Characterisation of RNA uridylyltransferases in *Schizosaccharomyces pombe*

A dissertation presented for the degree of D. Phil.
Michaelmas Term 2012

Daniel Dehany Scott
Sir William Dunn School of Pathology
Exeter College
University of Oxford

So come up to the lab,

And see what's on the slab;

I see you shiver with anticip . . .

Abstract

Daniel D. Scott

Exeter College

Submitted for the degree of D. Phil.

Michaelmas Term 2012

Characterisation of RNA uridylyltransferases in *Schizosaccharomyces pombe*

The control of RNA stability and function via 3' end modification is a widely conserved and biologically important regulatory mechanism. Though this control has traditionally been considered to depend entirely on the addition or removal of long or short poly(A) tails, in recent years the post-transcriptional addition of uridylyl residues to the 3' ends of diverse RNAs has been identified as a physiologically relevant degradative signal in a range of different species. The *Schizosaccharomyces pombe* protein Cid1 has been previously shown to possess uridylylation activity on mRNAs and via this uridylylation to promote the decapping and degradation of mRNAs. This work investigates the presence of residual uridylylation activity in strains lacking the *cid1* uridylyltransferase and shows that a second *S. pombe* enzyme, *cid16*, possesses robust and highly specific poly(U) polymerase (PUP) activity *in vitro* and is able to uridylylate the *act1* mRNA *in vivo* Characterisation of Cid16 shows that it possesses greater processivity and selectivity than Cid1 both for RNA substrates and for UTP, suggesting that Cid16 is a more stringent PUP than is Cid1. Deletion of *cid16* causes few changes in mRNA levels during exponential growth, suggesting that *cid16* may act on other targets or during other phases of the *S. pombe* life cycle, including a potential role in the regulation of RNAi factors during meiosis. Separate experiments show that, while *cid1* is dispensable for the stability of most RNAs during exponential growth, it unexpectedly regulates the transcription of genes in extended subtelomeric regions of the *S. pombe* genome, suggesting a hitherto unknown role for *cid1* in the regulation of subtelomeric heterochromatin formation and/or propagation. These observations suggest that the process of uridylylation in *S. pombe* is significantly more complicated than previously suspected and may regulate a range of different targets in diverse biological pathways.

Acknowledgments

Thanks must first of all go to my supervisor Chris, who has provided an enthusiastic, optimistic and above all calming influence at the times when it was most needed. I must also thank all the members of the Norbury lab past and present, and in particular Sophie Fleurdépine and Marie-Joëlle Schmidt, who have made the long days in the lab friendly and enjoyable – and often provided that much-needed coffee break/sounding board. During this D.Phil. I was lucky enough to collaborate with Dr. Juan Mata and his lab, in particular Ms. Ayesha Hasan, who were always helpful and friendly, and tolerated my minimal knowledge of microarray methodology with the greatest of fortitude. Monika Gullerova and Sneha Shah always provided a kind ear and the help of a fellow pombologist when required.

Outside of the lab, there have been many groups who have made my time in Oxford a memorable and immensely fun experience – the Donut Kings who were brave (foolish?) enough to entrust me with their direction; the Exeter College Boat Club, who taught me that a horse can run itself to death but a human will pass out first; the volunteers as Nightline, who welcomed me in to one of the kindest and most supportive groups in town; and my many Australian friends who provided a safety net, a terrible sense of humour and a welcome slice of home when it was needed. I will remember their friendship and camaraderie with great fondness.

To Chris, Katie and Laura: I couldn't have hoped for three better flatmates to wander around South Oxford with – thank you for your friendship, hugs, cider, humour, mad-cap bakery and for not singing (too many . . .) Rolf Harris songs. Finally, I must thank most of all my partner Amy, and my family – Mum, Dad, Steve and Peter. Between you, you saw me through the best times and the worst, provided unconditional support and love, and made everything worthwhile. I wouldn't have finished this without you.

