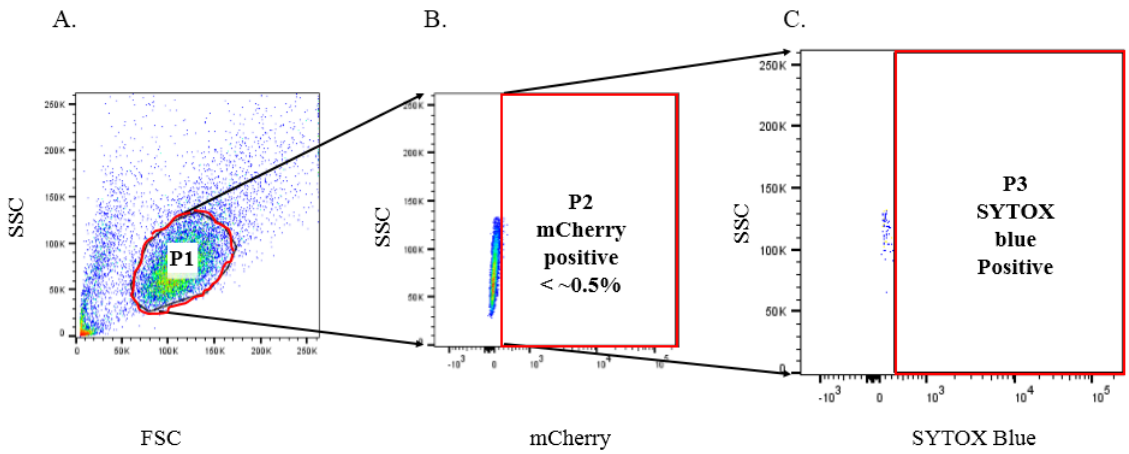


FIGURE 3.8: **Selecting the population of interest in flow cytometry data.(A.)** The parent population (P1) is first set up around a population with similar size and internal complexity based on their FSC and SSC value. **(B.)** P1 is further gated to include cell populations that are mCherry positive (P2). This value should be $< \sim 0.5\%$ of the negative control sample because it has not been transfected with the mCherry-C1_*CCNF* constructs. **(C.)** The mCherry positive population is further gated for SYTOX blue fluorescence, allowing the exclusion of the majority of the population.

4 Results

4.1 Overview

This project aimed to characterise the cyclin F protein to investigate the potential effects of ALS- and FTD-linked mutations on protein function. The first aim of the project was to identify the 3D structural features of the cyclin F using bioinformatic software to gain insight into function of both the wild-type and mutant protein. The second aim of the project was to examine the function of the different domains of the cyclin F by conducting domain-specific deletion cloning and expression of the clones in living neuronal cells.

4.2 Homology modelling and 3D visualisation of cyclin F in SWISS-MODEL

Homology modelling of the full length cyclin F was conducted in the online 3D modelling program SWISS-MODEL (**Arnold et al.**, 2006; **Bordoli et al.**, 2009). Over 200 matches were found for cyclin F, most of which had similarity only to the cyclin box domain of cyclin F. The top ranked homologues were cyclin B1

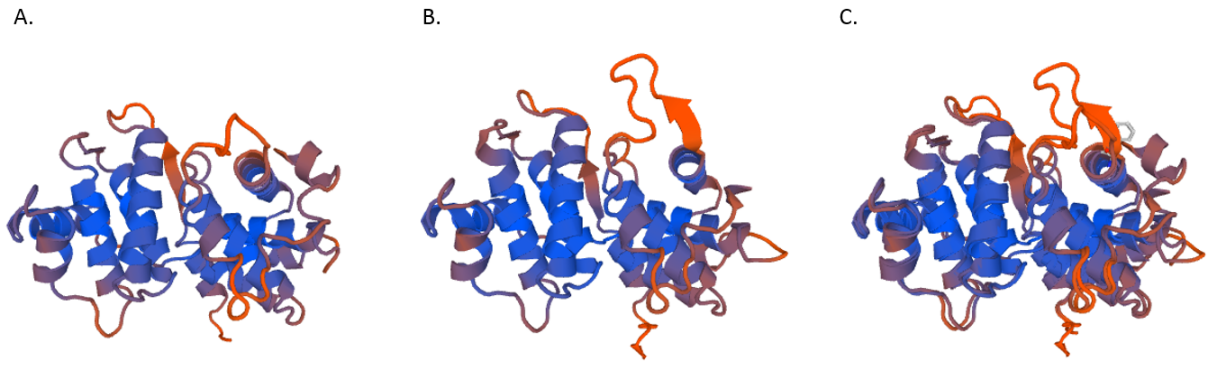


FIGURE 4.1: **3D models of two cyclin F homologues, cyclin B1 and cyclin A2, generated using homology modelling in Swiss Model.** Blue regions indicate conserved sequences with cyclin F, Red regions indicate divergent sequences with cyclin F. **(A.)** A model of G2/mitotic-specific cyclin-B1 (PDB # 4y72.1B). QMEAN4 = -4.41. **(B.)** A model of Cyclin A2 in complex with phosphorylated CDK1 (PDB # 3ddp.1). QMEAN4 = -3.96. **(C.)** A composite of the two homologues (cyclin B1 and cyclin A2) showing their similarity to each other.

(PDB:4y72.1B) (Figure 4.1A) with a QMEAN4 value of -4.41, followed by cyclin A2 (PDB:3ddp.1)(Figure 4.1B), with a QMEAN4 value of -3.96. Both of these values are indicated as being of low confidence by the inbuilt SWISS-MODEL program indicator. As shown in Figure 4.1, cyclin A2 and cyclin B1 are composed of two clusters of alpha helices, one containing five alpha helices, and one containing seven alpha helices. These alpha helices have the highest amino acid conservation with cyclin F while the loop regions have the lowest conservation. While several template homologues were found for the F-box domain, they were not modelled due to the lack of ALS-linked mutations found in this region.

4.3 Threading and 3D visualisation of cyclin F in Aquaria

Threading alignments were conducted for the full length Cyclin F in Aquaria (**O'Donoghue et al.**, 2015). The most confident templates were cyclin A2 and cyclin B1. Cyclin F aligned to cyclin A2 (PDB 4cfu-B) with 29% sequence identity, and an HHblits E-value of 2.7 $10^{-}44$ (Figure 4.2. Similarly to the models built in SWISS-MODEL, high levels of conservation were found in the alpha helices, especially

the hydrophobic patch region, while the areas with the lowest conservation were the loop regions. As shown in Figure 4.2, the two ALS-linked mutations, p.R392T and p.S509P, (coloured orange) lie in close proximity to the two alpha helices that make up the hydrophobic patch (coloured in pink).

4.4 3D visualisation of full length cyclin F using I-TASSER

To investigate whether a model could be obtained for the full structure of the cyclin F protein, the FASTA protein sequence (NP_41002) was uploaded to I-TASSER (**Roy et al.**, 2010) for predictive modelling. The best predicted 3D structure model of cyclin F (shown in Figure 4.3) gave a low C-score (-2.7) and TM-score (0.40 ± 0.14), indicating a lack of confidence in the structure. When individual domains were submitted to I-TASSER, confident models could be obtained for the cyclin box and F-Box domains, but not for the PEST domain. The top-ranked model obtained for the F-box domain had an acceptable C-score of 0.25 and estimated TM=score of 0.75 ± 0.11 (Figure 4.4). The top threading templates included other components of SCF ligases, including Skp1, which is known to be structurally similar to the F-box domain of cyclin F (**Kipreos and Pagano**, 2000). The top-ranked model for the cyclin box domain had a C-score of 1.47 and a TM-score of 0.92± 0.06, with the top threading templates aligned to cyclin A2, cyclin B1 and cyclin A3 as found with the other programs (Figure 4.5). Like the other models of cyclin A2 and cyclin B1 (4.1 & 4.2), the I-TASSER model of the cyclin box also contained two clusters of alpha helices: a cluster of five alpha helices encompassing the hydrophobic patch, and another cluster with seven alpha helices. The model obtained for PEST had low confidence, with a C-score = -3.49 and a TM-score = 0.29 ±0.09. The F-box and cyclin box domains were individually modelled in PyMOL Molecular Graphics System (Version 1.7.4 Schrdinger, LLC.Pymol) (Figures 4.3, 4.4 & 4.5).

Cyclin B1 (Uniprot#P14635), cyclin A2 (uniprot#P20248) and cyclin F (Uniprot#P41002) were aligned in Clustal Omega (**Sievers et al.**, 2011) to find areas of significant conservation. The α-1 and α-3 helices containing the two sections of the hydrophobic patch had the highest level of conservation (red boxes in Figure

4.6). Therefore the I-TASSER cyclin F cyclin box model was visualised in Pymol to show close ups of the region (Figure 4.7A). The mutant cyclin box models were visualised in PyMOL to show atomic contacts between the hydrophobic patch and the two residues substituted by ALS-linked mutations,(p.R392T and p.S509P). The models were compared to the same region of the wild type cyclin box (Figure 4.7B-E). This model confirms the prediction that the cyclin box mutations lie close to the hydrophobic patch in 3D space. Of particular interest, the substitution of a serine for a proline at position 509 is predicted to change the conformation of a threonine residue in the adjacent MYRIL sequence of the hydrophobic patch (Figure 4.7D & E)

4.5 Mutational analysis of ALS-linked mutations in cyclin F

The ALS- and FTD-linked *CCNF* mutations were analysed in four programs that predict the pathogenicity of gene variants (Table 4.1). Cyclin F is predicted to have 53 potential phosphorylation sites (39 serine, 8 threonine & 6 tyrosine sites) as determined by NetPhos 2.0. Almost half of these sites lie in the PEST sequence (18 serine, 2 tyrosine and 3 threonine potential phosphorylation sites (Figure 4.8). This high level of phosphorylation in the PEST sequence is similar to other PEST sequences (**Meyer et al.**, 2011). Two of the missense mutations (p.S509P, and p.S621G) result in a predicted loss of phosphorylation site, and are considered pathogenic by at least 2 of the programs (Table 4.1).

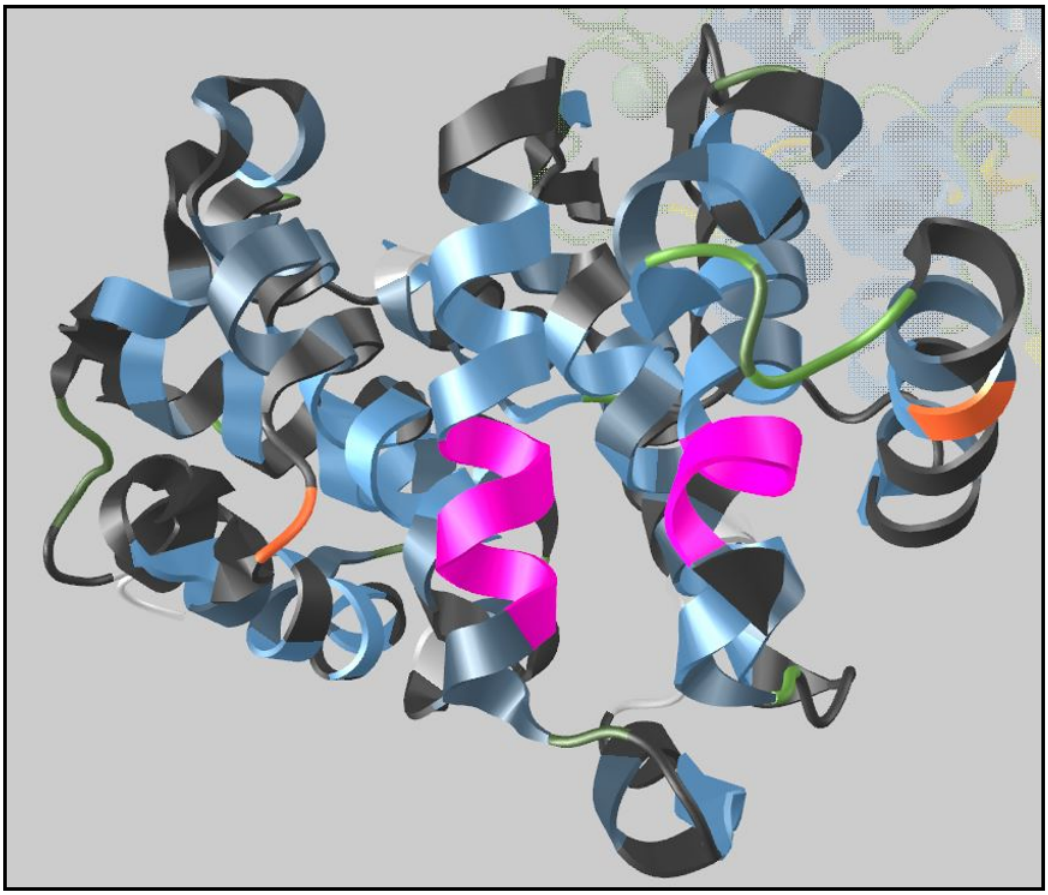


Figure 4.2: **3D model of the related cyclin F structure cyclin A2 generated by Aquaria.**
. The figure depicts the related cyclin F construct, cyclin A2. Two ALS-linked mutations, identified by our laboratory (unpublished data) are coloured dark pink. Bright colours indicate identical residues between the cyclin A2 template and cyclin F. Darker colours indicate conserved substitutions, and very dark colours indicate non-conserved substitutions. The two alpha helices of the hydrophobic patch are shown coloured bright pink.

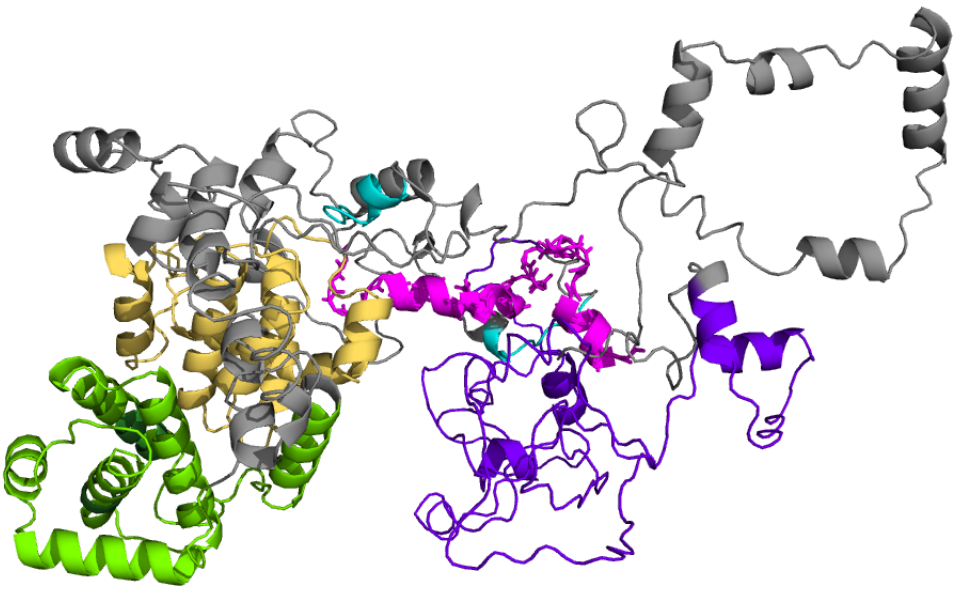


FIGURE 4.3: **A 3D model of cyclin F as determined using I-TASSER and visualised in pymol.** C-score = -2.7; estimated TM-score = 0.40 ± 0.14). The domains are coloured as follows: aqua = NLS1 and NLS2; pink= F-Box domain; green= cyclin N domain; yellow = cyclin C domain; purple = PEST domain; grey=intervening segments.

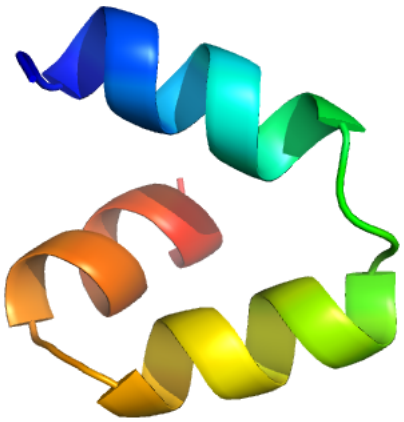


FIGURE 4.4: **The 3-helix cluster structure of the F-box domain of cyclin F, generated by I-TASSER and visualised in Pymol.** C-score=0.25, TM=0.75±0.11. Structures are coloured from blue at the C-terminal to red at the N-terminal.

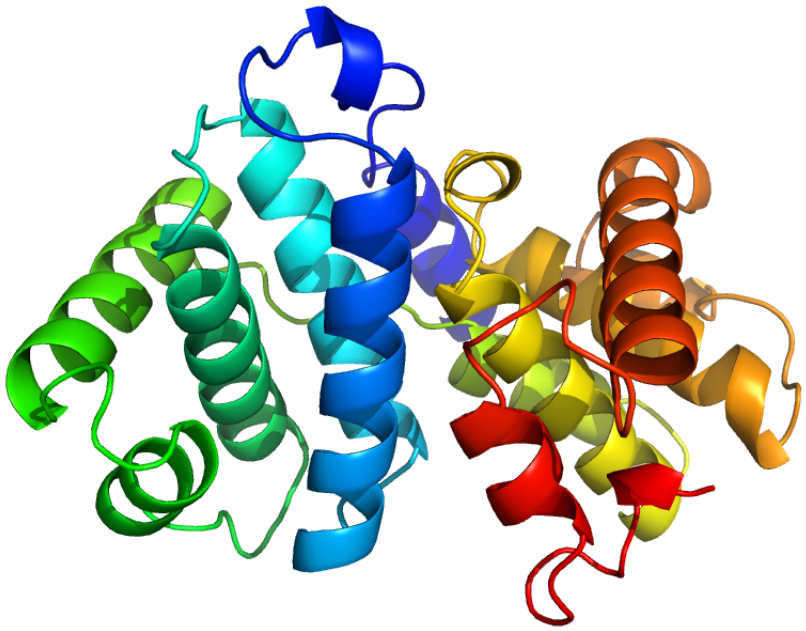


Figure 4.5: **The helical turns of the cyclin box domain, generated by I-TASSER and visualised using Pymol.** C-score = 1.47 and a TM-score = 0.92± 0.06. Structures are coloured from blue at the C-terminal to red at the N-terminal.

```
sp|P41002|CCNF_HUMAN      A----NANQ----LGLEVRASSEIVCQLFQAS-----QAVSKQQVFSVQKGLNDTMRYIL 313
sp|P20248|CCNA2_HUMAN     --------ILEDEKPVSVNEVPDYHEDIHTYLREMEVKCKPKVGYMKKQPDITNSMRAIL 214
sp|P14635|CCNB1_HUMAN     SDVILAVNDVDAEDGADPNLCSEYVKDIYAYLRQLEEEQAVRPKYLL-GREVTGNMRAIL 205
                                    . .   :   ::.       :   :   :    :. .** **

sp|P41002|CCNF_HUMAN      IDWLVEVATMKDFTSLCLHLTVECVDRYLRRRLVPRYRLQLLGIACMVICTRFISKEILT 373
sp|P20248|CCNA2_HUMAN     VDWLVEVGEEYKLQNETLHLAVNYIDRFLSSMSVLRGKLQLVGTAAMLLASKFEEIYPPE 274
sp|P14635|CCNB1_HUMAN     IDWLVQVQMKFRLLQETMYMTVSIIDRFMQNNCVPKKMLQLVGVTAMFIASKYEEMYPPE 265
                          :*****:*    : .  ::::*. :**::    * :  ***:* :.*.:::: .
```

Figure 4.6: **A snapshot of the Clustal Omega output comparing alignment of the 3 cyclin proteins: cyclin F (encoded by *CCNF*); cyclin A2 (encoded by *CCNA2*); and cyclin B1 (encoded by *CCNB1*).** The red boxes enclose the α-1 and α-3 helices. The two hydrophobic patches are underlined in black: MRYIL in cyclin F in the α-1 helix and LQLL in cyclin F in the α-3 helix. An asterisk (*) denotes positions which have conserved residues. A (:) (colon) denotes conservation between groups with strongly similar properties. A (.) (full stop) denotes conservation between groups with weakly similar properties.