Table of Contents

Abbreviations

'	minute
"	second
+/-RT	with/without reverse transcriptase
µg	microgramme
µL	microlitre
µM	micromolar
4-SU	4-thiouridine
Å	Ångstrom
A_{600}	absorbance at 600nm
Ago	Argonaute
AMP	adenine monophosphate
ARC	Argonaute siRNA chaperone complex
bp/kb	base pairs/kilobases
BSA	bovine serum albumin
cDNA	copy DNA
CGs	convergent genes
Ci	Curie
CLRC	Clr4 complex
cm	centimetre
cRACE	circularised rapid amplification of cDNA ends
CTD	C-terminal domain
CUT	cryptic unstable transcript
(d)ATP	(deoxy)adenosine triphosphate
dCMP	deoxycytidine monophosphate
(d)CTP	(deoxy)cytosine triphosphate
(d)GTP	(deoxy)guanosine triphosphate
dH_2O	distilled H2O
DNAse	deoxyribonuclease
(d)NTP	(deoxy)nucleotidyl triphosphate
dsRNA	double-stranded RNA
DTT	Dithiothreitol
dTTP	deoxythiamine triphosphate
EBV	Epstein-Barr virus
EDTA	ethylenediaminetetraacetic acid
EMM	Edinburgh minimal media
ERI	enhanced RNAi
EV	empty vector
g	gramme

x *g*	times gravity
GO	gene ontology
gRNA	guide RNA
GST	glutathione-S-transferase
hr	hour
HU	hydroxyurea
IgG	Immunoglobulin G
IPTG	isopropyl-β-D-thiogalactopyranoside
IR	inverted repeat
kDa	kiloDalton
L	litre
LB broth	Luria-Bertani broth
M	molar
Me	methyl group
ME4S	malt extract medium containing 4 supplements
mg	miilligramme
miRNA	microRNA
miR:miR*	miRNA:miRNA* dsRNA duplex
mJ	milliJoules
mL	millilitres
mM	millimolar
mRNA	messenger RNA
M_w	molecular weight
ncPAP	non-canonical poly(A) polymerase
ng	nanogramme
nm	nanometre
nt	nucleotide(s)
NTAP	N-terminal tandem affinity purification
NTD	N-terminal domain
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PAP	poly(A) polymerase
PCP	poly(C) polymerase
PCR	polymerase chain reaction
PGP	poly(G) polymerase
piRNA	Piwi-interacting RNA
pmol	picomole
PNPase	polynucleotide phosphorylase
PTGS	post-transcriptional gene silencing
PUP	poly(U) polymerase

qRT-PCR	quantitative reverse transcriptase PCR
RDE	RNAi defective
RDRC	RNA-dependent RNA polymerase complex
RdRP	RNA-dependent RNA polymerase
RISC	RNA-induced silencing complex
RITS complex	RNA-induced transcriptional silencing complex
RNAi	RNA interference
RNAse	ribonuclease
rpm	rotations per minute
rRNA	ribosomal RNA
RT	room temperature
S	Svedberg unit
SAP	shrimp alkaline phosphatase
scnRNAs	scan RNA
s.d.	standard deviation
SDS	sodium dodecylsulfate
siRNA	small interfering RNA
siR:siR*	siRNA:siRNA* dsRNA duplex
SLBP	stem-loop binding protein
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
sRNA	small RNA
ssRNA	single-stranded RNA
$t_{½}$	half-life
TAP	tobacco acid pyrophosphatase
TEMED	N,N,N',N'-tetramethylethylene-1,2-diamine
TGS	transcriptional gene silencing
TRAMP complex	Trf4p-Air1/2p-Mtr4p polymerase complex
tRNA	transfer RNA
TSS	transcriptional start site
TUTase	terminal U transferase
U	enzymatic unit
UMP	uracil monophosphate
UTP	uracil triphosphate
UTR	untranslated region
UV	ultraviolet
V	Volt
W	Watt
wt	*wild-type*
YE5S	yeast extract medium containing 5 supplements

Chapter 1
Introduction

The coordinate control of RNA transcription and decay in eukaryotic cells is essential to maintain appropriate levels of RNA species and coordinate myriad biological functions. Though decay pathways are controlled by many different factors and regulate many different RNAs, each involves the production of a free 5' and/or 3' end that is acted upon by processive exoribonucleases (Andrade *et al.* 2009; Tomecki *et al.* 2010). Control of RNA degradation at the 3' end was previously thought to rely either on the removal of the stabilising poly(A) tail added to mRNAs by the canonical poly(A) polymerase (PAP; Lund and Dahlberg 1992; Chen *et al.* 2005; Li *et al.* 2005) or on the addition of a destabilising oligo(A) tail by nucleoplasmic nucleotidyltransferases (LaCava *et al.* 2005; Vaňáčová *et al.* 2005; Wyers *et al.* 2005). However, in the last 10 years evidence has accumulated for widespread and biologically important RNA regulation by a second nucleotidyltransferase activity – the addition of uridylyl residues to RNAs by an evolutionarily conserved but functionally diverse class of nucleotidyltransferases, termed the non-canonical poly(A) polymerases (ncPAPs; Aravind and Koonin 1999; Stevenson and Norbury 2006; Wilusz and Wilusz 2008).

In this introduction, the prevalence and varied functions of RNA uridylylation will be reviewed, starting with the diverse pathways of uridylylation-mediated RNA degradation and continuing on to the involvement of uridylylation in non-degradative cellular processes. The family of nucleotidyltransferases responsible for RNA uridylylation, and the current knowledge regarding the functions of family members in the fission yeast *S. pombe*, will be discussed. Finally, a brief introduction into the formation of heterochromatin in *S. pombe* and the role of RNA species in this process will be provided.

1.1 Uridylylation and degradation of mRNAs

Pol II-transcribed mRNAs are modified cotranscriptionally with two key determinants of stability and translatability, the 5' methyl-7-guanosine cap and 3' homopolymeric adenylyl (poly(A)) tail (Proudfoot *et al.* 2002). Removal of either of these modifications is sufficient to prevent translation of the mRNA and elicits rapid degradation; however, for most organisms studied, the primary and rate-limiting step of mRNA degradation is removal of the poly(A) tail by one or more deadenylases (Goldstrohm and Wickens 2008; Chen and Shyu 2011). Deadenylation of an mRNA promotes degradation either by stimulating decapping and 5'→3' exonucleolysis or by allowing the exosome to continue 3'→5' exonucleolytic decay (Chen and Shyu 2011). However, the recent identification of several deadenylation-independent degradation pathways relying on the action of cytoplasmic uridylyltransferases suggests that regulation of mRNA degradation by 3' modification may be more complicated than previously suspected.

1.1.1 Uridylylation of miRNA 5' cleavage products

Unlike in animals, *Arabidopsis* micro RNAs (miRNAs) bind with perfect complementarity to their mRNA template and elicit endonucleolytic cleavage by the Argonaute protein (Ago), forming 5' and 3' cleavage products which must be degraded (Guleria *et al.* 2011). Whilst the 3' fragment is degraded by the Xrn1-like 5'→3' exonuclease AtXRN4 (Souret *et al.* 2004), sequencing analysis demonstrated that, instead of being exonucleolytically degraded 3'→5' by the exosome, the 5' cleavage product is oligouridylylated and degraded by decapping and 5'→3' exonucleolysis (Shen and Goodman 2004). This report represented the first evidence for the uridylylation of eukaryotic mRNAs, and while it initially seemed a surprising mechanism for the degradation of an RNA possessing a free 3'-OH end available for exonucleolytic degradation, such uridylylation-mediated decapping and 5'→3' degradation may be important as a means to prevent translation of cleaved mRNA products which would produce aberrant, C-terminally truncated proteins (Shen and Goodman 2004).

Uridylylation of miRNA cleavage products has also been observed on the 5' moiety of murine *HoxB8* mRNA cleaved by miR-196 (Shen and Goodman 2004) and, potentially, on a truncated Epstein-Barr virus (EBV) *Pol* mRNA cleaved by a virally-encoded miRNA prior to polyadenylation in a marmoset (*Saiguinus oedipus*) cell line (Furnari *et al.* 1993; Pfeffer *et al.* 2004; Shen and Goodman 2004), supporting a potentially ubiquitous role for this degradation mechanism. In a similar mechanism, the *Chlamydomonas* small RNA (sRNA) poly(U) polymerase (PUP) Mut68 has been implicated in the oligoadenylation of miRNA 5' cleavage products (Ibrahim *et al.* 2006; Ibrahim *et al.* 2010), raising the possibility that this enzyme may act simultaneously on the closely juxtaposed 3' ends of the miRNA and the 5' cleavage product.