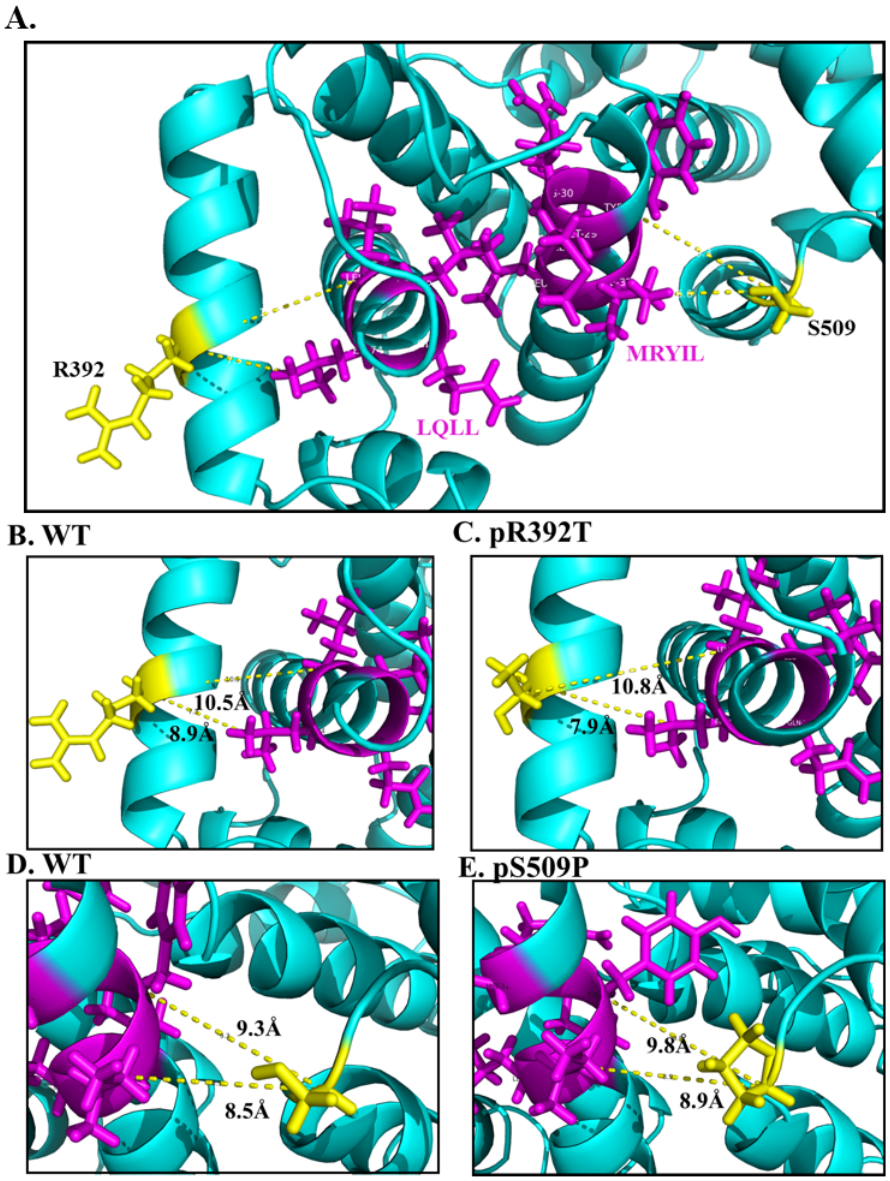


FIGURE 4.7: **A close up of the hydrophobic patch in cyclin box in wild type and mutant cyclin F as modelled in I-TASSER and visualised in Pymol. (A.)** The hydrophobic patch is shown coloured magenta. Amino acids 392 (Arg) and 509 (Ser), mutated in ALS patients (unpublished data), are shown in yellow. **(B.)**p.R392 is shown with the LQLL segment of the hydrophobic patch. **(C.)** The same segment is shown in the p.R392T mutant. **(D.)** p.S509 is shown with the MYRIL segment of the hydrophobic patch. **(E.)** The same segment is shown in the p.S509P mutant.

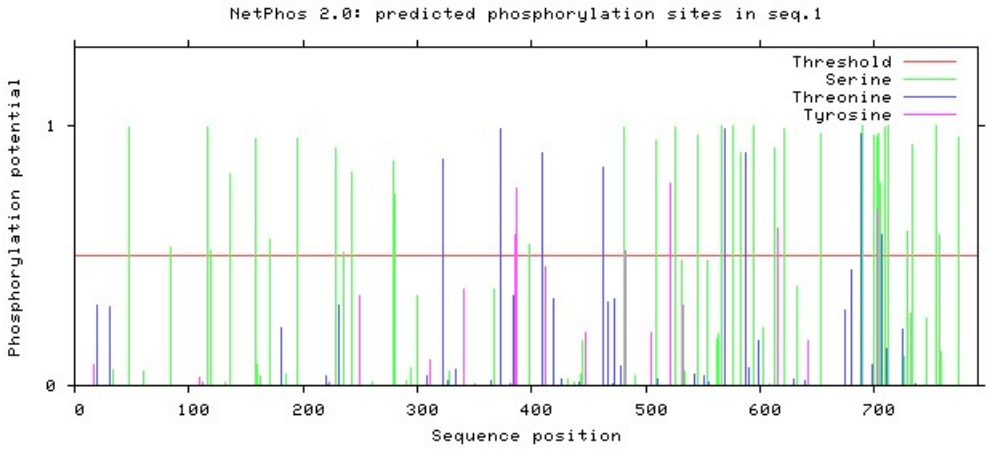


Figure 4.8: **A screenshot of the predicted phosphorylation sites of wild type cyclin F as determined in NetPhos.** Twenty-three potential phosphorylation sites cluster in the PEST domain from residues 581 to 766.

TABLE 4.1: ***In silico*** **prediction of the likelihood of pathogenesis of ALS-linked *CCNF* mutations occurring in the cyclin box and PEST domains.**

Amino acid change	**SIFT**	**PolyPhen2**	**MutationTaster**	**NetPhos**
R392T	Tolerated; 0.07	Probably damaging: 0.990	Disease causing; 0.995	No change
S509P	Affect*;0.01	Possibly damaging: 0.668	Disease causing; 0.999999	Loss of phosphorylation site
S621G	Affect*;0.00	Benign, 0.227	Disease causing; 0.9999	Loss of phosphorylation site
E624K	Affect*;0.00	Prob Damaging; 1.00	Disease causing; 0.9999	No change

* Denotes predictions have low confidence.

4.6 Developing *CCNF* deletion clones for functional analysis

4.6.1 Overview

The second part of this project aimed to generate cyclin F deletion constructs, with specific domains removed from internal regions of the cyclin F protein, leaving the rest of the protein intact (Figure 4.9). To generate these deletion fragments, mutagenesis via inverse PCR (*i*PCR) was employed.

4.6.2 TA cloning *CCNF* cDNA into the pGEM-T entry vector

Wild-type human *CCNF* cDNA was successfully cloned into the pGEM-T vector with *XhoI* and *EcoRI* restriction sites introduced to flank the open reading frame (see the vector map 4.10). This was confirmed by the correct amplification of the predicted band size (Figure 4.11) and by Sanger sequencing the entire construct. The ligated control insert failed to grow on the agar plates so could not be PCR amplified for use as a positive control.

4.6.3 Site-directed mutagenesis to create deletion constructs

The pGEM-T_*CCNF*_WT plasmid construct was subjected to Q5 Site-directed mutagenesis (Section 3.3.11) with primers designed to remove entire protein domains (see Figure 4.9). The *i*PCR mutagenesis products were subsequently transformed (Section 3.3.7) and propagated (Section 3.3.9). To confirm the mutagenesis reaction had successfully deleted the required protein domains, several transformed colonies were selected to be PCR amplified using the T7 promoter_F and SP6 primers, to amplify the deletion *CCNF* cDNA (Section 3.3.8). PCR products were visualised on an agarose gel (Figure 4.12) to ensure selected colonies were of the correct size and therefore expressed the desired gene (Section 3.3.8). Estimated *CCNF* deletion cDNA PCR product sizes are shown in Figure 4.12. Samples with sizes similar to the WT construct were run on a gel with the WT to show relative size.

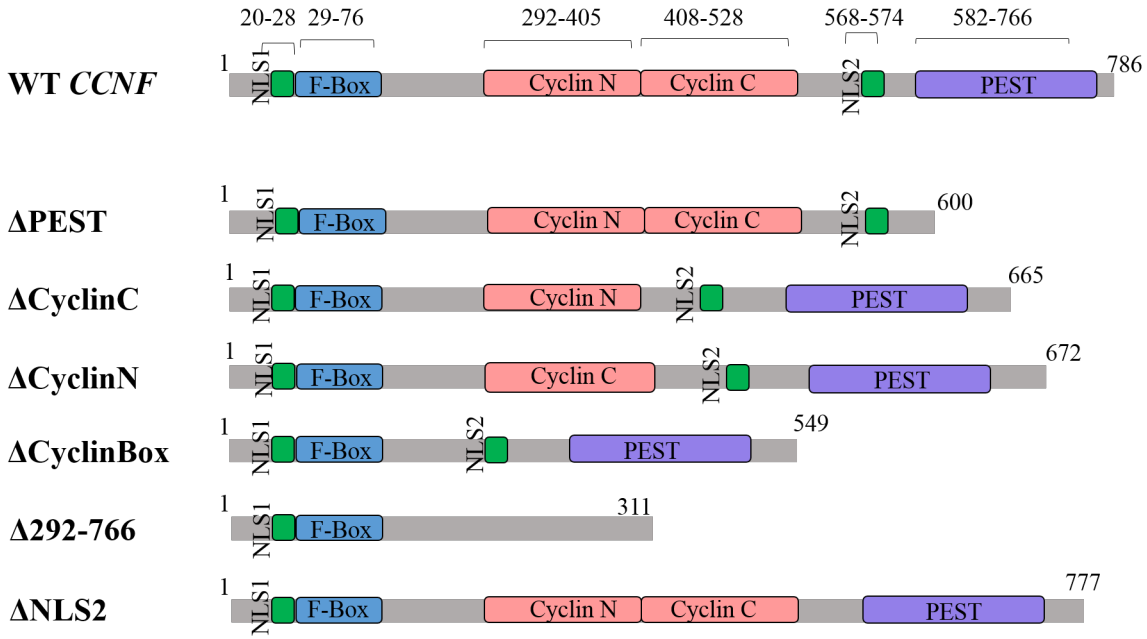


FIGURE 4.9: **Schematic representation of wild type *CCNF* and the six deletion *CCNF* cDNA fragments to be generated in this project.** WT:Wild type *CCNF*; Δ indicates the removal of the given domain or amino acids. F-box: F-box motif responsible for binding to the SCF complex; NLS1: nuclear localisation signal 1; NLS2: nuclear localisation signal 2; Cyclin N: cyclin N region of the cyclin box, containing the substrate binding region; cyclin C region of the cyclin box; PEST: the PEST sequence containing a stretch of proline, glutamic acid, serine and threonine amino acids bordered by positively charged residues.

The pGEM-T_*CCNF*_WT. plasmid DNA was purified and quantitated (Section 3.3.9), with sufficient concentrations and sufficient A260/A280 values to continue with the deletion cloning (Table 4.2).

TABLE 4.2: **Concentrations of purified pGEM-T_*CCNF* deletion constructs.**

Construct	Concentration(ng/μl)	A260/A280
pGEM-T_*CCNF*_ΔPEST	89.3	1.84
pGEM-T_*CCNF*_ΔCyclinN	67.1	1.91
pGEM-T_*CCNF*_ΔCyclinC	61.4	1.91
pGEM-T_*CCNF*_ΔCyclinBox	108	1.88
pGEM-T_*CCNF*_Δ262-766	162	1.86
pGEM-T_*CCNF*_ΔNLS2	201.8	1.86

Sequencing primers were designed previously in our laboratory to create overlapping sequence fragments spanning the entire *CCNF* gene region and pGEM-T vector entry

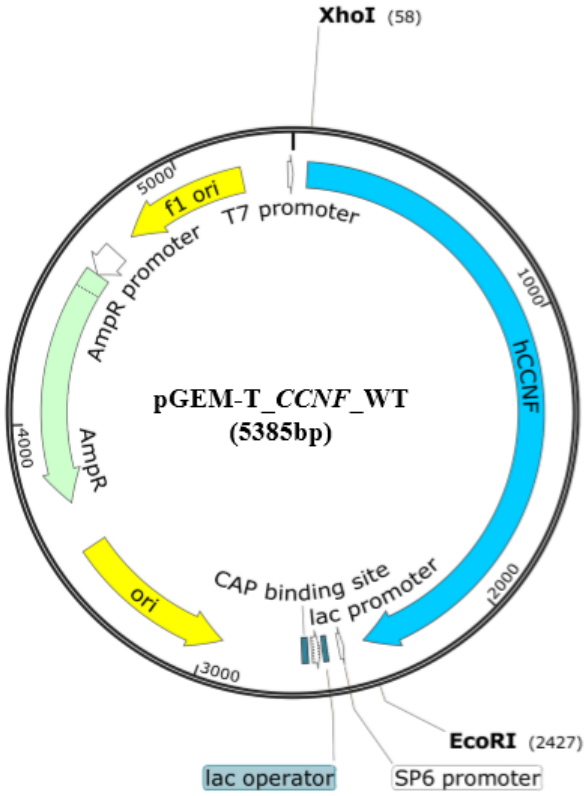


Figure 4.10: **Vector map of the generated wild type entry vector, pGEM-T_*CCNF*_WT.** The map shows the *CCNF* gene with *XhoI* restriction site at the 5′ end, and *EcoRI* restriction site at the 3′ end. The vector retains most of the intrinsic pGEM-T features such as T7 and SP6 promoters flanking the *CCNF* gene insert, and the ampicillin resistance gene (AmpR). Figure generated using SnapGene Viewer software 2.8.1 (GSL Biotech; available at snapgene.com)

points (Section 3.3.10), to ensure that the correct segment of *CCNF* gene had been deleted and that no undesired mutations had been introduced into the vector backbone restriction sites. As several of the *CCNF* deletion constructs were missing regions where these primers bind, new primers were designed to cover these areas. At least one colony from each pGEM-T_*CCNF* construct was confirmed to carry the correct sequence. Figure 4.13 shows the sequencing chromatograms spanning the deletion segments of the pGEM-T_*CCNF* deletion constructs.

As can be seen in Figure 4.14, the size of the *CCNF* deletion fragments corresponded closely to the expected sizes. However, the pGEM-T_*CCNF* constructs all had an extra band measuring between 4–6kb in size. For ΔPEST, ΔCyclinBox, ΔCyclinN and

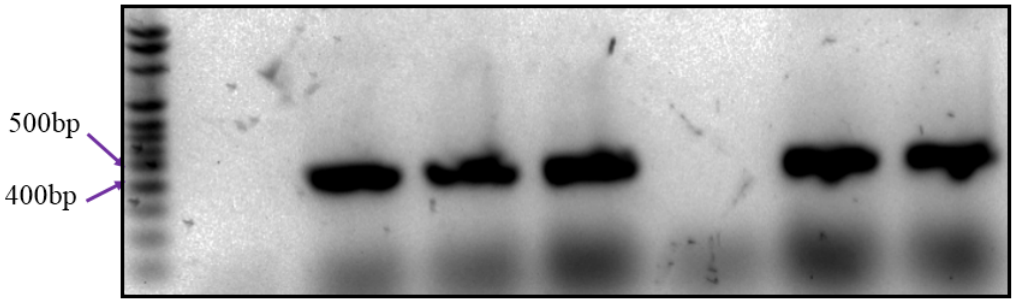


FIGURE 4.11: **Check colony PCR to confirm *CCNF* correctly inserted into pGEM-T_*CCNF*_WT.** PCR products amplified using primers T7 and K97_B_R are shown on gels electrophoresed at 100V for 45 minutes. All lanes contain PCR products from different colonies of pGEM-T_*CCNF*_WT

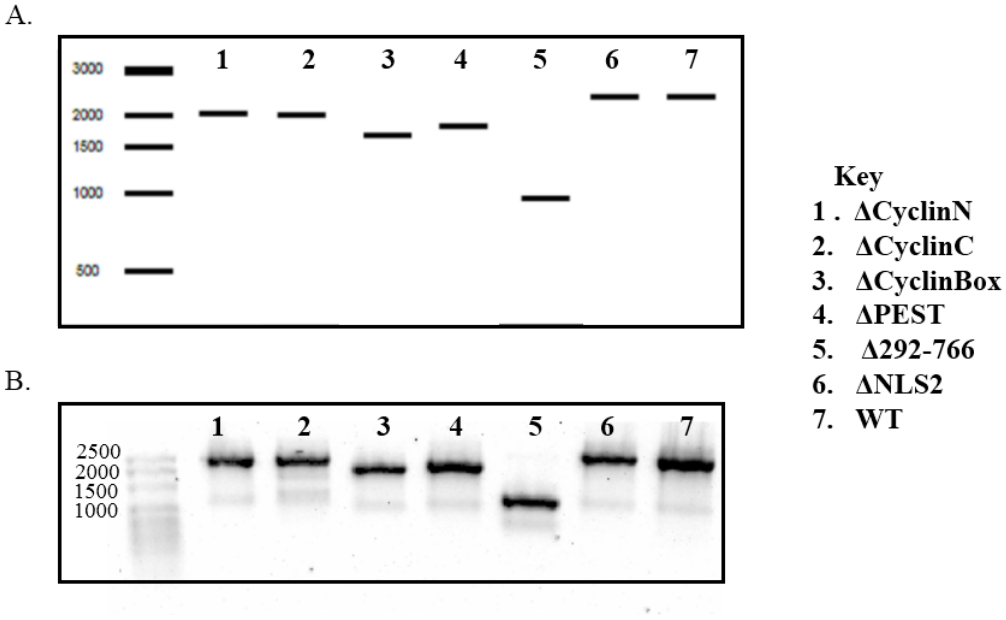


FIGURE 4.12: **PCR amplification of *CCNF* cDNA from pGEM-T_*CCNF* deletion constructs.** The seven different pGEM-T_*CCNF* constructs were PCR amplified using primers T7 & SP6. The PCR products were electrophoresed on an agarose gel at 100V for 45 minutes. **(A.)** A schematic showing the expected band sizes. **(B.)** An image of the agarose gel showing the band sizes closely match the expected band sizes. WT: pGEM-T_*CCNF*_WT; Δ: denotes the removal of a given sequence.