1.1.2 Deadenylation-independent decapping in fungi

The *cid1* gene of *S. pombe* was initially identified in an overexpression screen for rescue of DNA replication checkpoint defects and was subsequently found to play subtle roles in the maintenance of this checkpoint in response to genetic or pharmacological inhibition of DNA replication (Wang *et al.* 1999; Wang *et al.* 2000). Initial bioinformatic and biochemical analyses suggested that Cid1 was a cytoplasmic poly(A) polymerase (Wang *et al.* 2000; Read *et al.* 2002); however, subsequent work has demonstrated Cid1 to possess robust PUP activity which outcompetes its PAP activity *in vitro* and *in vivo* (Rissland *et al.* 2007). Surprisingly, Cid1 was able to add mono- or di-uridylyl tails to mature polyadenylated mRNA transcripts *in vivo* in a manner independent of poly(A) tail length (Rissland *et al.* 2007). As subsequently reported for histone mRNAs (Mullen and Marzluff 2008), these tails were found mainly on decapped mRNAs but accumulated on capped transcripts upon deletion or inhibition of decapping factors *lsm1* and *dcp1* respectively, suggesting that Cid1 uridylylation may promote a novel, decapping-independent pathway of bulk mRNA decay (Rissland *et al.* 2007; Rissland and Norbury 2009). Importantly, Cid1-dependent uridylylation also accumulated upon deletion of the major deadenylase *ccr4*, whilst poly(A) tails on decapped transcripts were significantly shorter in *cid1Δ* strains, suggesting that uridylylation and deadenylation operate

redundantly to promote degradation of mRNAs. In support of this hypothesis, deletion of either *cid1* or *ccr4* partially stabilised the uracil-responsive *urg1* transcript but neither did so to the same extent as deletion of the common decapping activator *lsm1* (Rissland and Norbury 2009).

In *Aspergillus nidulans*, recent analyses of deadenylation enzymes illuminated a similar pathway to that in *S. pombe*. Morozov *et al*. (2010a) demonstrated that the previously uncharacterised ncPAP CutA (CU nucleotidyltransferase) of *Aspergillus* was able to add heteropolymeric oligo (C/U) tails with a consensus CUCU sequence to mRNAs; however, unlike Cid1, CutA exhibited exclusive specificity for deadenylated mRNAs in wild-type cells. Deletion of this enzyme impaired bulk mRNA decay (Morozov *et al.* 2010a) and resulted in the disassembly of P-bodies (Morozov *et al.* 2010b), suggesting that CUCU addition represents an intermediate step in decapping-mediated degradation of mRNAs subsequent to deadenylation rather than eliciting an independent, deadenylation-independent pathway like Cid1. However, findings that cells deleted for deadenylase enzymes exhibited increased CutA activity and that this activity was extended to target mRNAs possessing mature poly(A) tails suggest that *A. nidulans* is capable of activating a CutA-dependent pathway of mRNA degradation upon inhibition of the dominant deadenylation pathway (Morozov *et al.* 2010a).

One surprising observation was the extremely short nature of the uridyl tails added to mRNA by Cid1 (almost entirely U_{1-2}) and to a lesser extent CutA (all <9 nt in length), especially given the long poly(U) tails added by Cid1 *in vitro* (Rissland *et al.* 2007). However, short oligo(U) tails have been reported to inhibit 3'→5' exonucleolysis in human cells (Ford and Wilusz 1999) and to efficiently recruit the Lsm1-7 decapping complex; even a U_1 tail is sufficient for some enhancement of decapping (Song and Kiledjian 2007). Thus, even mono- or diuridylyl additions to the 3' terminus of an mRNA polyadenylyl tail may be sufficient to override the preference of Lsm1-7 for oligo(A) sequences over poly(A) sequences (Chowdhury *et al.* 2007) and facilitate efficient decapping of mRNAs with mature poly(A) tails. Alternatively, cell cycle- or transcript-specific potentiation of Cid1 activity could facilitate the addition of longer oligo(U) tails, though such activity regulation has not so

far been observed. Importantly, this pathway represents a mechanism of mRNA degradation found in *S. pombe* but not in the more widely investigated *S. cerevisiae*, which lacks ncPAP family members possessing uridylyltransferase activity; while *S. cerevisiae* possesses deadenylation-independent mRNA degradation pathways, these are either gene-specific and regulated by *cis* elements (Badis *et al.* 2004; Muhlrad and Parker 2005) or occur in mutants compromised for stable mRNP formation and, consequentially, for efficient protection of the 5' cap (Morrissey *et al.* 1999). However, many other species possess a large complement of ncPAPs, many of which have been demonstrated to possess uridylyltransferase activity (Aravind and Koonin 1999; Kwak and Wickens 2007; Rissland *et al.* 2007; Figure 1.1, Table 1.1), suggesting that the fungal mRNA uridylylation pathway may be more widely conserved than previously suspected. Importantly, the widespread use of oligo(dT) primers in microarray and sequencing studies will prevent the detection of oligo(U) tails distal to the poly(A) tail, suggesting that this mRNA modification may have gone unobserved in many previous studies.

The observation of residual uridylylation of mRNAs in *S. pombe cid1Δ* strains suggests that another of the six Cid1 family members in *S. pombe* (Aravind and Koonin 1999; Wang *et al.* 2000) may also possess mRNA uridylyltransferase activity (Rissland *et al.* 2007). Low-level homopolymeric oligo(U) tails were also observed in *cutAΔ* strains of *A. nidulans* (Morozov *et al.* 2010a). This is surprising given that the *A. nidulans* genome encodes only one other ncPAP family member, *AN5694.2*, which exhibits significant homology to Trf4/5-like nuclear poly(A) polymerases (See Figure 1.1, Table 1.1; Morozov *et al.* 2010a). While the source of the residual uridylyltransferase activity in these species is currently unknown, the existence of multiple mRNA uridylyltransferase enzymes could allow different regulatory pathways of mRNA uridylylation and/or redundant activity on mRNAs that may mask more severe phenotypes that would result from complete loss of mRNA uridylylation.

1.1.3 Degradation of histone mRNAs in mammals

Mammalian histone mRNAs are unique in that they are not polyadenylated, terminating instead with a conserved stem-loop structure which is bound by the stem-loop binding protein (SLBP) and which

promotes pre-mRNA processing, nuclear export and efficient translation (Dominski and Marzluff 1999; Sanchez and Marzluff 2002; Sullivan *et al.* 2009). In order to maintain accurate DNA:histone stoichiometry, histone mRNAs are expressed coincident with DNA replication and are rapidly degraded upon conclusion of S phase or upon inhibition of DNA replication (Gallwitz and Mueller 1969; Heintz *et al.* 1983; Morris *et al.* 1991). Early work demonstrated that the terminal stem-loop is a *cis*-regulatory element necessary and sufficient for regulated histone mRNA degradation (Pandey and Marzluff 1987); however, it is only active when found within ~70 nt of the terminating ribosome, suggesting a possible interaction with the translation machinery (Kaygun and Marzluff 2005). Degradation additionally requires direct interaction of the nonsense-mediated decay (NMD)-related helicase Upf1 with SLBP and, in the case of DNA replication inhibition, SLBP phosphorylation by the checkpoint kinases ATR and/or DNA-PK (Kaygun and Marzluff 2005; Müller *et al.* 2007). 3'hExo, a conserved single stranded RNA (ssRNA) exoribonuclease implicated in sRNA degradation (Kennedy *et al.* 2004) and biogenesis (Duchaine *et al.* 2006; Pavelec *et al.* 2009) as well as in the maturation of the 5.8S ribosomal RNA (rRNA; Ansel *et al.* 2008) is able to interact with the stem-loop simultaneously and cooperatively with SLBP and trim the 5 nt 3' overhang, leading to suggestions that exonucleolysis by 3'hExo could mediate degradation of the stem-loop and expose the body of the mRNA to exosomal degradation (Dominski *et al.* 2003). However, histone mRNA degradation was subsequently found to utilise both decapping and exosomal pathways (in some cases simultaneously) and to function efficiently in the absence of 3'hExo, suggesting an alternative degradative signal must exist (Mullen and Marzluff 2008). Crucially, in addition to retrieving 5' and/or 3'-degraded transcripts, direct sequencing of histone mRNA ends retrieved a significant proportion of sequences exhibiting oligouridylylation at the 3' ends of mature transcripts. This uridylylation was stimulated upon the conclusion of S-phase or DNA replication inhibition and was largely restricted to decapped transcripts, supporting an oligouridylylation-mediated decapping pathway for histone mRNA degradation (Mullen and Marzluff 2008). Though Mullen and Marzluff (2008) somewhat surprisingly identified the human mitochondrial PAP (Tomecki *et al.* 2004; Nagaike