ΔCyclinC, the band measured 5kB; while for ΔNLS2 it measured >5kb. These sizes match the approximate size of the full pGEM-T_*CCNF* constructs, suggesting a partial restriction digestion. The Δ292-766 construct restriction digest displayed only two clear bands. However, the larger band was denser compared to all others, measuring between 3-4kb, indicating it may include a mixture of the digested 3kB pGEM-T vector as well as undigested 4kB pGEM-T_*CCNF*_Δ292-766 construct. Despite this, the insert bands were clearly enough delineated to allow excision via scalpel, with subsequent gel purification resulting in sufficient concentrations to continue with the ligation (Table 4.3). The digested pmCherry-C1 vector (Clontech,USA) was the correct size, and was gel purified separately for each ligation reaction.

Table 4.3: **Concentrations of gel-purified restriction digest products.**

Construct	**Concentration(ng/μl)**	**A260/A280**
*CCNF*_ΔPEST	10.6	1.92
*CCNF*_ΔCyclinN	19.8	1.98
*CCNF*_ΔCyclinC	22.6	1.87
*CCNF*_ΔCyclinBox	12.3	2.07
*CCNF*_Δ262-766	12.3	1.96
*CCNF*_ΔNLS2	17.2	1.89
$pmCherry - C1^{*}$	40.1	1.87
$pmCherry - C1^{**}$	60.2	1.85
$pmCherry - C1^{***}$	33.7	1.93

* used with ΔCyclinN, ΔCyclinC; ** used with ΔPEST, ΔCyclinBox, Δ292-766, *** used with ΔNLS2

4.6.4 PCR validation and purification of pmCherry-C1_*CCNF* deletion constructs

To confirm that *CCNF* had successfully ligated into the pmCherry-C1 destination vector, a forward primer in the 5′ mCherry sequence and an internal *CCNF* reverse primer were chosen to PCR amplify a segment 425bp in length (Figure 4.15). The PCR products were subsequently electrophoresed on an agarose gel to ascertain whether the correct fragment had been amplified. The restriction cloning was successful for all constructs, except for the pmCherry-C1_*CCNF*_ΔNLS2 construct

(Figure 4.16). A repeat of the PCR using additional colonies also failed, suggesting a failure at the gel extraction stage, likely due to an inability to successfully separate the digested and undigested entry vector.

Two colonies were chosen from each of the five successfully amplified constructs, the DNA purified and sequenced. Sequencing primers were designed to create overlapping segments covering the whole *CCNF* insert as well as the mCherry sequence at the 5′ end of *CCNF* insert and the vector sequence at the 3′ end of the insert. A vector map and the five *CCNF* deletion inserts are displayed in Figure 4.17

The five successfully generated pmCherry-C1_*CCNF* deletion constructs, as well as pmCherry-C1_*CCNF*_WT, were all propagated and purified (Section 3.3.16) with sufficiently high yields (Table 4.4) for use in *in vitro* studies.

TABLE 4.4: **Concentrations of purified pmCherry-C1_*CCNF* constructs.**

Construct	**Concentration(ng/μl)**	**A260/A280**
pmCherry-C1_*CCNF*_ΔPEST	821.1	1.89
pmCherry-C1_*CCNF*_ΔCyclinN	980	1.89
pmCherry-C1_*CCNF*_ΔCyclinC	762.5	1.88
pmCherry-C1_*CCNF*_ΔCyclinBox	1474.5	1.89
pmCherry-C1_*CCNF*_Δ262-766	2244.1	1.9
pmCherry-C1_*CCNF*_WT	518.1	1.89

Therefore, a total of 11 *CCNF* constructs were generated in the current project (Table 4.5) for use in *in vitro* assays.

Table 4.5: **Summary of all *CCNF* constructs generated.**

Construct	Antibiotic resistance
pGEM-T_*CCNF*_ΔPEST	Ampicillin
pGEM-T_*CCNF*_ΔCyclinN	Ampicillin
pGEM-T_*CCNF*_ΔCyclinC	Ampicillin
pGEM-T_*CCNF*_ΔCyclinBox	Ampicillin
pGEM-T_*CCNF*_Δ262-766	Ampicillin
pGEM-T_*CCNF*_ΔNLS2	Ampicillin
pmCherry-C1_*CCNF*_ΔPEST	Kanamycin
pmCherry-C1_*CCNF*_ΔCyclinN	Kanamycin
pmCherry-C1_*CCNF*_ΔCyclinC	Kanamycin
pmCherry-C1_*CCNF*_ΔCyclinBox	Kanamycin
pmCherry-C1_*CCNF*_Δ292-766	Kanamycin

4.7 Deletion of domains in cyclin F leads to nuclear or cytoplasmic aggregations as well as toxicity

4.7.1 Sub-cellular localisation of wild type, mutant and deletion Cyclin F

To study the effect of deletions in cyclin F on its sub-cellular localisation, two neuronal cell lines, NSC-34 and SH-SY5Y cells, were transfected with six pmCherry-*CCNF* plasmid constructs (Table 4.4). Due to the higher transfection efficiency of the SH-SY5Y cells (approximately 24%) compared to NSC-34 cells (approximately 8%), counting was conducted on the SH-SY5Y samples to identify the typical composition of cell populations. Populations were distinguished depending on whether cyclin F expression occurred in the cytoplasm, nucleus or both, and whether aggregates were present.

SH-SY5Y cells

It has been previously reported that cyclin F is expressed in both the nucleus and cytoplasm of cells (**Bai et al.**, 1994). Similarly, the current study found that the wild type cyclin F localised to the nucleus in ~55% of cells, to the cytoplasm in ~25% of cells and to both in ~20% of cells (Figures 4.18A & 4.19A-C). These expression patterns were not significantly different for the ΔCyclinN, ΔCyclinC or ΔCyclinBox

cyclin F constructs (Figure 4.19A-C). In contrast, the percentage of SH-SY5Y cells with predominantly nuclear expression was significantly higher for the ΔPEST mutant than for the wild type cyclin F (p<0.01)(Figure 4.19A). This was accompanied by significantly lower cytoplasmic expression (p<0.01) (Figure 4.19B). Conversely, the Δ292-766 construct had significantly lower nuclear expression (p<0.0001) and higher cytoplasmic expression (p<0.01) than wild type cyclin F (Figure 4.19A & B).

The ΔPEST construct also showed evidence of large and punctate nuclear aggregations (Figure 4.18B & C). Approximately 47.97% (± 6.5%) of cells had intranuclear aggregations, a significantly higher proportion than for wild type cyclin F (p<0.0001) (Figure 4.19D).

NSC-34 cells

Transfections of the six pmCherry-*CCNF* plasmid constructs (Table 4.4) were repeated in NSC-34 cells because this cell line has motor neuron features (**Cashman et al.**, 1992). Similar to the SH-SY5Y cells, wild type cyclin F localises to both the cytoplasm (Figure 4.20A) and the nucleus (Figure 4.20B) in NSC-34 cells, with little evidence of protein aggregation. The ΔCyclinN, ΔCyclinC or ΔCyclinBox cyclin F constructs localised to both the cytoplasm and nucleus like the wild type cyclin F, while theΔPEST cyclin F was predominantly nuclear and Δ292-766 predominantly cytoplasmic. Due to the low level of transfection efficiency, these cells were not counted to determine if these differences were statistically significant.

The deletion of the PEST domain resulted in both punctate (arrow) and large (arrow head) protein aggregations in the nucleus(Figure 4.20C). Interestingly, the ΔCyclin C construct also showed large intranuclear aggregates (arrow head) in a small number of cells (Figure 4.20E). The cellular localisation of the ΔCyclinN and ΔCyclin Box constructs was very similar, with some punctate aggregation formation in the cytoplasm (Figure 4.20D & F). The Δ292-766 construct showed predominantly weak diffused cytoplasmic expression and no aggregates were found for this construct (Figure 4.20G).

The presence of aggregates in ΔCyclinC, ΔCyclinN and ΔCyclinBox cyclin F-transfected cells suggests that the formation of aggregates could be motor neuron

specific. As a consequence, the NSC-34 cell line was used for the following assays.

4.7.2 ΔPEST and ΔCyclinBox are more toxic than WT Cyclin F when expressed in NSC-34 cells

Gating parameters were established using the negative control (mock transfected cells) (Figure 4.21), and applied to all samples transfected with cyclin F constructs. Representative SYTOX Blue positive populations are shown in Figure 4.22A.

A significant increase in toxicity in ΔPEST (p<0.001) and ΔCyclinBox (p<0.05) cyclin F-transfected cells was observed compared to wild type-transfected cells (Figure 4.22B). Although not significant, ΔCyclin N-cyclin F transfected cells showed a higher proportion of SYTOX Blue population compared to the wild type population (p=0.05154) (Figure 4.22B). Conversely, the percentage of SYTOX Blue positive population in Δ292-766-transfected cells was significantly smaller than wild type (p<0.05) (Figure 4.22B).

Differences in mCherry expression between wild type and deletion cyclin F were then determined (Figure 4.23). mCherry expression was presented as a percentage of mCherry positive cells to all cells in P1 gate. As shown in Figure 4.23, the ΔPEST mutant had a significantly higher mCherry expression than the wild type (p<0.01). The three cyclin mutants also showed significantly higher mCherry expression than the wild type (p<0.05). Conversely, the Δ292-766-cyclin F transfected cells had a significantly lower transfection rate than WT-transfected cells (p<0.0001).

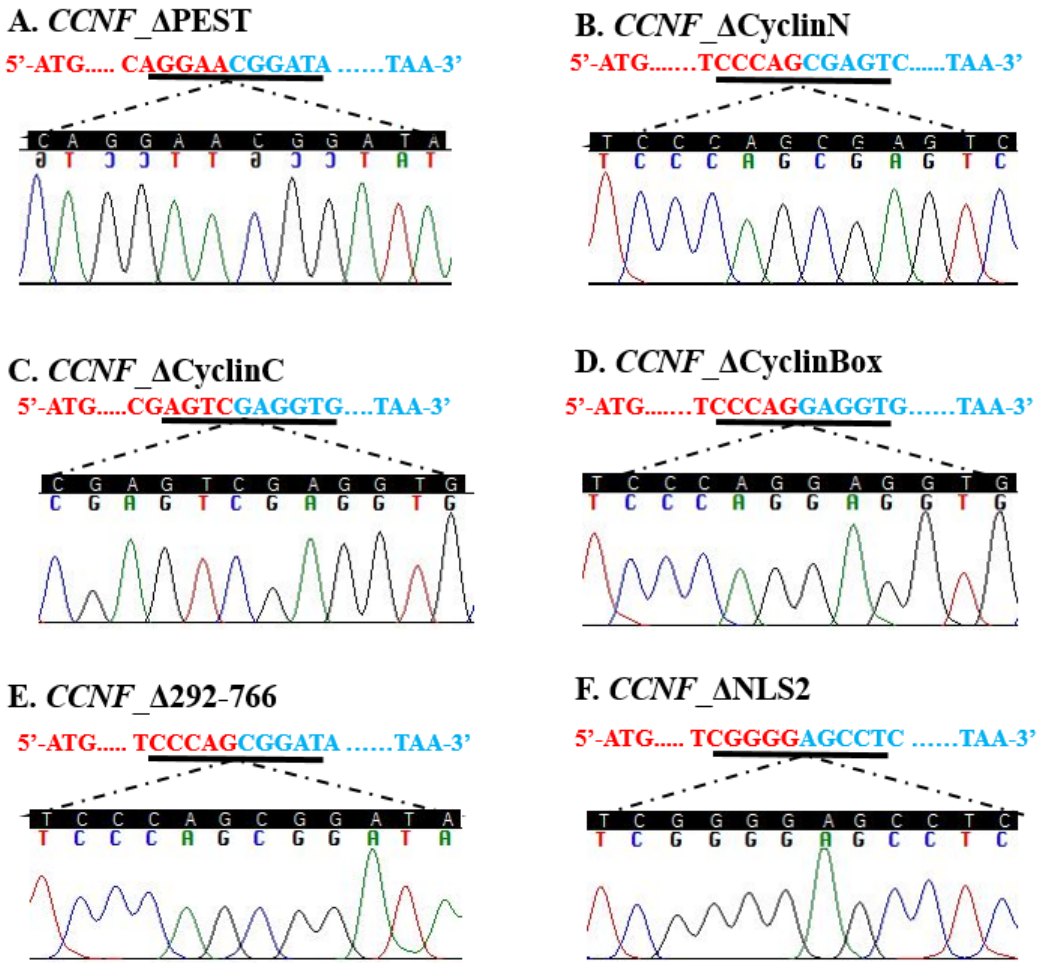


FIGURE 4.13: **Sequencing chromatograms of the nucleotides flanking the deleted region for the six *CCNF* deletion constructs.**The sequence above each chromatogram shows the expected sequence of the construct spanning the deletion region. The sequence in blue denotes sequence at the 5' end of the deleted segment. The sequence in red denotes the sequence at the 3' end of the deleted segment.

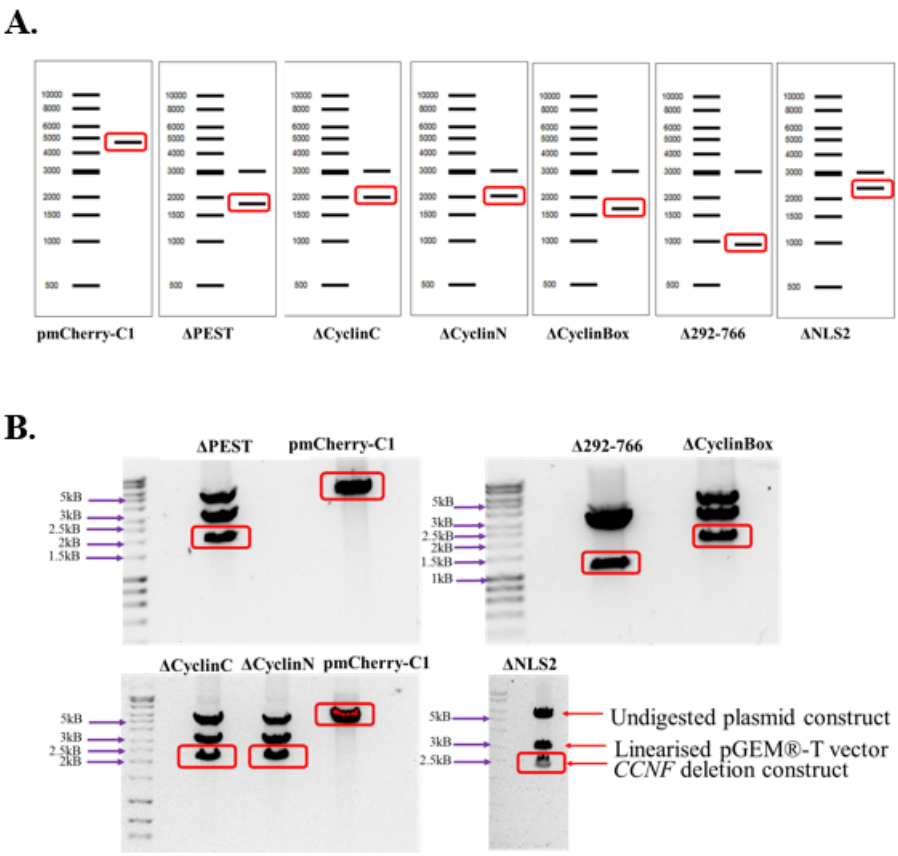


Figure 4.14: **The expected and actual gel band sizes of *XhoI* and *EcoRI* restriction digested pGEM-T_*CCNF* deletion clones and pmCherry-C1 empty vector. A.** The bands to be excised for gel purification (Section 3.3.4) as determined in Serial cloner (version 2.6) are shown circled in red. For the *CCNF* deletion constructs, the band of 3000kb demarks the pGEM-T vector backbone. **B.** Red boxes indicate the section of gel excised and purified for ligation. The extra band at the top of the pGEM-T_*CCNF* clones is likely undigested or partially digested plasmid construct.

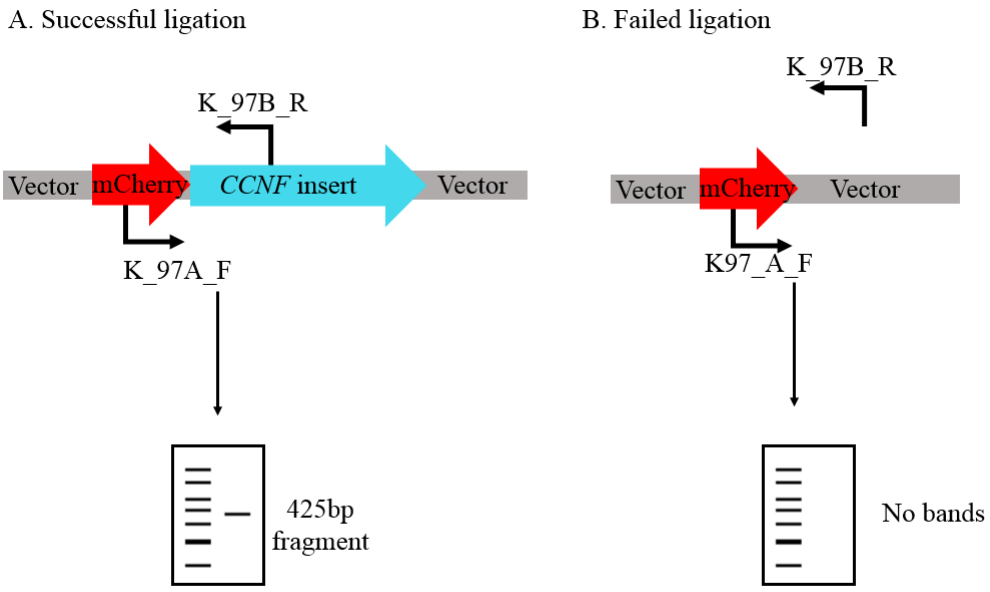


FIGURE 4.15: **Check colony PCR. (A.)** If the ligation has been successful, the forward primer will bind in the mCherry sequence, and the reverse primer will bind in the *CCNF* insert. A 425bp amplicon will be generated by PCR and measured on an agarose gel. **(B.)** If the ligation has failed, the reverse primer cannot bind, and no PCR product will be generated resulting in a lack of visible bands on an agarose gel.

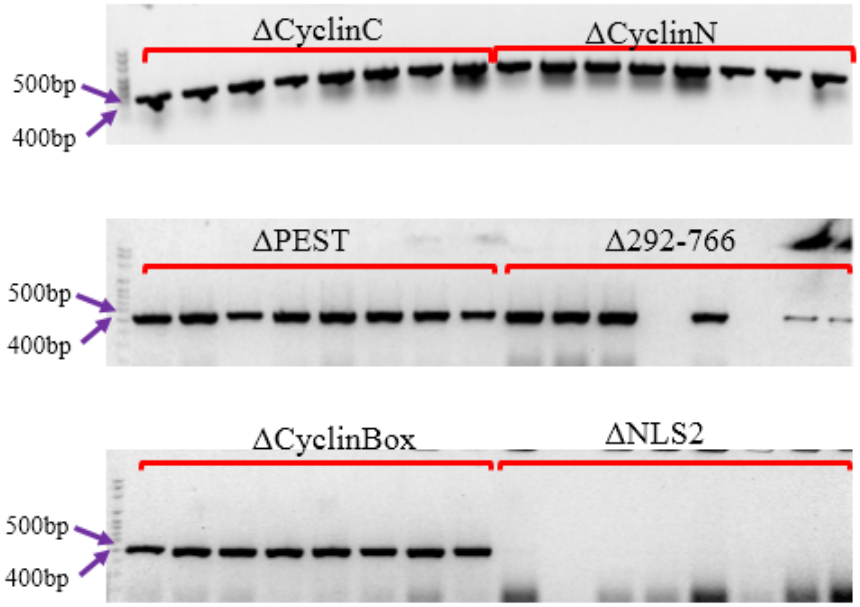


Figure 4.16: **PCR amplification of *CCNF* deletion fragments from pmCherry-C1 vector.** A representative agarose gel showing the PCR products of *CCNF* generated using a forward internal primer in the mCherry sequence and a reverse internal *CCNF* primer, with the Hyperladder 100bp plus ladder (Bioline, Australia). The expected band size was 425bp. PCR amplification was successful for five constructs, but not for pmCherry-C1_*CCNF*_ΔNLS2.