et al. 2005) and hTRF4-2 orthologue (Shcherbik *et al.* 2010) as required for histone mRNA uridylylation, subsequent work has demonstrated that the responsible enzyme is ZCCHC11 (Schmidt *et al.* 2011), a known cytoplasmic uridylyltransferase which also acts on mature and precursor miRNAs (discussed below).

A number of aspects of the histone mRNA uridylylation degradation pathway still remain unresolved, in particular the role (if any) of 3'hExo and the mechanism by which ZCCHC11 is recruited to and/or activated on the histone mRNA. Given that 3'hExo on histone mRNAs *in vivo* is restricted to trimming the last 3 nt of the stem-loop 3' overhang (Dominski *et al.* 2003) and has similarly been shown to selectively trim ssRNA 3' overhangs of both siRNAs (Yang *et al.* 2006) and 5.8S rRNA (Ansel *et al.* 2008; Gabel and Ruvkun 2008), an intriguing possibility may be that the constitutive presence of 3'hExo on histone mRNAs throughout S phase trims any precocious uridylylation of histone mRNAs by ZCCHC11 prior to the completion of S phase; the subsequent recruitment of the Upf1 helicase (or some other factor) may stimulate release of 3'hExo and/or SLBP or may promote oligouridylylation in such a way as to facilitate decapping and consequent degradation. Such a mechanism would be highly reminiscent of the opposing poly(A) polymerase/deadenylase activities of GLD-2/PARN on oocyte mRNAs in *Xenopus laevis* and, potentially, other organisms (Kim and Richter 2006; Radford *et al.* 2008). In such a model, 3'hExo would in fact be a stabilising partner of histone mRNAs, explaining the lack of accumulation of histone mRNA upon its depletion (Mullen and Marzluff 2008). Alternatively, the observation that 3'hExo binds to histone stem-loops succeeded by an AU_4 tail some 20-fold less than to those with the templated AC_3A tail (Yang *et al.* 2006) suggests that recruitment of ZCCHC11 and its PUP activity may be sufficient to instigate release of 3'hExo and/or other factors and allow degradation of the histone mRNA. It is important to note that both of these models are extremely tentative and still require refinement; however, they provide testable hypotheses regarding the composition of stem-loop-bound multiprotein complexes and the interaction partner(s) of ZCCHC11, which it is hoped may ultimately direct a clearer understanding of the degradation of these tightly regulated mRNAs.

1.2 Uridylylation and degradation of small RNAs

In recent years, an explosion in high-throughput sequencing studies has demonstrated widespread and abundant uridylylation of sRNAs and their precursors in many species including *Drosophila* (Seitz *et al.* 2008; Fernandez-Valverde *et al.* 2010; Berezikov *et al.* 2011), mammals (Landgraf *et al.* 2007; Wu *et al.* 2007; Morin *et al.* 2008; Chiang *et al.* 2010; Newman *et al.* 2011; Wyman *et al.* 2011; Choi *et al.* 2012), *C. elegans* (Ruby *et al.* 2006) and protozoa (Couvillion *et al.* 2009). Until recently a detailed understanding of the biogenesis and physiological relevance of these uridylylation events has been missing; however, recent work is beginning to elucidate the machinery involved in small RNA uridylylation and how this modification changes the landscape of small RNAs in diverse organisms.

1.2.1 The Hen1 RNA methyltransferase

1.2.1.1 HEN1 and sRNA uridylylation in Arabidopsis thaliana

hen1 (*hua enhancer 1*) was first identified in a screen for mutants enhancing plant reproductive organ development defects in the *Arabidopsis* double mutant *hua1-1 hua2-1* (Chen *et al.* 2002). The protein is highly conserved across many species (Park *et al.* 2002; Tkaczuk *et al.* 2006), though the structure is noticeably different between animal and plant kingdoms (Park *et al.* 2002). Consistent with a conserved role of small RNAs in plant flower development (Nag and Jack 2010), analysis of the mutant alleles retrieved from the mutagenesis screen showed that HEN1 was required for stability of both miRNAs (Park *et al.* 2002) and sense post-transcriptional gene silencing (S-PTGS)-mediating small interfering RNAs (siRNAs) (Boutet *et al.* 2003; Yu *et al.* 2005) in *Arabidopsis*, suggesting a role at the nexus of small RNA processing.

Several biochemical studies demonstrated that HEN1 was an RNA methylase *in vitro* which specifically targeted the 3'-terminal residues of miR:miR* (Yu *et al.* 2005) and siR:siR* duplexes (Li *et al.* 2005) and that this methylation specifically occurred on the 2'-OH of the 3'-terminal nucleotide.

This activity was highly specific for double-stranded precursors of miRNAs/siRNAs, showing specificity for 21-24 nt double-stranded (dsRNAs) with a 3' 2 nt overhang characteristic of Dicer family cleavage reactions (Yang *et al.* 2006). Addition of a 2'-O-methyl group to synthetic miRNAs *in vitro* was shown to inhibit several 3' modifying enzymes (Yang *et al.* 2006), and loss of 2'-O-methylation *in vivo* coincided with destabilisation of endogenous miRNAs (Yu *et al.* 2005). A mechanistic basis for the stabilising effects of HEN1 was elucidated by Li *et al*. (2005) who showed that loss of HEN1 *in vitro* resulted in increased 3' oligouridylylation and 3'→5' exonucleolytic trimming of miRNAs and siRNAs, suggesting HEN1 methylation was a protective modification preventing degradation of sRNAs by uridylylation-coupled exonucleolysis. Despite HEN1 targeting the dsRNA forms of sRNAs, no uridylylation was found on passenger miR* or siR* strands, suggesting the uridylylation was happening after release of the passenger strand, i.e. on mature RNA-induced silencing complexes (RISCs) and was thus decoupled from the activity of HEN1 (Li *et al.* 2005). A pair of recent papers have identified *Arabidopsis* HESO1 (*hen1 suppressor1*) as a nucleocytoplasmic ncPAP able to oligouridylylate and promote degradation specifically of unmethylated siRNAs and miRNAs (Ren *et al.* 2012; Zhao *et al.* 2012).

1.2.1.2 The HEN1 methyltransferase and uridylylation in Caenorhabditis elegans

Consistent with the extensive conservation of HEN1 orthologues across phyla, the HEN1-catalysed methylation-uridylylation pathway is conserved and biologically relevant in diverse organisms. Three concurrent papers have recently identified bioinformatically a single *Caenorhabditis elegans* (*C. elegans*) homologue of *Arabidopsis* HEN1 (*hen1 of nematodes 1, henn-1*) (Billi *et al.* 2012; Kamminga *et al.* 2012; Montgomery *et al.* 2012) and demonstrated it possesses *Arabidopsis* HEN1-like 2'-O-methyltransferase activity; however, unlike *Arabidopsis* HEN1, HENN1 preferentially methylates ssRNA rather than dsRNA substrates *in vitro* (Kamminga *et al.* 2012). Mutations in *henn1* were shown to destabilise *C. elegans* germline piRNAs (also known as 21U sRNAs; germline-specific 24-31nt sRNAs antisense to selfish genetic elements; Ruby *et al.* 2006) and additionally to destabilise