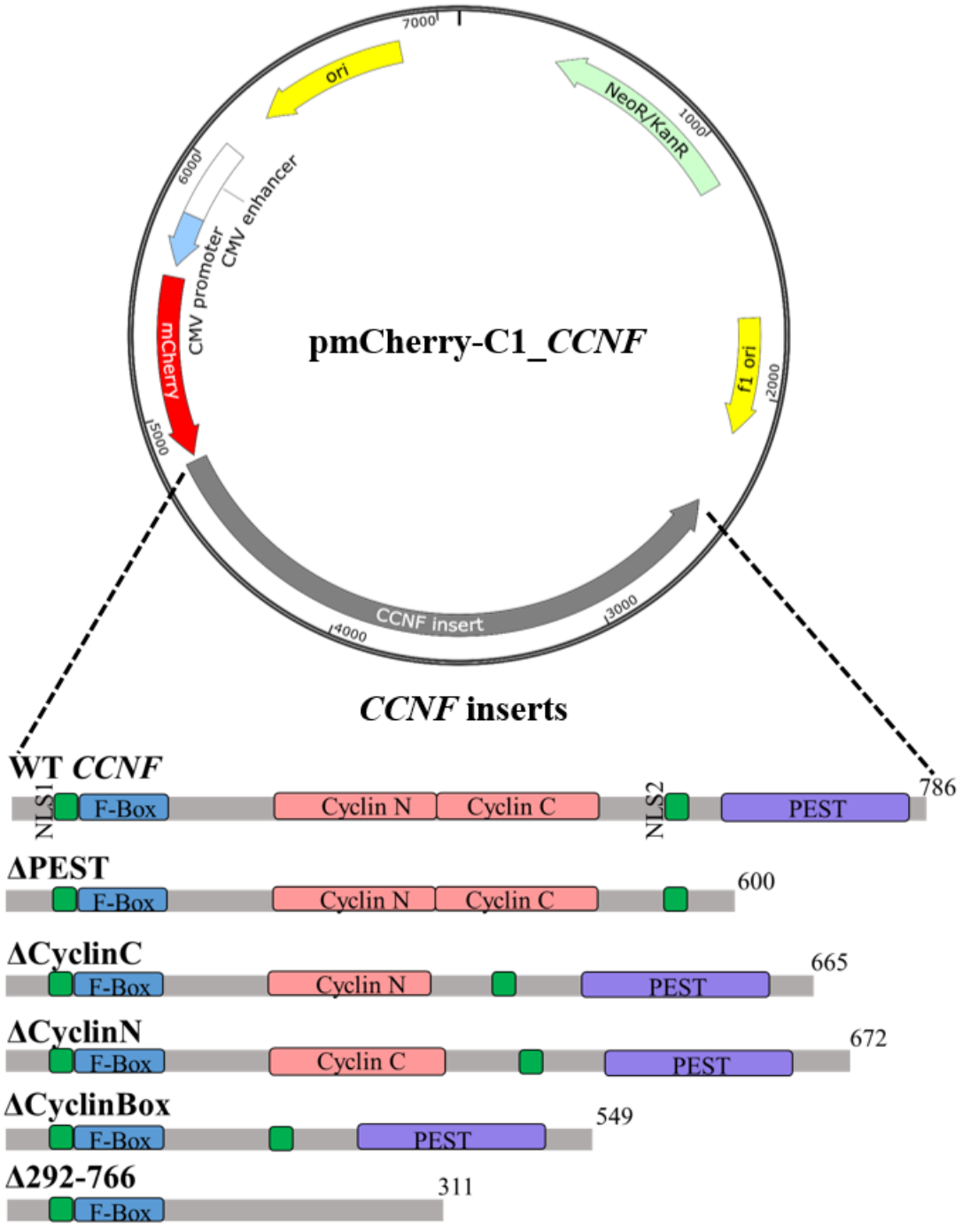


FIGURE 4.17: **The pmCherry-C1_*CCNF* vector with *CCNF* inserts shown underneath.** WT: wild type cyclin F. Δ denotes the deletion of a particular domain. Vector constructed using SnapGene Viewer software 2.8.1 (GSL Biotech; available at snapgene.com).

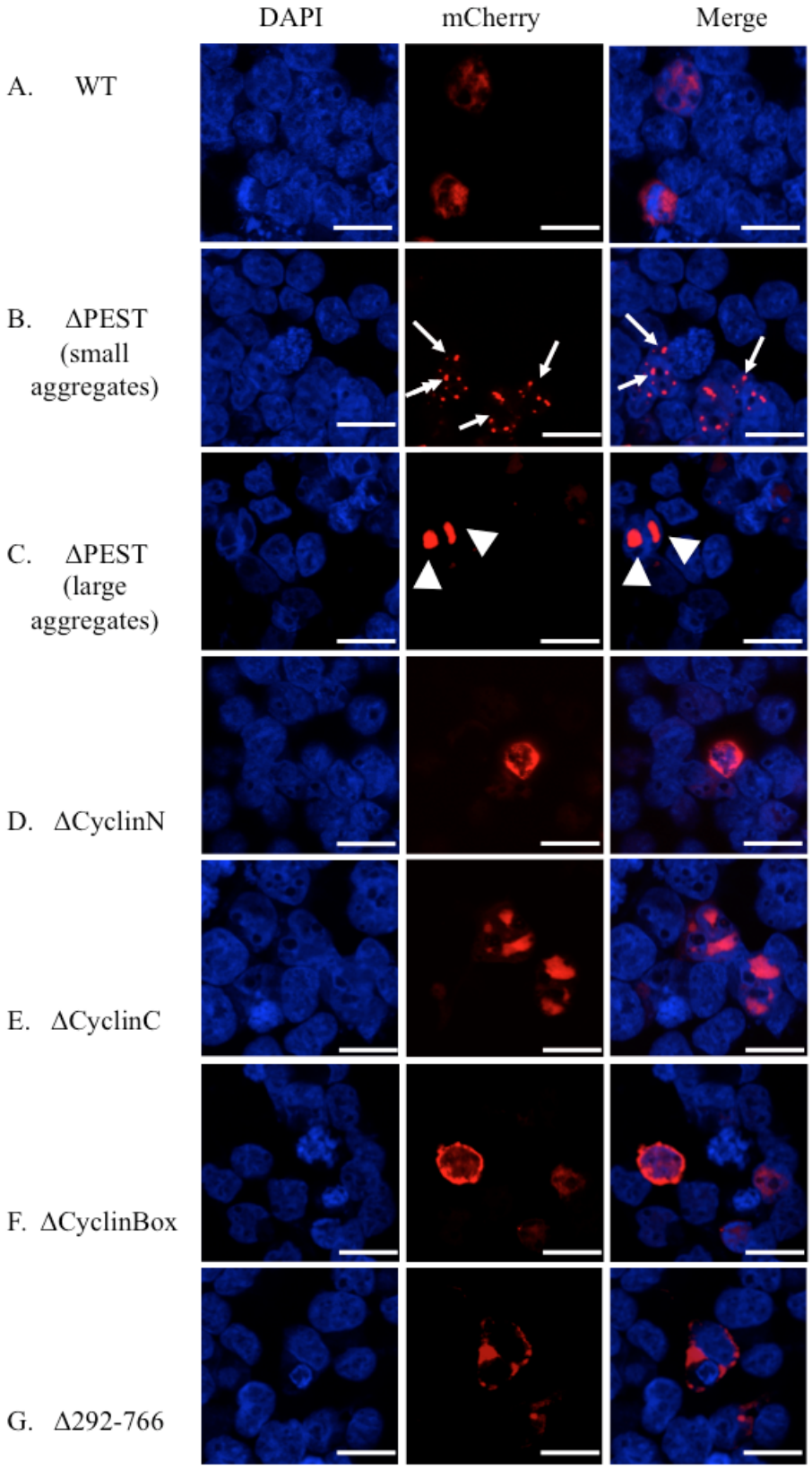


Figure 4.18: **Sub-cellular localisation of transfected wild-type and mutant cyclin F in SH-SY5Y cells.** From left to right, images are shown with the DAPI filter, the mCherry filter and merged. Arrows indicate punctate aggregations. Arrow heads indicate large aggregations. Scale bar=20μM

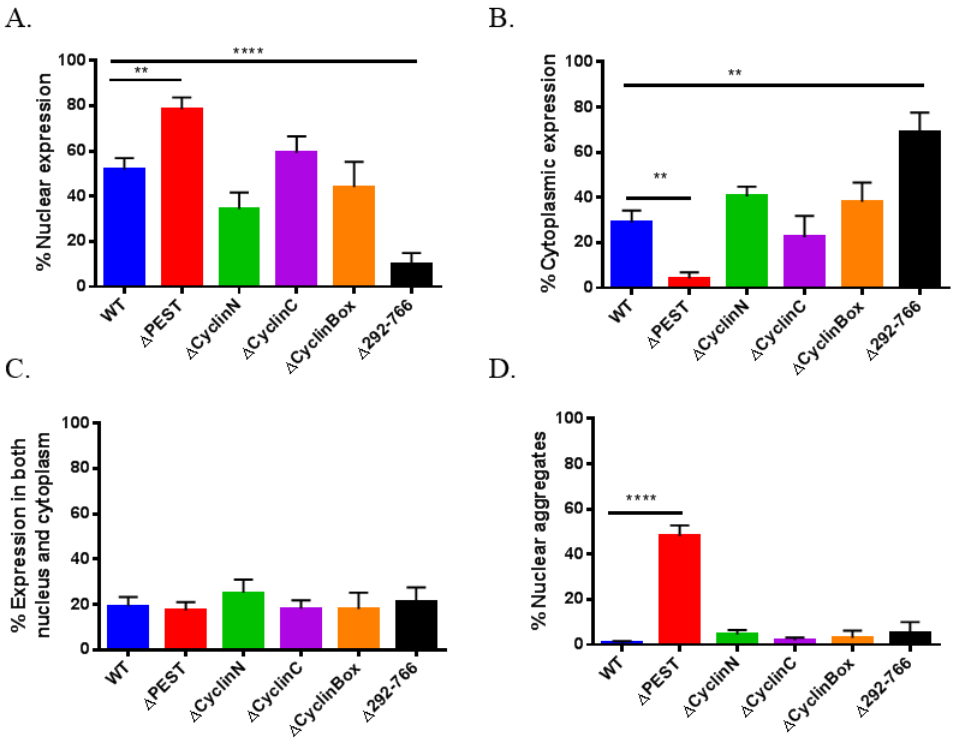


FIGURE 4.19: **Sub-cellular localisation and aggregation formation in cyclin F constructs in SH-SY5Y cells.** Cell counting was conducted in SH-SY5Y cells. Mean values are based on at least eight images randomly acquired from two experimental repeats. At least 80 cells were counted. Error bars indicate S.E.M. $^{**}p<0.01$; **** $p<0.0001$.

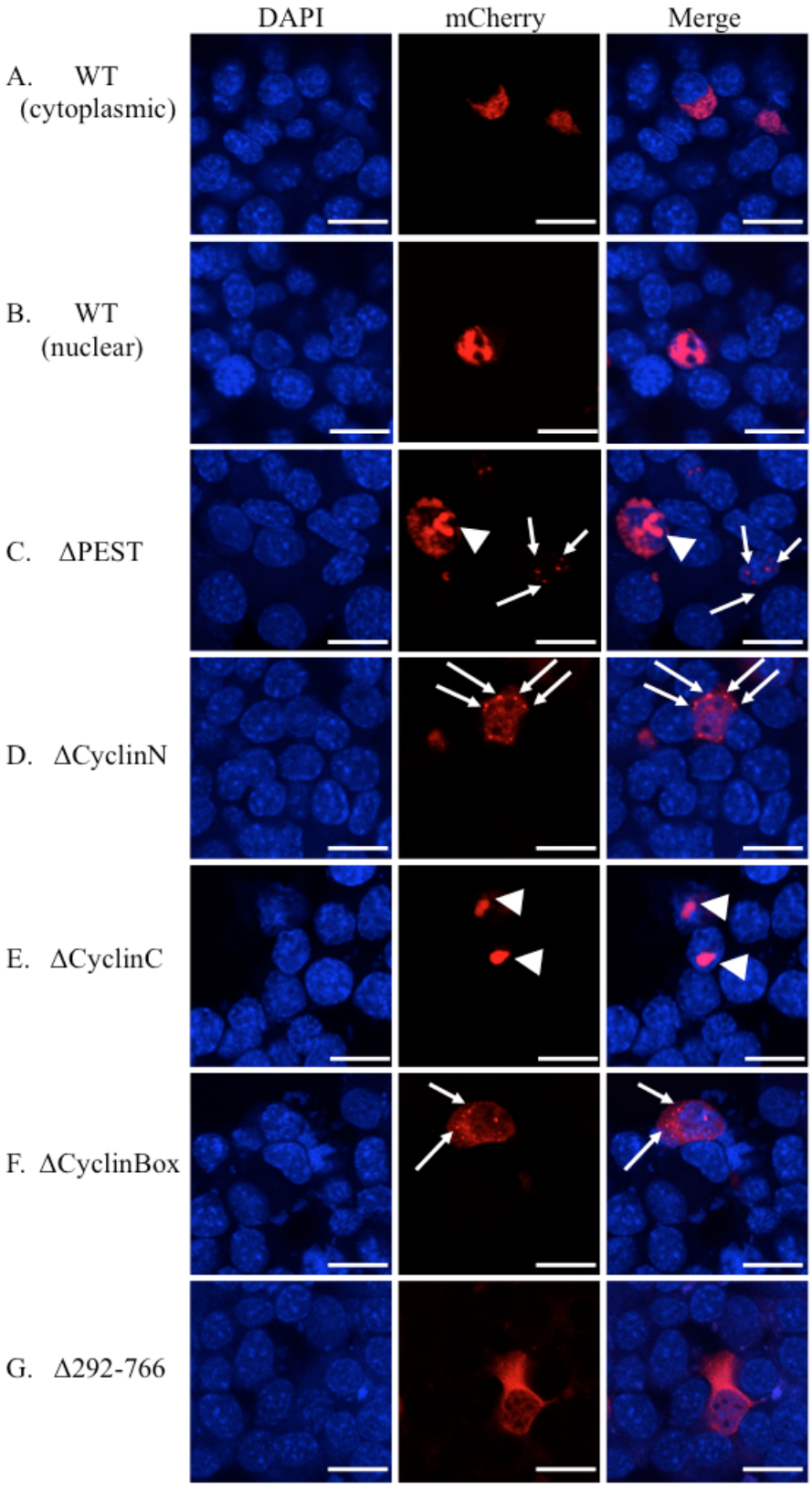


Figure 4.20: **Sub-cellular localisation of transfected wild-type and mutant cyclin F in NSC-34 cells.**From left to right, images are shown with the DAPI filter, the mCherry filter and merged. Arrows indicate punctate aggregations. Arrow heads indicate large aggregations. Scale bar=20μM

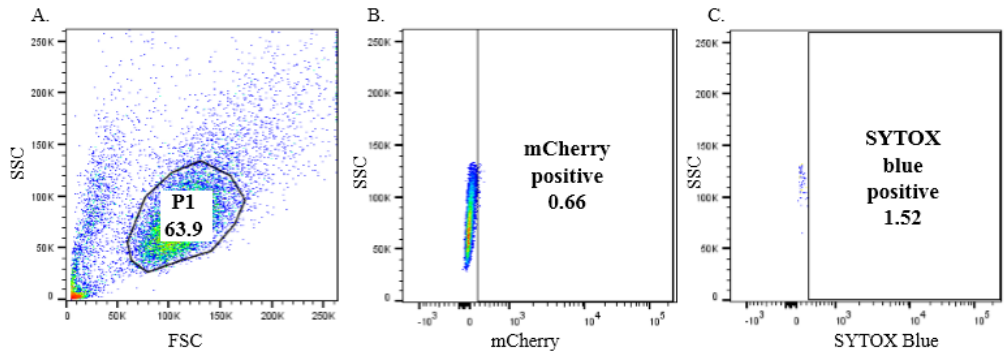


FIGURE 4.21: **Gating set up using negative controls (lipofectamine (Invitrogen) with no DNA). (A.)** A gate for the parent population (denoted P1) was first applied to exclude debris and doublets by plotting FSC data versus SSC data in the negative control. **(B.)** Gate mCherry positive is a subpopulation of P1. This gate defines the mCherry positive population, in which the majority of the negative control population is negative for mCherry expression. **(C.)** Gate SYTOX Blue positive is a subpopulation of gate mCherry positive. This gate defines the SYTOX Blue population, in which the majority of the population is negative for SYTOX Blue staining.

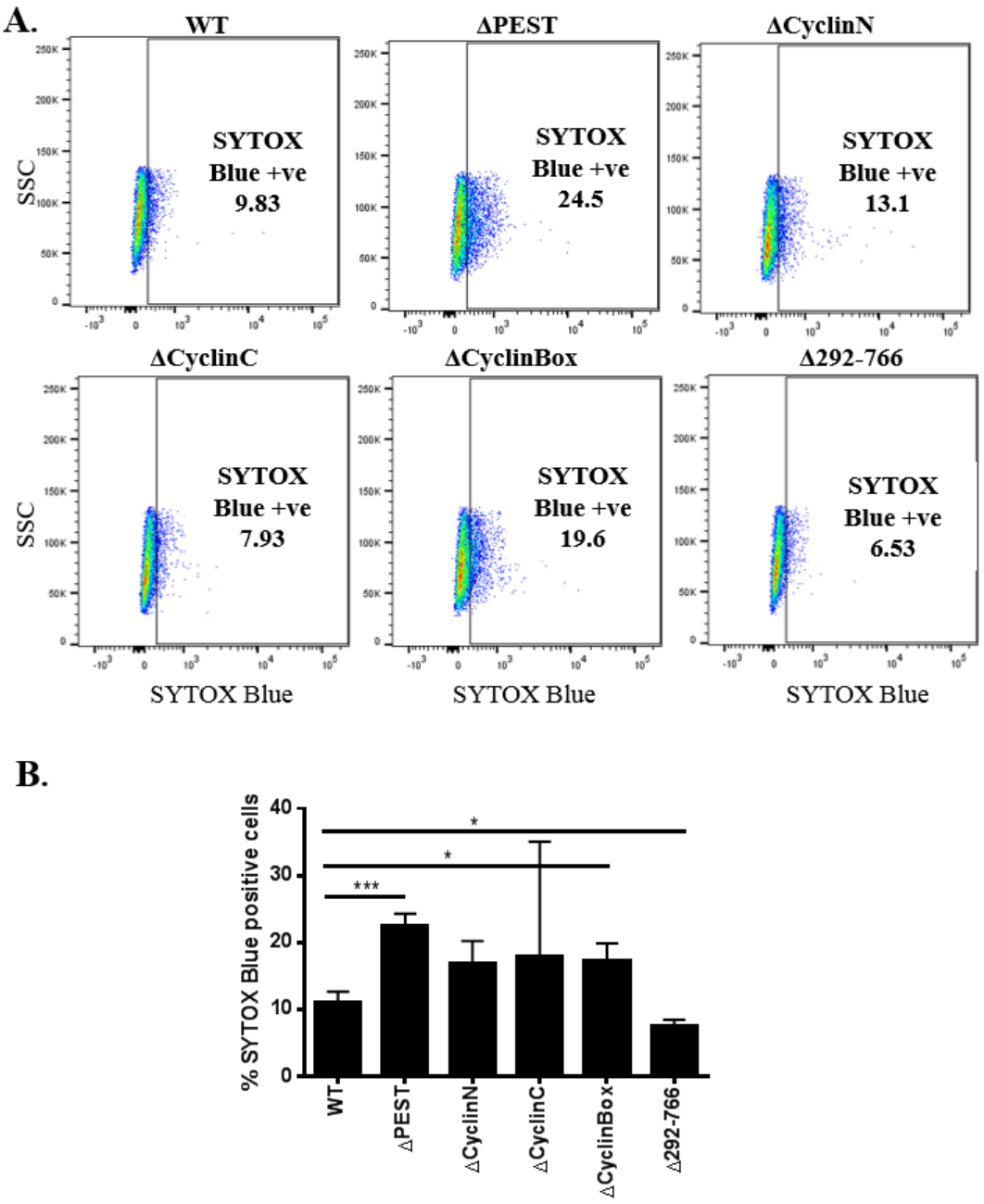


Figure 4.22: **Toxicity of cyclin F-transfected NSC-34 cells. (A.)** Representative dot plots of SYTOX Blue staining in mCherry positive cell populations. **(B.)** Quantification of SYTOX Blue positive population. The ΔPEST and ΔCyclinBox-transfected cells had a significantly higher SYTOX Blue staining compared to wild type. The Δ292-766-transfected NSC-34 cells had a significantly lower staining of cell death compared to wild-type. The values represent mean ± S.E.M obtained by pooling 3 samples together. Error bars show S.E.Ms (*** $p<0.001$, * $p<0.05$. 10,000 events were measured, n=3).

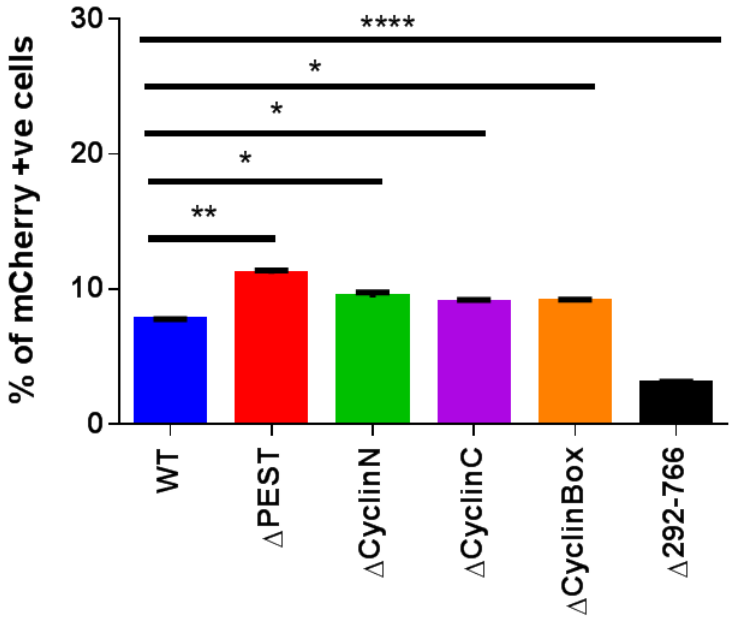


FIGURE 4.23: **mCherry expression of cyclin F constructs in NSC-34 cells.** The values represent mean ± S.E.M obtained by pooling 3 samples together. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$. 10000 events were measured. n=3.

5 Discussion

5.1 ALS

ALS is a devastating disease that is characterised by the progressive degeneration of motor neurons leading to paralysis and death within 2-5 years of diagnosis (**Robberecht and Philips**, 2013). At present diagnosis is difficult; and no treatments are available to slow progression. The pathological mechanisms of ALS and FTD are poorly understood, with the only known causes being genetic mutations that lead to motor neuron degeneration (**Renton et al.**, 2014). By studying the effects of such gene mutations on protein structure and function, it is hoped that a general pathophysiological model of ALS disease progression can be built. Understanding these disease mechanisms could potentially open up avenues for future research into more effective diagnostic, prognostic and treatment options. In particular, finding methods for earlier diagnosis would transform the health care provided to patients; giving certainty of diagnosis to patients and their families from the outset and allowing clinical trials to be conducted before patient neural tissues have been irreparably damaged (**Rosenfeld and Strong**, 2015).

Our laboratory recently identified a mutation in the cyclin F gene, *CCNF*, that segregated in an Australian ALS family (Table 1.2, Figure 1.1) (unpublished data.)

Worldwide collaborative efforts to screen FALS, SALS and FTD cohorts uncovered *CCNF* mutations at a prevalence of 0.6 to 3.3% globally, which were absent from public SNP databases and Australian control individuals (unpublished data). *CCNF* codes for an E3 ubiquitin ligase, cyclin F, which mediates the ubiquitination of proteins targeted for destruction by the ubiquitin proteasome pathway (**Skaar et al.**, 2013). Mutations in cyclin F may lead to impairment in the ubiquitination of proteins targeted for destruction, which could lead to the accumulation of cyclin F, its substrates or both into aggregates. Defining how mutations in cyclin F could impair protein degradation and cause ALS forms the basis of the current study. The first component of the study focused on creating 3D models of wild-type and mutant cyclin F using bioinformatics tools. The second component involved the generation of several cyclin F deletion constructs, which were expressed in neuronal cell lines to assess the effect of domain deletions on sub-cellular localisation and toxicity.

5.2 3D modelling of cyclin F

Bioinformatic tools were utilised successfully to create confident 3D models of the cyclin box domain of cyclin F. These models demonstrate the proximity of the ALS-linked mutations to the hydrophobic patch, suggesting they have the potential to affect substrate binding. Unfortunately, a confident model of the full-length cyclin F could not be obtained.

3D models of cyclin F were initially built using two template-based protein structure modelling programs, SWISS-MODEL and Aquaria (Section 4.2, Figures 4.1 & 4.2). The highest homology was found with cyclin A2 and cyclin B1, with 29% and 24% sequence identity to cyclin F respectively (Section 4.2, Figure 4.2). Between 20-35% sequence identity is considered the twilight zone of sequence alignment, and models based on these templates carry an increased risk of false positives (**Rost**, 1999). However, cross validation estimates in Aquaria suggest there is an 89% chance that cyclin F has a similar structure to cyclin A2. The fact that only the cyclin box domain of cyclin F sequence aligned to cyclin A2 and cyclin B1 was most probably the reason for the low confidence values given to the 3D models obtained with SWISS MODEL (Figure 4.1). This is supported by the fact that the 3D model constructed in I-TASSER when the cyclin box domain sequence was uploaded separately had a high confidence score, with a TM-score of 0.92 ± 0.06 (Figure 4.5).

As shown in Figures 4.1 & 4.2, the cyclin box domain of cyclin A2 and cyclin B1 consists of two clusters of alpha-helices: a five alpha helice-cluster containing the hydrophobic patch and seven alpha helice-cluster. The models show that the amino acids that are most highly conserved with cyclin F occur in the alpha helices, and in particular, the hydrophobic patch, suggesting a conservation in the binding behaviour of the three proteins (Figures 4.1 & 4.2). The areas of lowest conservation occur in the loop regions. The predicted 3D model of the cyclin domain of cyclin F as determined in I-TASSER (Figure 4.5) mimics these models closely (Figure 4.5). Manual alignments of cyclin F, cyclin A2 and cyclin B1 in ClustalOmega (**Sievers et al.**, 2011) confirmed the most highly conserved region of the three proteins was the hydrophobic patch (Figure 4.6). The hydrophobic patch is comprised of two short sequences of hydrophobic amino acids, MYRIL and LQLL, in adjacent α-helices (Figure 4.7). In cyclin A2 and cyclin B1, this hydrophobic patch is responsible for binding to an RXL (CY) motif in their corresponding CDK substrate (**D'Angiolella et al.**, 2013). While cyclin F does not have a known CDK binding partner, studies have confirmed that cyclin F binds substrates in a similar fashion (**D'Angiolella et al.**, 2010, 2012). The known cyclin F substrates, CP110 and RRM2, have each been shown to carry an RXL motif that is essential for binding to cyclin F (**D'Angiolella et al.**, 2010, 2012). Furthermore, co-immunoprecipitation studies conducted on cyclin F have confirmed that loss of the cyclin region, particularly cyclin N abolishes substrate binding, while mutations in the hydrophobic patch (p.M309A & p.L352A) significantly reduced substrate binding (**D'Angiolella et al.**, 2010, 2012; **Klein et al.**, 2015).

Two ALS-linked mutations have been found within the cyclin box domain of cyclin F: p.R392T in the cyclin N region and p.S509P in the cyclin C region (unpublished data) (Figure 1.1). One of the major benefits of 3D protein modelling is that it estimates the proximity of residues in real space. Amino acids far apart in amino acid sequence position can be physically close together when the protein is folded into its 3D structure. Therefore, to determine if cyclin box substitutions (p.R392T & p.S509)) could alter the structure and function of the cyclin box, the mutant sequences were uploaded to I-TASSER for individual modelling. The resulting models indicate that the two ALS-linked mutations lie in close proximity to the substrate binding domain in 3D space despite their distance from each other in the amino acid sequence (Section 4.2, Figures 4.7 & 5.1). It is conceivable that these mutations could alter the conformational structure of the two alpha helices that contain the hydrophobic patch, potentially altering substrate binding (Figure 4.7). It is also important to remember

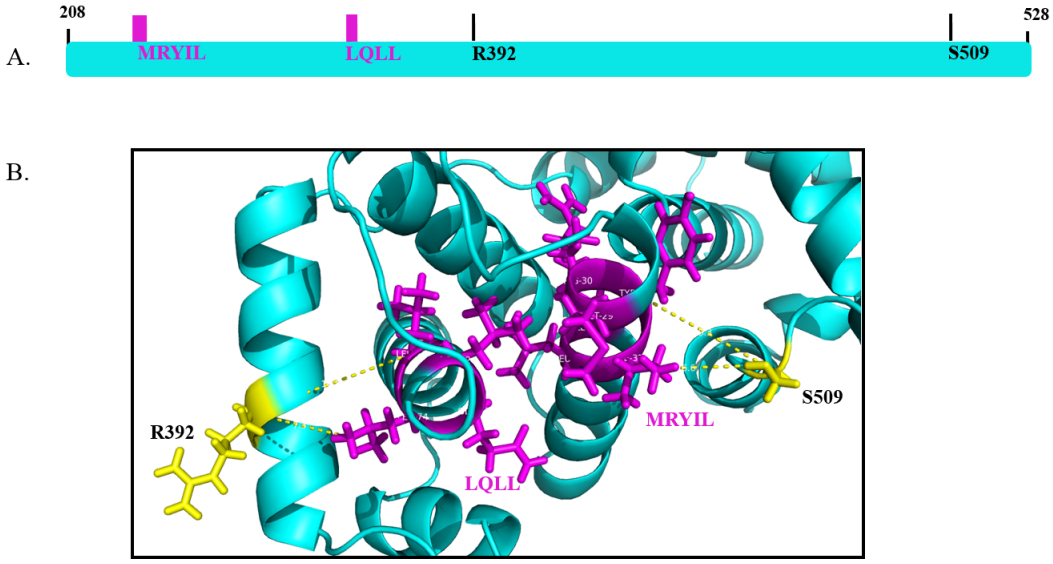


FIGURE 5.1: **The relative 2D and 3D distance between the hydrophobic patch and position of ALS mutations p.R392T and p.S509P.(A.)** A schematic of the amino acid sequence of the cyclin box domain of cyclin F. The hydrophobic patch residues are shown in magenta. The positions of the mutations are shown in black. The schematic is to scale. **(B.)** The 3D representation of the cyclin box, focusing particularly on the hydrophobic patch (shown in magenta), and positions 392 and 509 (shown in yellow), which are the locations of the ALS-linked mutations.

that even if changes to an amino acid are remote from the binding or active site, it can still have a dramatic effect on the thermodynamic stability or activity of a protein (**Freeman et al.**, 2011).

The p.S509P mutation is of particular interest as it occurs in an ALS family and involves a change from a polar serine to a non-polar proline residue (Figure 5.2A). Amino acids with polar side-chains, like serine, form H bonds with other polar side-chains. These H-bonds have been shown to be essential for stabilising the 3D structure of proteins so loss of serine could alter the stability of the 3D structure. Conversely, proline has a cyclic side-chain which gives it a conformational rigidity that could affect the secondary structure of its neighbouring proteins. Modelling of this protein in I-TASSER and PyMol indicate this substitution could confer a change in structure of the tyrosine in the neighbouring hydrophobic patch (Figure 4.7D). Mutational analysis using prediction software programs could not conclusively determine if either this mutation or p.R392T (Figures 5.2B & 4.1) were pathogenic;

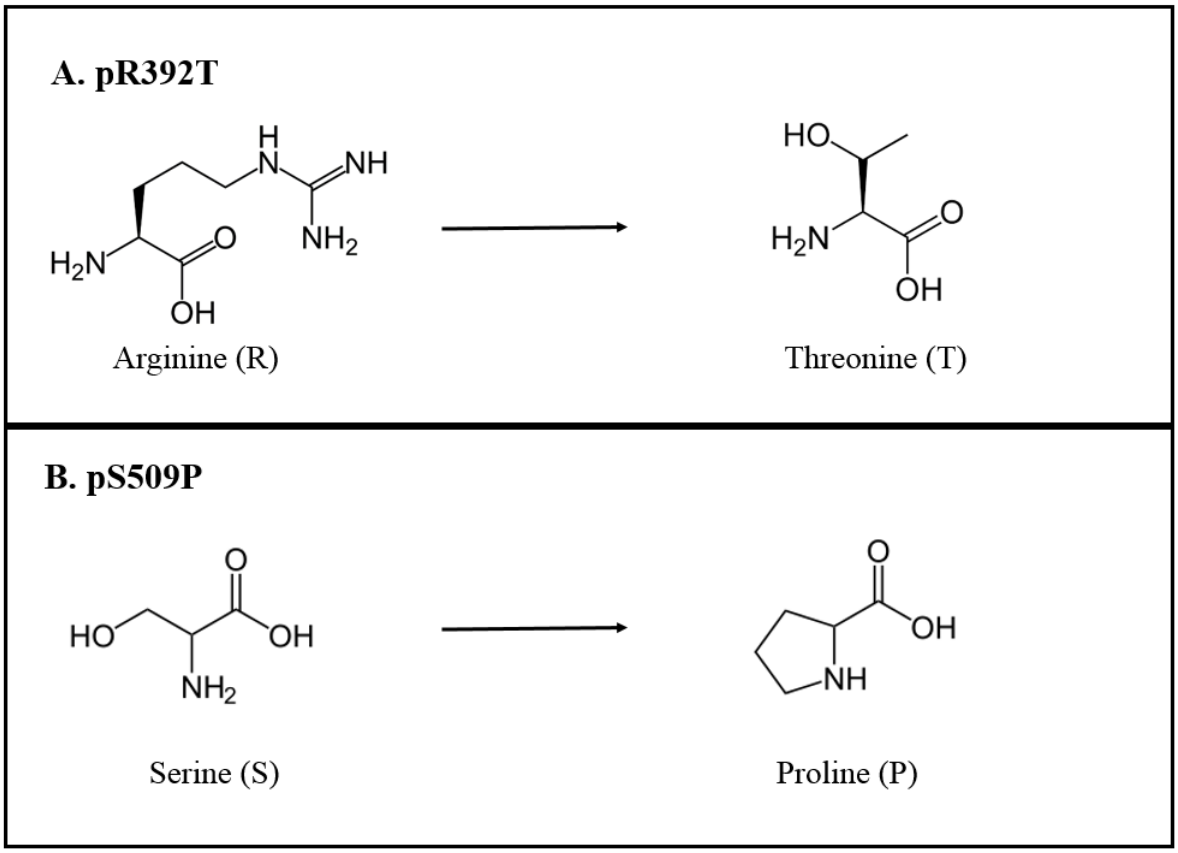


FIGURE 5.2: **Mutations found in the cyclin domain of Cyclin F.(A.)** At residue 392, a polar arginine is replaced with another polar threonine. **(B.)** At residue 509, polar serine is replaced with a chiral proline residue.

however, discrepancies between prediction programs were not unexpected.

Most prediction programs use alignments of the query sequence to homologous sequences to identify whether mutations occur at evolutionarily-conserved sites (**Flanagan et al.**, 2010; **Schwarz et al.**, 2014). The more highly conserved the site of the mutation, the more likely a modification at that site is to be deleterious (**Flanagan et al.**, 2010). Polyphen2 also uses annotation data obtained from UniProt to determine whether mutations occur in functionally active regions where they are more likely to impede protein activity (**Adzhubei et al.**, 2010). However, comparative analysis of the two most popular of these programs, SIFT and PolyPhen2, have shown only a 40% overlap in the predictions (**Castellana and Mazza**, 2013). Both programs are also better at predicting loss-of-function than gain-of-function mutations

(**Flanagan et al.**, 2010), and are conservative when estimating pathogenicity of amino acid substitutions (**Castellana and Mazza**, 2013). Thus a prediction that a substitution is benign is not necessarily correct. This is especially true of dominant diseases, which are more likely to be caused by gain-of-function mutations. Taking these limitations into account, it has been recommended that a combination of several programs be used in conjunction (**Castellana and Mazza**, 2013). It is also important to remember that although conservative mutations are typically deemed to have a low likelihood of being deleterious, this is not always the case (**Jonson and Petersen**, 2001), and multiple neutral mutations can have a combinatory deleterious effect (**Liu et al.**, 2015). Therefore, although these programs are useful for prioritising potentially pathogenic gene variants, the results should not be taken as definitive. Only experimental validation can determine whether mutations are likely to have a pathogenic effect. Therefore, in the future, co-localisation studies should be used to determine if p.S509P and p.R392T mutations affect binding to a known cyclin F substrate, RRM2.

Unfortunately a full length model of cyclin F could not be obtained using SWISS-MODEL or Aquaria, as outside the cyclin box and F-box domains, the cyclin F sequence diverges considerably from all known templates. Therefore, the full length sequence was uploaded to I-TASSER, which uses a composite approach of threading and *ab initio* techniques to build models. Such composite approaches have been deemed the best for protein structure prediction by recent Critical Assessment of Structure Prediction (CASP) experiments, and I-TASSER is consistently ranked amongst the best prediction programs (**Moult et al.**, 2014; **Zhang et al.**, 2015). However, I-TASSER has limitations. It was designed for the analysis of single domain proteins of $<$ 1500bp (**Roy et al.**, 2010), and as such, the 3D model of the full-length cyclin F protein generated by I-TASSER had a very low confidence score (C-value = -2.7) (Figure 4.3). It has been suggested that each domain of multi-domain proteins be submitted separately for analysis (**Roy et al.**, 2010). These separate models can then be stitched together at a later date using docking software, with intervening regions modelled by *ab initio* modelling (**Roy et al.**, 2010). Therefore, the F-box, cyclin box and PEST domains were submitted to I-TASSER again for separate analyses.

The F-box domain of cyclin F showed the highest similarity to its binding partner Skp1 as has been suggested previously (Figure 4.4) (**Kipreos and Pagano**, 2000). Despite the evolutionary conservation of the F-box motif, the F-box protein

sequences diverge considerably ((**Cenciarelli et al.**, 1999)). Therefore, although a reasonable confidence level was found for the F-box 3D structural model (C-score =0.25; TM=score of 0.75 ± 0.11), it was not as high as for the cyclin box determined in I-TASSER (C-score = 1.47; TM-score = 0.92± 0.06). A confident 3D model could not be obtained for the PEST region in I-TASSER (C-score = -3.49; TM-score = 0.29 ±0.09). This was because no confident homologue could be found for the PEST domain despite the existence of PEST sequences in many other proteins (**Meyer et al.**, 2011). This is not surprising as PEST sequences are by their very nature, disordered sequences (**Meyer et al.**, 2011). Without a confident alignment to a template sequence, *ab initio* modelling must be used to build the 3D model. However, the PEST region is ~ 200 amino acids long, larger than the 120 amino acid maximum length optimal for *ab initio* modelling (**Zhang et al.**, 2015). Without a confident PEST model, it was therefore impossible to obtain a complete cyclin F model with high confidence.

5.3 Generation of a deletion construct series of *CCNF* for the study of ALS pathogenesis

Over the last 20 years, several research groups have used deletion cloning to elucidate the structure and function of cyclin F domains. These studies have typically employed deletion clones generated by truncating the 5′ and 3′ ends of *CCNF* (**Fung et al.**, 2002; **Bai et al.**, 1994; **D'Angiolella et al.**, 2010, 2012; **Klein et al.**, 2015), or by deletion of very small internal segments that code for the NLS1 or NLS2 (**Kong et al.**, 2000) (Figure 1.4). The current study is unique in that precise deletions have been made in *CCNF* to generate cyclin F constructs with whole domains removed, leaving the rest of the protein intact. This should enable the more accurate characterisation of cyclin F protein domain function.

The method used to create the deletion clones was *i*PCR mutagenesis. One potential pitfall of this technique is that even high-fidelity, proof-reading polymerases can still lead to unwanted mutations in critical areas of the vector backbone. This issue can be overcome by cloning the original cDNA into a small entry vector and sub cloning into a destination vector after mutatgenesis (**Zawaira et al.**, 2012). In the current study, *CCNF* cDNA was first inserted into the pGEM-T vector, which is only 3kb in size, prior to *i*PCR mutagenesis. The *CCNF* cDNA fragment was PCR amplified with primers designed to insert *XhoI* and *EcoRI* restriction

sites at the 5′ and 3′ ends respectively. This facilitated the later sub-cloning, after mutagenesis, of the *CCNF* insert into the pmCherry-C1 expression vector for use in *in vitro* studies. A high-fidelity DNA polymerase was used to generate the *CCNF* gene fragment because it reduces the number of unwanted mutations. However, it also generates blunt ends on synthesised DNA, which complicated the ligation into pGEM-T. pGEM-T has thymine overhangs to allow the efficient ligation of inserts via TA cloning. Thus, it was necessary to add adenine tails to the *CCNF* PCR product prior to ligation into the pGEM-T vector (Section 3.3.5). *i*PCR was then carried out on the pGEM-T_*CCNF*_WT construct, resulting in the successful generation of six pGEM-T_*CCNF* deletion constructs as shown in Table 4.5.

Prior to the *in vitro* studies, the *CCNF* deletion fragments were sub-cloned into the pmCherry-C1 vector using restriction cloning. A validation PCR demonstrated that sub-cloning into the pmCherry-C1 vector was successful for all deletion fragments except for ΔNLS2 (Figure 4.16). For this construct, an additional 16 transformed colonies underwent repeated PCR validation to determine if the original PCR had failed. *CCNF* was not amplified from any colony (data not shown), suggesting that the *CCNF*_ΔNLS fragment did not successfully ligate into the pmCherry-C1 vector, perhaps due to a failure at the gel extraction stage. Gel extraction may have failed due to inadequate digestion as suggested by the presence of three bands on the gel (Figure 4.14).

Rather than repeat the sub-cloning protocol again for pGEM-T_*CCNF*_ΔNLS, it was decided that the Δ292-766 construct can act as a surrogate for the ΔNLS2 as it also lacks the NLS2 region. Localisation studies confirm predominantly cytoplasmic expression of this construct (see Figures 4.18, 4.19 & 4.20).

5.4 Deletion of the PEST and cyclin domains lead to protein aggregation in the nucleus

Previous *in vitro* studies utilising cyclin F deletion clones were carried out in non-neuronal cell lines, including COS-1, U-2OS, HeLa and HEK293 cells (see Tables 1.4, 1.5 & 1.6). For this project, two different neuronal cell lines (the NSC-34 mouse motor neuron-like cell line and the SH-SY5Y human neuroblastoma cell line) were selected to specifically investigate the function of cyclin F domains in neurons. Analysis of the transfected NSC-34 and SH-SY5Y cells using confocal

microscopy showed that expression of cyclin F was both cytoplasmic and nuclear, similar to previous studies in other cell types (**Bai et al.**, 1994; **Kong et al.**, 2000). Interestingly, the expression of the ΔPEST cyclin F appeared to be predominantly nuclear in NSC-34 and SH-SY5Y cells (Figures 4.18 & 4.20). Cell counting in SH-SY5Y cells confirmed that ΔPEST cyclin F had a significantly higher level of nuclear localisation than wild type cyclin F (Figure 4.19). Nuclear localisation was accompanied by single large or multiple punctate intranuclear aggregates (Figure 4.18B & C). In contrast, Δ292-766 cyclin F was significantly more cytoplasmic than wild type cyclin F, confirming the feasibility of using NLS2 knockouts as a reference for cytoplasmic localisation in sub-cellular localisation studies (Figures 4.18, 4.19 & 4.20). This result was expected since Δ292-766 cyclin F has no NLS2 region, which is partially responsible for localising cyclin F to the nucleus (**Kong et al.**, 2000).

Almost half of the ΔPEST cyclin F-transfected SH-SY5Y cells had intranuclear aggregates, which was significantly greater than seen in cells transfected with wild-type cyclin F ($p<0.0001$) (Figures 4.18, 4.19 & 4.20). Previous studies utilising PEST deletion mutants do not mention aggregate formation in non-neuronal cell lines (**Bai et al.**, 1994; **Fung et al.**, 2002; **D'Angiolella et al.**, 2010, 2012) (Tables 1.4, 1.5 & 1.6), so aggregate formation may be neuron-specific. The PEST sequence is a common feature of short-lived proteins (**Bai et al.**, 1994; **Kong et al.**, 2000). Phosphorylation of the PEST sequence marks the protein for degradation, typically via ubiquitin-mediated or calpain proteolysis (**Martinez et al.**, 2003; **Fecto et al.**, 2011). Therefore, the accumulation of nuclear aggregates in ΔPEST cyclin F-transfected cells could be due to abnormally increased protein stability. However, the degradation of cyclin F is ubiquitin and calpain-independent (**D'Angiolella et al.**, 2013). While the mechanisms of cyclin F degradation are still to be fully determined, several studies have confirmed that PEST is required for its degradation (see Table 1.4) (**D'Angiolella et al.**, 2013). Increased stabilisation of cyclin F would presumably inhibit its proper turnover, which could result in the accumulation of cyclin F into pathogenic aggregates.

Interestingly, mutations have previously been found in the PEST region of p62 (sequestosome), another protein associated with familial ALS (**Fecto et al.**, 2011). p62, which is encoded by the *SQSTM1* gene, binds to ubiquitinated proteins and directs them to the UPS or autophagy pathways (**Fecto et al.**, 2011). Round, ubiquitinated protein aggregates were found in the spinal cords of patients carrying *SQSTM1* mutations, and immunostaining indicated that they were positive for both p62 and TDP-43 (**Pankiv et al.**, 2007). p62 has also been shown to co-localise

with ubiquitinated protein aggregates in non-SQSTM1 ALS individuals, including non-SOD1 related FALS, and SALS cases (**Keller et al.**, 2012; **Fecto et al.**, 2011; **Gal et al.**, 2007; **Deng et al.**, 2010). Loss of phosphorylation at either one of two serine residues in p62, p.S507P within the second PEST sequence or p.S318P between the two PEST sequences, is thought to be the cause of p62 stabilisation and accumulation into aggregates (**Teyssou et al.**, 2013). Interestingly, one of the mutations found in the PEST region of cyclin F, p.S621G, is also predicted to result in a loss of a phosphorylation site (Table 4.1). For future studies, it will be important to examine if a p.S621G mutant construct also leads to the formation of nuclear aggregates *in vitro*.

A loss of phosphorylation could also be the reason that ΔPEST cyclin F is trapped in the nucleus. Phosphorylation is a common post-translational modification that influences the sub-cellular localisation of some proteins (**Hung and Link**, 2011). It is often required for the nuclear export of proteins. For example, the loss of one phosphorylation site is sufficient to impede translocation of cyclin D1 from the nucleus to the cytoplasm (**Alt et al.**, 2000). If cyclin F also relies upon phosphorylation to be exported, loss of PEST could trap cyclin F in the nucleus.

Mis-localisation of proteins and subsequent aggregate formation are frequently implicated in neurodegeneration: TDP-43 (**Winton et al.**, 2008), FUS (**Schwartz et al.**, 2014) and ubiquilin 2 (**Deng et al.**, 2011) are relocated to the cytoplasm in the motor neurons of ALS patients, and tau is relocated to the dendritic spines in AD patients (**Hoover et al.**, 2010). There is some speculation that protein mis-localisation could be the cause of pathogenesis rather than the protein aggregation that ensues from it (**Liu et al.**, 2013). Given that cyclin F shuttles between the nucleus and the cytoplasm during the cell cycle (**D'Angiolella et al.**, 2013), it is likely that mis-localisation disturbs normal cyclin F function. Sequestration of cyclin F into neuronal aggregates would restrict the amount of protein available for cytoplasmic activity. During the G2 phase, cyclin F not only associates with the RRM2 protein in the nucleus (**D'Angiolella et al.**, 2012), but also translocates from the nucleus to the cytoplasm where it associates with the centrosomal duplication protein, CP110 (**D'Angiolella et al.**, 2010). It mediates the proteolytic destruction of both proteins, a requirement for the maintenance of chromosomal and genomic stability. **D'Angiolella et al.** (2013) have demonstrated that NLS2 knockouts are unable to bind RRM2 due to cyclin F redistribution from the nucleus to the cytoplasm. At present, it is difficult to predict the effect of the loss of cytoplasmic cyclin F in terminally differentiated motor neurons since all cytoplasmic substrates identified to date are involved in the cell

cycle (**D'Angiolella et al.**, 2013). It is possible that other unidentified substrates of cyclin F, unrelated to the cell cycle, associate with cyclin F in the cytoplasm of neurons.

A pressing question is whether ALS-linked PEST mutations in cyclin F mimic the Δ PEST cyclin F phenotype. Two ALS-linked mutations have been identified in the PEST domain of cyclin F; one in multiple affected members of an ALS family (p.S621G) and one in a sporadic ALS case (p.E624K) (unpublished data). The p.S621G mutation substitutes a polar serine residue with a hydrophobic glycine (chemical structures are shown in Figure 5.3). This missense mutation is predicted to be deleterious by two prediction programs (Table 4.1). Glycine is the smallest amino acid, lacks a side-chain, and provides flexibility in protein loops. A substitution of glycine for serine could presumably result in a change of 3D shape. Two out of the three prediction programs predicted p.S621G substitution to be deleterious. The p.E624K mutation substitutes a basic lysine residue for an acidic glutamic acid. All three prediction programs predicted this mutation to be deleterious although there is no loss of phosphorylation (Table 4.1). Without a confident 3D model of PEST, it is difficult to predict the effect of either mutation on the structure and function of cyclin F. Therefore, in the future, localisation assays will be repeated with pmCherry-C1_*CCNF*_S621G and pmCherrry-C1_*CCNF*_E624K substitution constructs to see if they result in a similar phenotype to ΔPEST cyclin F.

NSC-34 cells transfected with other cyclin box cyclin F deletion mutants also displayed aggregates, although less obvious than for Δ PEST cyclin F. ΔCyclinC cyclin F had some single large nuclear aggregates, and ΔCyclinN cyclin F and ΔCyclinBox cyclin F had multiple, punctate cytoplasmic aggregates (Figures 4.20). All three cyclin box cyclin F deletion mutants could still bind to the Skp1 protein through the F-box domain (**D'Angiolella et al.**, 2012). This could result in the formation of non-functional SCF complexes, which could lead to toxic protein aggregation. The location of the hydrophobic patch within the cyclin N segment of the cyclin box (**D'Angiolella et al.**, 2012) may be responsible for the differing phenotypes observed among the three Δcyclin constructs. Even with cyclin C deleted, it is plausible that substrates could still bind through the cyclin N region of ΔCyclinC cyclin F although it may lose optimal function. This, in turn, could alter its behaviour, including localisation and the formation of larger aggregates. aggregation propensity. In contrast, the cyclin N and cyclin box deletion mutants would have lost the ability to bind substrates. In the future, NSC-34 cell counts will need to be conducted to determine if aggregate formation is significantly greater in these deletion constructs to

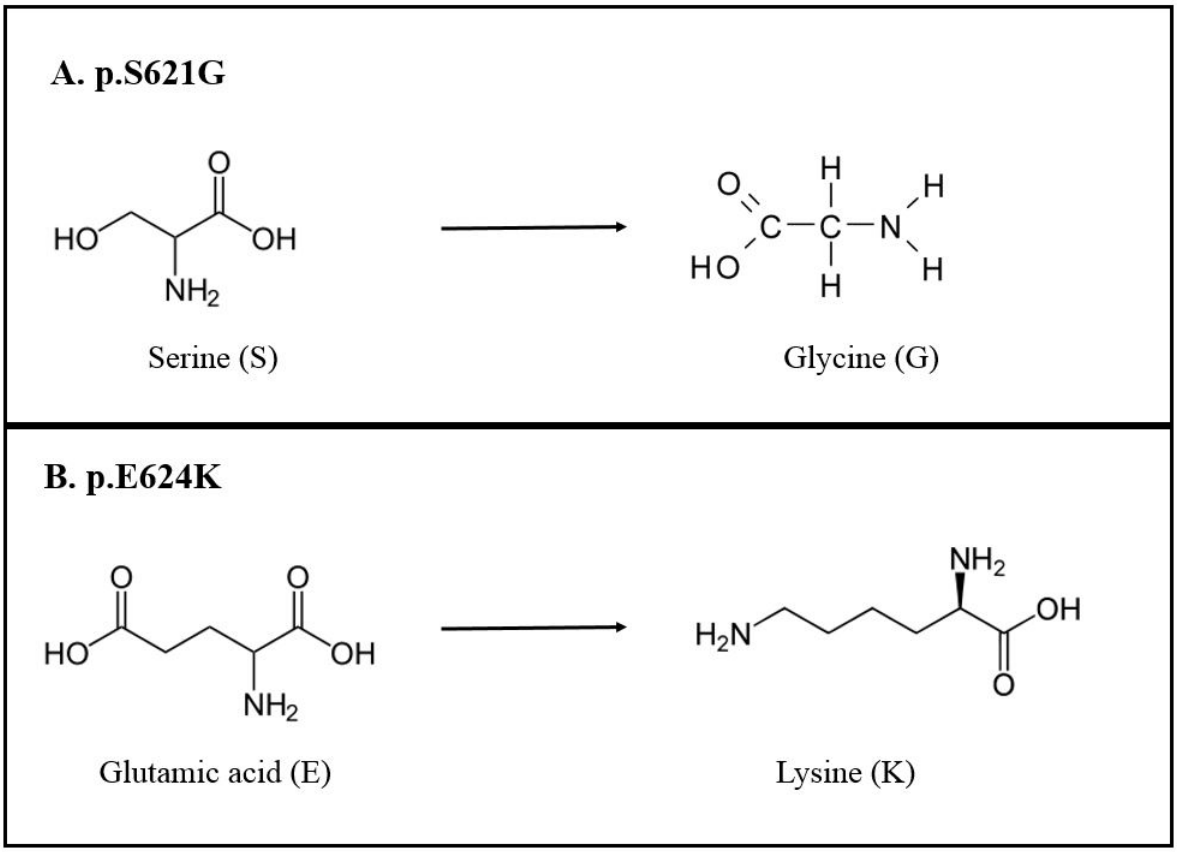


FIGURE 5.3: **Mutations found in the PEST domain of Cyclin F.(A.)** At residue 621, a serine residue is replaced with a glycine residue. **(B.)** At residue 624, glutamic acid residue is replaced with a lysine residue.

wild type cyclin F.

Aggregates were not observed to the same extent in SH-SY5Y cells, compared to NSC-34 cells, when transfected with the cyclin box deletion constructs, suggesting that aggregates could be specific to cells with motor neuron-like properties. NSC-34 cells are hybrids of neuroblastoma and mouse spinal cord motor neurons from enriched primary cultures (**Cashman et al.**, 1992). They have been shown to display characteristics of motor neurons, including acetylcholine synthesis and storage; extension of processes; support of action potentials and expression of neurofilaments (**Tovar-Y-Romo et al.**, 2009). In comparison, the SH-SY5Y cell line is derived from human neuroblastoma. It has a similar biochemistry to immature catecholaminergic neurons, lacking mature neuronal markers. Both NSC-34 and SH-SY5Y cell lines are commonly used in ALS studies. Other studies utilising TDP-43 mutants in SH-SY5Y

and NSC-34 cells have also found differences in expression patterns in these different cell lines (**Wu et al.**, 2013b). For example, there were higher apoptotic death rates in NSC-34 cells compared to SH-SY5Y cells after transfection with mutant TDP-43 (**Wu et al.**, 2013b). As such, the cell toxicity studies were conducted using NSC-34 cells as they more closely resemble motor neuron behaviour (**Cashman et al.**, 1992).

Protein aggregates are a hallmark feature of ALS neuropathology. Whether aggregates are the cause or consequence of neurodegenerative diseases is still debated (**Al-Chalabi et al.**, 2013). The recent discovery of FALS-linked mutations in genes involved in protein clearance pathways could indicate that the aggregates themselves are pathogenic (**Robberecht and Philips**, 2013; **Renton et al.**, 2014). The presence of UPS and autophagy proteins in neuronal cytoplasmic inclusions further supports their causal role (**Bendotti et al.**, 2012). Functioning protein degradation pathways are essential for the proper turnover of abnormal and misfolded proteins, including those proteins implicated in ALS neuropathology. In fact, studies have demonstrated that obstruction of the UPS leads to the accumulation of several ALS disease proteins, including SOD1, TDP-43 and optineurin, in neuronal protein aggregates (**Bendotti et al.**, 2012). However, recent experiments on mutant TDP-43-transfected HeLa cells have demonstrated that reducing protein aggregation does not mitigate cell toxicity, suggesting that protein aggregation may be part of the cell death process rather than the cause of cell death (**Liu et al.**, 2013).

The precise mechanism by which mutations in UPS genes disrupt protein clearance and contribute to ALS pathogenesis is still unknown. One hypothesis is that mutations within UPS genes could impede the delivery of aberrant proteins to the proteasome for destruction, thereby leading to the build-up of protein aggregates (**Ling et al.**, 2013). Several ALS-linked proteins are known to be involved in the transport of ubiquitin-tagged proteins to the proteasome, including ubiquilin 2, p62 and VCP. Mutations have been found in the ubiquitin binding regions of each of these proteins, which could presumably impair ubiquitin-labelled substrate binding (**Deng et al.**, 2011; **Johnson et al.**, 2010; **Williams et al.**, 2012; **Fecto et al.**, 2011; **Teyssou et al.**, 2013). Other mutations in these genes are found in regions essential for proper protein folding, which could lead to general protein dysfunction due to protein misfolding (**Deng et al.**, 2011; **Johnson et al.**, 2010; **Williams et al.**, 2012; **Fecto et al.**, 2011; **Teyssou et al.**, 2013).

5.5 Loss of PEST sequence and cyclin box leads to increased toxicity

Deletion of the PEST sequence or the entire cyclin box domain led to significantly higher rates of cell death than the wild type as assessed by the SYTOX blue flow cytometry assay (Figure 4.22). This was expected as both ΔPEST and ΔCyclinBox cyclin F showed evidence of aggregates (Figure 4.20B,C & F), which are known to have toxic effects on cells (**Wu et al.**, 2013b). Cells expressing ΔCyclinN cyclin F also had small cytoplasmic aggregates, and the toxicity assay approached significance (p=0.05151) (Figure 4.20D). In contrast, lower rates of toxicity were found for cells transfected with Δ292-766 cyclin F compared to WT cyclin F (Figure 4.22) likely due to insufficient protein expression. The transcript derived from this construct is only 1000bp long and the resultant protein lacks both the cyclin box and PEST domains (Figure 4.9). Δ292-766 cyclin F had lower rates of expression as visualised by fluorescent/confocal microscopy, and had significantly lower mCherry expression as assessed by flow cytometry (Figures 4.20G & 4.23).

Cells transfected with the ΔPEST cyclin F had a significantly higher proportion of mCherry expression than all other constructs (Figure 4.23). This could reflect increased protein stability, further supporting the role that the PEST domain plays in the destabilisation of cyclin F although it is important to point out that increased mCherry expression could also reflect increased transfection efficiency. In addition, ΔCyclinN, ΔCyclinC and ΔCyclinBox cyclin F had significantly higher mCherry expression than wild type cyclin F (Figure 4.23). However, only ΔCyclin Box cyclin F had higher toxicity, suggesting that higher expression levels were not the direct cause of toxicity (Figure 4.22). Data from experiments previously conducted in our laboratory showed increased toxicity in the pmCherry-C1 p.S621G mutant compared to wild-type cyclin F in a similar SYTOX blue flow cytometry assay (unpublished data). Repeating this assay with both the p.S621G and p.E624K cyclin F mutants could confirm whether these missense mutations induce more significant cell toxicity than wild type cyclin F. Future studies will assess whether stabilisation of cyclin F due to loss of PEST leads to cell toxicity by studying the half-life of all cyclin F constructs following protein synthesis inhibition.

5.6 Limitations of using cell culture to assess cyclin F function

There are several limitations with *in vitro* studies that could affect the reliability of the results and their relevance to ALS. Firstly, all immortalised cell lines, including NSC-34 and SH-SY5Y cells are cycling cells whereas motor neurons are terminally differentiated and therefore no longer capable of cell division (**Currais et al.**, 2009). Studying the effect of a cell-cycle protein in mitotic cells may not accurately reflect its role in post-mitotic motor neurons. However, cyclin F acts not only as a cell cycle protein, but also as an E3 ubiquitin ligase (**D'Angiolella et al.**, 2013). E3 ligases have been implicated in other neurodegenerative diseases, and are expressed in non-mitotic cells (**Staropoli**, 2008). In addition, cell-cycle re-entry has been demonstrated in some neurodegenerative diseases, especially after excitotoxicity, oxidative stress, ischemia and DNA damage (**Currais et al.**, 2009; **Folch et al.**, 2012). Re-entry into the cell cycle could lead to the apoptosis that is commonly found in neurodegenerative diseases (**Folch et al.**, 2012). Some alternatives to immortalised cell lines include the use of enriched primary motor neurons (PMNs) (**Wu et al.**, 2013b) or chemically-differentiated NSC-34s (**Maier et al.**, 2013). For follow-up studies, it would be ideal to express species-specific *CCNF* deletion constructs in *in vivo* models, such as zebrafish or *C.elegans*, which have the potential to more accurately reproduce a motor neuron disease phenotype (**Tovar-Y-Romo et al.**, 2009).

The second limitation is the use of animal-derived cell lines to model human disease. The NSC-34 cell line is derived from mouse cells, and it is possible that species differences could affect the phenotype seen in cell culture. However, without an effective human motor neuron cell line, NSC-34s are currently the best available option for studying the effect of ALS phenotype in *in vitro* studies (**Tovar-Y-Romo et al.**, 2009). To determine if a similar phenotype could be seen in human cells, the human-derived SH-SY5Y cell line was also used. SH-SY5Y cells do not display motor neuron features, and they also lack the expression of mature neuronal markers (**Krishna et al.**, 2014). However, a recent assessment of the cell line using a systems genomic approach indicate that it is a suitable model to assess perturbations caused by mutations in ALS-linked genes (**Krishna et al.**, 2014). Both NSC-34 and SH-SY5Y cell lines have been extensively used in published ALS studies. The similarity observed in the sub-cellular localisation of cyclin F deletion constructs supports their use in the current study.

The third limitation is that wild type genes can cause toxicity when introduced into cell lines due to overexpression of the protein (**Wu et al.**, 2013b). Genome engineering has been gaining popularity as a strategy to overcome this issue. Over the last decade, more precise genome editing techniques that use engineered endonucleases have been developed. These include protein-based nucleases, such as transcription activator-like effector nucleases (TALENs) and Zinc finger nucleases (ZFNs), and more recently, a modified version of the RNA-based clustered regularly interspaced short palindromic repeats and CRISPR-associated (CRISPR/Cas9) (All reviewed in **Gaj et al.** (2013)). While all three endonucleases have been used for genome engineering in cell lines and animal models, specific deletion within genes remains challenging. Furthermore, a serious concern with endonucleases is the incidence of off-target genome editing, which ranges from low to considerable depending on experimental conditions (**Gasiunas and Siksnys**, 2013; **Gupta and Musunuru**, 2014). While CRISPR/Cas9 is increasingly being adopted for mutagenesis studies, much of its use remains to be optimised and validated, and so was not used for the current study.

5.7 Summary and conclusions

Our laboratory identified four ALS-linked mutations within two major domains of cyclin F, two in the cyclin box and two in the PEST sequence. 3D modelling of the mutations within the cyclin box show that they may affect substrate binding, while deletion cloning studies show that mutations within PEST could affect protein stability and cause mis-localisation, leading to protein aggregation. This could lead to the enhanced cytotoxicity that was found in ΔPEST cyclin F-transfected cells.

In the current project, the functions of cyclin F domains were characterised using two different approaches. The first approach generated 3D models of cyclin F using homology modelling in multiple bioinformatic programs. A high-resolution model of the cyclin box was obtained with a high confidence value, indicating it is likely to have the correct topology (**Zhang et al.**, 2015). Exploration of this 3D model indicates that the two ALS-linked mutations (p.S509P and p.R392T) lie in close proximity to the hydrophobic patch in 3D space. Modelling of the two mutations highlight p.S509P in particular as potentially affecting the orientation of a residue in the hydrophobic patch. This was further supported by mutation prediction analysis, which universally deemed the p.S509P mutation as likely to be deleterious.

The second approach to characterise cyclin F domains involved the generation of a series of *CCNF* deletion clones, which were expressed in neuronal cell lines to assess the effect of domain deletion on sub-cellular localisation and cell toxicity. Eleven plasmid expression constructs were generated in total: six pGEM-T_*CCNF* and five pmCherry-C1_*CCNF* deletion constructs (see Table 4.5). *i*PCR mutagenesis proved to be an effective method for generating these deletion constructs, a process that could be extended to other ALS gene candidates in the future. The current series of cyclin F deletion constructs are now available for use in additional *in vitro* studies to further characterise cyclin F domain function. The most intriguing finding from the *in vitro* studies presented here was the presence of large intranuclear aggregates in almost half of the NSC-34 and SH-SY5Y cells expressing Δ-PEST cyclin F. Together with the observed increase in cell toxicity, this supports the supposition that loss of PEST leads to protein aggregation and cytotoxicity.

5.8 Future directions

Future studies should characterise the ΔPEST cyclin F nuclear aggregates using colocalisation studies. It will be particularly important to co-stain cells transfected with the cyclin F constructs with well-known ALS proteins, including TDP-43, ubiquitin, ubiquilin2 and p62 (**Blokhuis et al.**, 2013). If the cyclin F aggregates co-localise with any of these proteins, it would support the role of cyclin F in the induction of protein aggregation leading to ALS (**Wu et al.**, 2012). In addition, future studies should assess whether cyclin F with deletions alters UPS function by conducting a GFPu assay. GFPu is a UPS reporter molecule consisting of a 16 amino acid degron (CL1) derived from yeast fused to a C-terminal green fluorescent protein (GFP) (**Dantuma et al.**, 2000). It allows the assessment of protein degradation by the ubiquitin-proteasome-mediated proteolysis. If the ubiquitin proteasome pathway is inhibited, then the GFP molecule will not be degraded. The resulting GFP fluorescence can be measured using flow cytometry.

Abbreviations

AD	Alzheimer's disease
ALS	amyotrophic lateral sclerosis
ALSoD	amyotrophic lateral sclerosis online genetic database
BLAST	basic local alignment search tool
B-Myb	myb-related protein B
bp	base pair
C9ORF72	chromosome 9 open reading frame 72
CCNF	cyclin F gene
C33	a stable cell line derived from cervical cancer cells
cDNA	coding DNA
CG	cytosine and guanine
Co-IP	co-immunoprecipitation
COS-1	a fibroblast-like cell line derived from CV-1 simian cells.
CP110	centriolar coiled-coil protein of 110 kDa
Cul1	cullin 1 protein
DAPI	4',6-diamidino-2-phenylindole
dATP	deoxyadenine triphosphate
dbSNP	single nucleotide polymorphism database

DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dsDNA	double stranded DNA
dTTP	deoxytyrosine triphosphate
E1	E1 ubiquitin activating enzyme
E2	E2 ubiquitin conjugating enzyme
E3	E3 ubiquitin ligase
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
FALS	familial ALS
FSC	forward scatter channel
FTD	frontotemporal dementia
FUS	fused in sarcoma
G1 phase	gap phase 1 of the cell cycle
G2 phase	gap phase 2 of the cell cycle
GLT1	glial glutamate transporter 1
HEK293	a cell line derived from human embryonic kidney cells
HeLa	an immortal cell line derived from cervical cancer cells taken from Henrietta Lacks
hnRNP	heterogeneousnuclear ribonucleoprotein
IBM	inclusion body myopathy

iPCR	inverse polymerase chain reaction
LB	luria broth
M phase	mitotic phase of the cell cycle
MATR3	matrin 3
MND	motor neuron disease
mRNA	messenger RNA
NCBI	national center for biotechnology information
NLS	nuclear localisation signal
NuSAP1	nucleolar and spindle-associated protein 1
OMIM	online mendelian inheritance in man
OPTN	optineurin
p62	nucleoporin p62
PBS	phosphate buffered solution
PCR	polymerase chain reaction
PD	Parkinson's disease
Rbx1	Ring box protein 1
RRM2	ribonucleotide reductase M2
RNA	ribonucleic acid
S phase	synthesis phase of the cell cycle
SALS	sporadic ALS
SCF	Skp1-Cul1-F-box complex
Skp1	S-phase kinase-associated protein 1
SNP	single nucleotide polymorphism
SOD1	superoxide dismutase 1

SQSTM1	sequestosome 1
SSC	side-scatter channel
TARDBP	TAR DNA binding protein gene
TBE	tris-borate EDTA
TDP-43	TAR DNA binding protein of 43 kDa
TE	tris EDTA
TSAP	thermosensitive alkaline phosphatase
U2-OS	a cell line derived from human osteosarcoma cells
UBQN2	ubiquilin 2
UPS	ubiquitin proteasome system
VCP	valosin containing protein

References

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010), A method and server for predicting damaging missense mutations, *Nat. Methods*, 7, 4, 248–249 30, 87

Al-Chalabi, A., Jones, A., Troakes, C., King, A., Al-Sarraj, S., and van den Berg, L. H. (2012), The genetics and neuropathology of amyotrophic lateral sclerosis, *Acta Neuropathol.*, 124, 3, 339–352 1

Al-Chalabi, A., Kwak, S., Mehler, M., Rouleau, G., Siddique, T., Strong, M., et al. (2013), Genetic and epigenetic studies of amyotrophic lateral sclerosis, *Amyotroph Lateral Scler Frontotemporal Degener*, 14 Suppl 1, 44–52 95

Alt, J. R., Cleveland, J. L., Hannink, M., and Diehl, J. A. (2000), Phosphorylation-dependent regulation of cyclin D1 nuclear export and cyclin D1-dependent cellular transformation, *Genes Dev.*, 14, 24, 3102–3114 92

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997), Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, 25, 17, 3389–3402 27

Anderson, S. (1981), Shotgun DNA sequencing using cloned DNase I-generated fragments, *Nucleic Acids Res.*, 9, 13, 3015–3027

Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., et al. (2006), TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis, *Biochem. Biophys. Res. Commun.*, 351, 3, 602–611 1, 2

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006), The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, *Bioinformatics*, 22, 2, 195–201 10, 28, 53

Aronica, E., Baas, F., Iyer, A., ten Asbroek, A. L., Morello, G., and Cavallaro, S. (2015), Molecular classification of amyotrophic lateral sclerosis by unsupervised clustering of gene expression in motor cortex, *Neurobiol. Dis.*, 74, 359–376 8

Ash, P. E., Zhang, Y. J., Roberts, C. M., Saldi, T., Hutter, H., Buratti, E., et al. (2010), Neurotoxic effects of TDP-43 overexpression in C. elegans, *Hum. Mol. Genet.*, 19, 16, 3206–3218 11, 15

Australian Institute of Health and Welfare (2011), Analysis of Australian Institute of Health and Welfare National National Mortality Database, Technical report, Australian Institute of Health and Welfare, Canberra, NSW 1

Bai, C., Richman, R., and Elledge, S. J. (1994), Human cyclin F, *EMBO J.*, 13, 24, 6087–6098 4, 11, 12, 13, 14, 69, 89, 91

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J. W., et al. (1996), SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box, *Cell*, 86, 2, 263–274 8

Bairoch, A. and Apweiler, R. (1996), The SWISS-PROT protein sequence data bank and its new supplement TREMBL, *Nucleic Acids Res.*, 24, 1, 21–25 28

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., et al. (2015), UniProt: a hub for protein information, *Nucleic Acids Res.*, 43, Database issue, D204–212 29

Bendotti, C., Marino, M., Cheroni, C., Fontana, E., Crippa, V., Poletti, A., et al. (2012), Dysfunction of constitutive and inducible ubiquitin-proteasome system in amyotrophic lateral sclerosis: implication for protein aggregation and immune response, *Prog. Neurobiol.*, 97, 2, 101–126 2, 4, 95

Benkert, P., Tosatto, S. C., and Schomburg, D. (2008), QMEAN: A comprehensive scoring function for model quality assessment, *Proteins*, 71, 1, 261–277 28

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000), The Protein Data Bank, *Nucleic Acids Res.*, 28, 1, 235–242 27, 28

Blokhuis, A. M., Groen, E. J., Koppers, M., van den Berg, L. H., and Pasterkamp, R. J. (2013), Protein aggregation in amyotrophic lateral sclerosis, *Acta Neuropathol.*, 125, 6, 777–794 99

Blom, N., Gammeltoft, S., and Brunak, S. (1999), Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, *J. Mol. Biol.*, 294, 5, 1351–1362 10, 30

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2009), Protein structure homology modeling using SWISS-MODEL workspace, *Nat Protoc*, 4, 1, 1–13 28, 53

Brown, N. R., Korolchuk, S., Martin, M. P., Stanley, W. A., Moukhametzianov, R., Noble, M. E., et al. (2015), CDK1 structures reveal conserved and unique features of the essential cell cycle CDK, *Nat Commun*, 6, 6769 10

Caccamo, A., Majumder, S., and Oddo, S. (2012), Cognitive decline typical of frontotemporal lobar degeneration in transgenic mice expressing the 25-kDa C-terminal fragment of TDP-43, *Am. J. Pathol.*, 180, 1, 293–302 11

Carbain, B., Paterson, D. J., Anscombe, E., Campbell, A. J., Cano, C., Echalier, A., et al. (2014), 8-Substituted O(6)-cyclohexylmethylguanine CDK2 inhibitors: using structure-based inhibitor design to optimize an alternative binding mode, *J. Med. Chem.*, 57, 1, 56–70 10

Cashman, N. R., Durham, H. D., Blusztajn, J. K., Oda, K., Tabira, T., Shaw, I. T., et al. (1992), Neuroblastoma x spinal cord (NSC) hybrid cell lines resemble developing motor neurons, *Dev. Dyn.*, 194, 3, 209–221 47, 70, 94, 95

Castellana, S. and Mazza, T. (2013), Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools, *Brief. Bioinformatics*, 14, 4, 448–459 87, 88

Cenciarelli, C., Chiaur, D. S., Guardavaccaro, D., Parks, W., Vidal, M., and Pagano, M. (1999), Identification of a family of human F-box proteins, *Curr. Biol.*, 9, 20, 1177–1179 89

Chesi, A., Staahl, B. T., Jovi?i?, A., Couthouis, J., Fasolino, M., Raphael, A. R., et al. (2013), Exome sequencing to identify de novo mutations in sporadic ALS trios, *Nat. Neurosci.*, 16, 7, 851–855 3

Couthouis, J., Hart, M. P., Erion, R., King, O. D., Diaz, Z., Nakaya, T., et al. (2012), Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis, *Hum. Mol. Genet.*, 21, 13, 2899–2911 3

Couthouis, J., Hart, M. P., Shorter, J., DeJesus-Hernandez, M., Erion, R., Oristano, R., et al. (2011), A yeast functional screen predicts new candidate ALS disease genes, *Proc. Natl. Acad. Sci. U.S.A.*, 108, 52, 20881–20890 3

Currais, A., Hortobagyi, T., and Soriano, S. (2009), The neuronal cell cycle as a mechanism of pathogenesis in Alzheimer's disease, *Aging (Albany NY)*, 1, 4, 363–371 97

D'Angiolella, V., Donato, V., Forrester, F. M., Jeong, Y. T., Pellacani, C., Kudo, Y., et al. (2012), Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair, *Cell*, 149, 5, 1023–1034 5, 9, 11, 12, 14, 16, 85, 89, 91, 92, 93

D'Angiolella, V., Donato, V., Vijayakumar, S., Saraf, A., Florens, L., Washburn, M. P., et al. (2010), SCF(Cyclin F) controls centrosome homeostasis and mitotic fidelity through CP110 degradation, *Nature*, 466, 7302, 138–142 5, 9, 11, 12, 14, 16, 85, 89, 91, 92

D'Angiolella, V., Esencay, M., and Pagano, M. (2013), A cyclin without cyclin-dependent kinases: cyclin F controls genome stability through ubiquitin-mediated proteolysis, *Trends Cell Biol.*, 23, 3, 135–140 4, 85, 91, 92, 93, 97

Dantuma, N. P., Lindsten, K., Glas, R., Jellne, M., and Masucci, M. G. (2000), Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells, *Nat. Biotechnol.*, 18, 5, 538–543 99

DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., et al. (2011), Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS, *Neuron*, 72, 2, 245–256 2, 3

Deng, H. X., Chen, W., Hong, S. T., Boycott, K. M., Gorrie, G. H., Siddique, N., et al. (2011), Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia, *Nature*, 477, 7363, 211–215 2, 3, 92, 95

Deng, H. X., Zhai, H., Bigio, E. H., Yan, J., Fecto, F., Ajroud, K., et al. (2010), FUS-immunoreactive inclusions are a common feature in sporadic and non-SOD1 familial amyotrophic lateral sclerosis, *Ann. Neurol.*, 67, 6, 739–748 92

Edelheit, O., Hanukoglu, A., and Hanukoglu, I. (2009), Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies, *BMC Biotechnol.*, 9, 61 15

Emanuele, M. J., Elia, A. E., Xu, Q., Thoma, C. R., Izhar, L., Leng, Y., et al. (2011), Global identification of modular cullin-RING ligase substrates, *Cell*, 147, 2, 459–474 5, 9

Fecto, F. and Siddique, T. (2011), Making connections: pathology and genetics link amyotrophic lateral sclerosis with frontotemporal lobe dementia, *J. Mol. Neurosci.*, 45, 3, 663–675 2

Fecto, F., Yan, J., Vemula, S. P., Liu, E., Yang, Y., Chen, W., et al. (2011), SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis, *Arch. Neurol.*, 68, 11, 1440–1446 3, 91, 92, 95

Feiguin, F., Godena, V. K., Romano, G., D'Ambrogio, A., Klima, R., and Baralle, F. E. (2009), Depletion of TDP-43 affects Drosophila motoneurons terminal synapsis and locomotive behavior, *FEBS Lett.*, 583, 10, 1586–1592 11

Finn, R. D., Miller, B. L., Clements, J., and Bateman, A. (2014), iPfam: a database of protein family and domain interactions found in the Protein Data Bank, *Nucleic Acids Res.*, 42, Database issue, D364–373 29

Flanagan, S. E., Patch, A. M., and Ellard, S. (2010), Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations, *Genet Test Mol Biomarkers*, 14, 4, 533–537 87, 88

Folch, J., Junyent, F., Verdaguer, E., Auladell, C., Pizarro, J. G., Beas-Zarate, C., et al. (2012), Role of cell cycle re-entry in neurons: a common apoptotic mechanism of neuronal cell death, *Neurotox Res*, 22, 3, 195–207 97

Freeman, A. M., Mole, B. M., Silversmith, R. E., and Bourret, R. B. (2011), Action at a distance: amino acid substitutions that affect binding of the phosphorylated CheY response regulator and catalysis of dephosphorylation can be far from the CheZ phosphatase active site, *J. Bacteriol.*, 193, 18, 4709–4718 86

Fung, T. K., Siu, W. Y., Yam, C. H., Lau, A., and Poon, R. Y. (2002), Cyclin F is degraded during G2-M by mechanisms fundamentally different from other cyclins, *J. Biol. Chem.*, 277, 38, 35140–35149 11, 12, 13, 16, 89, 91

Gaj, T., Gersbach, C. A., and Barbas, C. F. (2013), ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering, *Trends Biotechnol.*, 31, 7, 397–405 98

Gal, J., Strom, A. L., Kilty, R., Zhang, F., and Zhu, H. (2007), p62 accumulates and enhances aggregate formation in model systems of familial amyotrophic lateral sclerosis, *J. Biol. Chem.*, 282, 15, 11068–11077 92

Gasiunas, G. and Siksnys, V. (2013), RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing?, *Trends Microbiol.*, 21, 11, 562–567 98

Gupta, R. M. and Musunuru, K. (2014), Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9, *J. Clin. Invest.*, 124, 10, 4154–4161 98

Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., and Galas, D. J. (1989), A simple method for site-directed mutagenesis using the polymerase chain reaction, *Nucleic Acids Res.*, 17, 16, 6545–6551 15

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989), Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene*, 77, 1, 51–59 15

Hoover, B. R., Reed, M. N., Su, J., Penrod, R. D., Kotilinek, L. A., Grant, M. K., et al. (2010), Tau mislocalization to dendritic spines mediates synaptic dysfunction independently of neurodegeneration, *Neuron*, 68, 6, 1067–1081 92

Hung, M. C. and Link, W. (2011), Protein localization in disease and therapy, *J. Cell. Sci.*, 124, Pt 20, 3381–3392 92

Imai, Y., Matsushima, Y., Sugimura, T., and Terada, M. (1991), A simple and rapid method for generating a deletion by PCR, *Nucleic Acids Res.*, 19, 10, 2785 15

Johnson, J. O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V. M., Trojanowski, J. Q., et al. (2010), Exome sequencing reveals VCP mutations as a cause of familial ALS, *Neuron*, 68, 5, 857–864 2, 3, 95

Johnson, J. O., Pioro, E. P., Boehringer, A., Chia, R., Feit, H., Renton, A. E., et al. (2014), Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis, *Nat. Neurosci.*, 17, 5, 664–666 3

Jonson, P. H. and Petersen, S. B. (2001), A critical view on conservative mutations, *Protein Eng.*, 14, 6, 397–402 88

Keller, B. A., Volkening, K., Droppelmann, C. A., Ang, L. C., Rademakers, R., and Strong, M. J. (2012), Co-aggregation of RNA binding proteins in ALS spinal motor neurons: evidence of a common pathogenic mechanism, *Acta Neuropathol.*, 124, 5, 733–747 92

Kim, Y., Kweon, J., Kim, A., Chon, J. K., Yoo, J. Y., Kim, H. J., et al. (2013), A library of TAL effector nucleases spanning the human genome, *Nat. Biotechnol.*, 31, 3, 251–258 3

Kipreos, E. T. and Pagano, M. (2000), The F-box protein family, *Genome Biol.*, 1, 5, REVIEWS3002 55, 88

Klein, D. K., Hoffmann, S., Ahlskog, J. K., O'Hanlon, K., Quaas, M., Larsen, B. D., et al. (2015), Cyclin F suppresses B-Myb activity to promote cell cycle checkpoint control, *Nat Commun*, 6, 5800 5, 9, 11, 12, 16, 85, 89

Kong, M., Barnes, E. A., Ollendorff, V., and Donoghue, D. J. (2000), Cyclin F regulates the nuclear localization of cyclin B1 through a cyclin-cyclin interaction, *EMBO J.*, 19, 6, 1378–1388 9, 11, 12, 14, 15, 16, 19, 89, 91

Kovalevich, J. and Langford, D. (2013), Considerations for the use of SH-SY5Y neuroblastoma cells in neurobiology, *Methods Mol. Biol.*, 1078, 9–21 47

Krishna, A., Biryukov, M., Trefois, C., Antony, P. M., Hussong, R., Lin, J., et al. (2014), Systems genomics evaluation of the SH-SY5Y neuroblastoma cell line as a model for Parkinson's disease, *BMC Genomics*, 15, 1154 97

Kumar, P., Henikoff, S., and Ng, P. C. (2009), Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm, *Nat Protoc*, 4, 7, 1073–1081 30

Kwiatkowski, T. J., Bosco, D. A., Leclerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., et al. (2009), Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis, *Science*, 323, 5918, 1205–1208 2, 3

Lagier-Tourenne, C., Polymenidou, M., Hutt, K. R., Vu, A. Q., Baughn, M., Huelga, S. C., et al. (2012), Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs, *Nat. Neurosci.*, 15, 11, 1488–1497 2

Lander, G. C., Saibil, H. R., and Nogales, E. (2012), Go hybrid: EM, crystallography, and beyond, *Curr. Opin. Struct. Biol.*, 22, 5, 627–635 8

Ling, S. C., Polymenidou, M., and Cleveland, D. W. (2013), Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis, *Neuron*, 79, 3, 416–438 2, 95

Liu, H. and Naismith, J. H. (2008), An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol, *BMC Biotechnol.*, 8, 91 15

Liu, M., Watson, L. T., and Zhang, L. (2015), Predicting the combined effect of multiple genetic variants, *Hum. Genomics*, 9, 18 88

Liu, R., Yang, G., Nonaka, T., Arai, T., Jia, W., and Cynader, M. S. (2013), Reducing TDP-43 aggregation does not prevent its cytotoxicity, *Acta Neuropathol Commun*, 1, 49 92, 95

Liu-Yesucevitz, L., Bilgutay, A., Zhang, Y. J., Vanderweyde, T., Vanderwyde, T., Citro, A., et al. (2010), Tar DNA binding protein-43 (TDP-43) associates with stress granules: analysis of cultured cells and pathological brain tissue, *PLoS ONE*, 5, 10, e13250 11

Lohmann, E., Coquel, A. S., Honore, A., Gurvit, H., Hanagasi, H., Emre, M., et al. (2015), A new F-box protein 7 gene mutation causing typical Parkinson's disease, *Mov. Disord.*, 30, 8, 1130–1133 8

Maier, O., Bohm, J., Dahm, M., Bruck, S., Beyer, C., and Johann, S. (2013), Differentiated NSC-34 motoneuron-like cells as experimental model for cholinergic neurodegeneration, *Neurochem. Int.*, 62, 8, 1029–1038 97

Martinez, L. O., Agerholm-Larsen, B., Wang, N., Chen, W., and Tall, A. R. (2003), Phosphorylation of a pest sequence in ABCA1 promotes calpain degradation and is reversed by ApoA-I, *J. Biol. Chem.*, 278, 39, 37368–37374 91

Maruyama, H., Morino, H., Ito, H., Izumi, Y., Kato, H., Watanabe, Y., et al. (2010), Mutations of optineurin in amyotrophic lateral sclerosis, *Nature*, 465, 7295, 223–226 2, 3

Meyer, R. D., Srinivasan, S., Singh, A. J., Mahoney, J. E., Gharahassanlou, K. R., and Rahimi, N. (2011), PEST motif serine and tyrosine phosphorylation controls vascular endothelial growth factor receptor 2 stability and downregulation, *Mol. Cell. Biol.*, 31, 10, 2010–2025 56, 89

Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., and Tramontano, A. (2014), Critical assessment of methods of protein structure prediction (CASP)–round x, *Proteins*, 82 Suppl 2, 1–6 88

Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., et al. (2006), Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis, *Science*, 314, 5796, 130–133 1, 2

Nonaka, T., Kametani, F., Arai, T., Akiyama, H., and Hasegawa, M. (2009), Truncation and pathogenic mutations facilitate the formation of intracellular aggregates of TDP-43, *Hum. Mol. Genet.*, 18, 18, 3353–3364 10, 11

O'Donoghue, S. I., Goodsell, D. S., Frangakis, A. S., Jossinet, F., Laskowski, R. A., Nilges, M., et al. (2010), Visualization of macromolecular structures, *Nat. Methods*, 7, 3 Suppl, 42–55 8, 10

O'Donoghue, S. I., Sabir, K. S., Kalemanov, M., Stolte, C., Wellmann, B., Ho, V., et al. (2015), Aquaria: simplifying discovery and insight from protein structures, *Nat. Methods*, 12, 2, 98–99 10, 28, 54

Ou, S. H., Wu, F., Harrich, D., Garcia-Martinez, L. F., and Gaynor, R. B. (1995), Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs, *J. Virol.*, 69, 6, 3584–3596

Pankiv, S., Clausen, T. H., Lamark, T., Brech, A., Bruun, J. A., Outzen, H., et al. (2007), p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy, *J. Biol. Chem.*, 282, 33, 24131–24145 91

Qi, D. and Scholthof, K. B. (2008), A one-step PCR-based method for rapid and efficient site-directed fragment deletion, insertion, and substitution mutagenesis, *J. Virol. Methods*, 149, 1, 85–90 15

Remmert, M., Biegert, A., Hauser, A., and Soding, J. (2012), HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment, *Nat. Methods*, 9, 2, 173–175 10, 28

Renton, A. E., Chio, A., and Traynor, B. J. (2014), State of play in amyotrophic lateral sclerosis genetics, *Nat. Neurosci.*, 17, 1, 17–23 1, 2, 83, 95

Renton, A. E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J. R., et al. (2011), A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD, *Neuron*, 72, 2, 257–268 2, 3

Robberecht, W. and Philips, T. (2013), The changing scene of amyotrophic lateral sclerosis, *Nat. Rev. Neurosci.*, 14, 4, 248–264 1, 2, 4, 8, 83, 95

Rosen, D. R. (1993), Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis, *Nature*, 364, 6435, 362 2, 3

Rosenfeld, J. and Strong, M. J. (2015), Challenges in the Understanding and Treatment of Amyotrophic Lateral Sclerosis/Motor Neuron Disease, *Neurotherapeutics*, 12, 2, 317–325 83

Rost, B. (1999), Twilight zone of protein sequence alignments, *Protein Eng.*, 12, 2, 85–94 84

Roy, A., Kucukural, A., and Zhang, Y. (2010), I-TASSER: a unified platform for automated protein structure and function prediction, *Nat Protoc*, 5, 4, 725–738 10, 28, 29, 55, 88

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012), NIH Image to ImageJ: 25 years of image analysis, *Nat. Methods*, 9, 7, 671–675 49

Schwartz, J. C., Podell, E. R., Han, S. S., Berry, J. D., Eggan, K. C., and Cech, T. R. (2014), FUS is sequestered in nuclear aggregates in ALS patient fibroblasts, *Mol. Biol. Cell*, 25, 17, 2571–2578 92

Schwarz, J. M., Cooper, D. N., Schuelke, M., and Seelow, D. (2014), MutationTaster2: mutation prediction for the deep-sequencing age, *Nat. Methods*, 11, 4, 361–362 30, 87

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011), Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol.*, 7, 539 55, 85

Skaar, J. R., Pagan, J. K., and Pagano, M. (2013), Mechanisms and function of substrate recruitment by F-box proteins, *Nat. Rev. Mol. Cell Biol.*, 14, 6, 369–381 4, 5, 6, 7, 84

Smith, B. N., Ticozzi, N., Fallini, C., Gkazi, A. S., Topp, S., Kenna, K. P., et al. (2014), Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS, *Neuron*, 84, 2, 324–331 3

Sreedharan, J., Blair, I. P., Tripathi, V. B., Hu, X., Vance, C., Rogelj, B., et al. (2008), TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis, *Science*, 319, 5870, 1668–1672 2, 3

Staropoli, J. F. (2008), Tumorigenesis and neurodegeneration: two sides of the same coin?, *Bioessays*, 30, 8, 719–727 8, 97

Sun, S., Ling, S. C., Qiu, J., Albuquerque, C. P., Zhou, Y., Tokunaga, S., et al. (2015), ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP, *Nat Commun*, 6, 6171 15

Takahashi, Y., Fukuda, Y., Yoshimura, J., Toyoda, A., Kurppa, K., Moritoyo, H., et al. (2013), ERBB4 mutations that disrupt the neuregulin-ErbB4 pathway cause amyotrophic lateral sclerosis type 19, *Am. J. Hum. Genet.*, 93, 5, 900–905 3

Takalo, M., Salminen, A., Soininen, H., Hiltunen, M., and Haapasalo, A. (2013), Protein aggregation and degradation mechanisms in neurodegenerative diseases, *Am J Neurodegener Dis*, 2, 1, 1–14 4

Teyssou, E., Takeda, T., Lebon, V., Boillee, S., Doukoure, B., Bataillon, G., et al. (2013), Mutations in SQSTM1 encoding p62 in amyotrophic lateral sclerosis: genetics and neuropathology, *Acta Neuropathol.*, 125, 4, 511–522 92, 95

Tovar-Y-Romo, L. B., Santa-Cruz, L. D., and Tapia, R. (2009), Experimental models for the study of neurodegeneration in amyotrophic lateral sclerosis, *Mol Neurodegener*, 4, 31 94, 97

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K. J., Nishimura, A. L., Sreedharan, J., et al. (2009), Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6, *Science*, 323, 5918, 1208–1211 2, 3

Williams, K. L., Warraich, S. T., Yang, S., Solski, J. A., Fernando, R., Rouleau, G. A., et al. (2012), UBQLN2/ubiquilin 2 mutation and pathology in familial amyotrophic lateral sclerosis, *Neurobiol. Aging*, 33, 10, 3–10 2, 3, 95

Winston, J. T., Koepp, D. M., Zhu, C., Elledge, S. J., and Harper, J. W. (1999), A family of mammalian F-box proteins, *Curr. Biol.*, 9, 20, 1180–1182 4

Winton, M. J., Igaz, L. M., Wong, M. M., Kwong, L. K., Trojanowski, J. Q., and Lee, V. M. (2008), Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation, *J. Biol. Chem.*, 283, 19, 13302–13309 15, 92

Wu, C. H., Fallini, C., Ticozzi, N., Keagle, P. J., Sapp, P. C., Piotrowska, K., et al. (2012), Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis, *Nature*, 488, 7412, 499–503 3, 99

Wu, J. and Jiang, R. (2013), Prediction of deleterious nonsynonymous single-nucleotide polymorphism for human diseases, *ScientificWorldJournal*, 2013, 675851

Wu, L. S., Cheng, W. C., and Shen, C. K. (2013a), Similar dose-dependence of motor neuron cell death caused by wild type human TDP-43 and mutants with ALS-associated amino acid substitutions, *J. Biomed. Sci.*, 20, 33 11

Wu, L. S., Cheng, W. C., and Shen, C. K. (2013b), Similar dose-dependence of motor neuron cell death caused by wild type human TDP-43 and mutants with ALS-associated amino acid substitutions, *J. Biomed. Sci.*, 20, 33 95, 96, 97, 98

Yang, C., Tan, W., Whittle, C., Qiu, L., Cao, L., Akbarian, S., et al. (2010), The C-terminal TDP-43 fragments have a high aggregation propensity and harm neurons by a dominant-negative mechanism, *PLoS ONE*, 5, 12, e15878 10, 11

Zawaira, A., Pooran, A., Barichievy, S., and Chopera, D. (2012), A discussion of molecular biology methods for protein engineering, *Mol. Biotechnol.*, 51, 1, 67–102 89

Zhang, W., Yang, J., He, B., Walker, S. E., Zhang, H., Govindarajoo, B., et al. (2015), Integration of QUARK and I-TASSER for Ab Initio Protein Structure Prediction in CASP11, *Proteins* 29, 88, 89, 98

Zhao, T., De Graaff, E., Breedveld, G. J., Loda, A., Severijnen, L. A., Wouters, C. H., et al. (2011), Loss of nuclear activity of the FBXO7 protein in patients with parkinsonian-pyramidal syndrome (PARK15), *PLoS ONE*, 6, 2, e16983 8

Zheng, L., Baumann, U., and Reymond, J. L. (2004), An efficient one-step site-directed and site-saturation mutagenesis protocol, *Nucleic Acids Res.*, 32, 14, e115 15