a particular subset of primary endo-siRNAs (26G siRNAs; Ruby *et al.* 2006) which bind to the oocyte/embryo-specific ERGO-1 Ago protein, but not those that bind to the spermatogenesis-specific ALG-3/4 Agos (Billi *et al.* 2012; Kamminga *et al.* 2012; Montgomery *et al.* 2012). Kamminga *et al*. (2012) show that, like in *Arabidopsis* (Li *et al.* 2005), this decrease in 26G siRNAs in *henn1* mutants is correlated with an increase in their rate of uridylylation and with 3' end shortening symptomatic of 3'→5' exonucleolysis. Whilst no analogous uridylylation was observed on piRNAs by this group, it is noted that they, unlike their competitors (Billi *et al.* 2012; Montgomery *et al.* 2012), observed no effect of *henn1* mutants on piRNA stability. This is likely due to the fact that their deep sequencing analysis was restricted to worms in early adulthood, at which developmental stage HENN1, whose regulation of piRNAs is developmentally regulated, shows no activity on piRNAs (Montgomery *et al.* 2012). The observation of *henn1* mutant-mediated destabilisation of piRNAs at different developmental stages (Billi *et al.* 2012; Montgomery *et al.* 2012) leaves open the possibility that a uridylylation-exonucleolysis pathway also operates on piRNAs at particular developmental stages.

The stabilising effects of *henn1* mutants on targets of piRNA and siRNA silencing are reproducible but modest, suggesting that residual 21U/26G levels are sufficient to establish silencing of these targets (Billi *et al.* 2012; Kamminga *et al.* 2012; Montgomery *et al.* 2012). This is somewhat surprising given that *henn1* mutants exhibit a temperature-dependent sterility phenotype similar to that observed in mutants of *prg1/2* (the PIWI-clade Ago proteins that complex with and cleave targets of 21U sRNAs) and suggests that some functionality of HENN1 may be going unobserved. *henn1* mutants also exhibit an unusual combination of ERI (Enhanced RNA Interference) phenotypes in somatic tissues but RDE (RNAi Defective) phenotypes in germline tissues (Billi *et al.* 2012; Kamminga *et al.* 2012; Montgomery *et al.* 2012). It is probable that the ERI phenotypes in somatic tissues are due to the reduced activity of 26G and secondary 22G siRNA pathways liberating rate-limiting RNAi factors, thereby allowing greater activity of the parallel exogenous RNAi pathways (Billi *et al.* 2012; Kamminga *et al.* 2012); however, the RDE phenotype in germline tissues is harder to

explain. It is possible that piRNAs in the *C. elegans* germline target and antagonise suppressors of exogenous RNAi (Kamminga *et al.* 2012); alternatively, HENN1 may methylate hitherto unidentified sRNAs with roles in germline exogenous RNAi (Billi *et al.* 2012).

1.2.1.3 DmHen1/Pimet and uridylylation in Drosophila melanogaster

HEN1 activity in animals was first suggested by the observation that *Drosophila* piRNAs, as well as those of mice and cattle, possessed a similar 2'/3'-OH 3' terminal modification to the *Arabidopsis* miRNAs/siRNAs (Vagin *et al.* 2006; Pélisson *et al.* 2007), and subsequently that this modification could also be found on siRNAs but not on miRNAs (Horwich *et al.* 2007; Pélisson *et al.* 2007; Saito *et al.* 2007; Ameres *et al.* 2010; Ameres *et al.* 2011). This modification was confirmed to be 2'-O-methylation by two different methods (Horwich *et al.* 2007; Saito *et al.* 2007) and was catalysed on ssRNA substrates by the lone *Arabidopsis* HEN1 orthologue in *Drosophila*, alternatively named DmHen1 (Horwich *et al.* 2007) or Pimet (piRNA methyltransferase; Saito *et al.* 2007). Loss of 2'-O-methylation in *Drosophila* was reported to cause increased size heterogeneity of piRNAs and upregulation of a highly sensitive downstream piRNA target transposon mRNA (Horwich *et al.* 2007), suggesting DmHen1/Pimet may be an important factor for *Drosophila* germline piRNA function. However, others did not observe such effects (Saito *et al.* 2007), suggesting that HEN1 may regulate a specific subset of piRNAs/piRNA targets, or that its activity on piRNAs may be regulated in a similar way to *C. elegans* HENN1 (Montgomery *et al.* 2012).

Though no mechanistic information regarding the role of HEN1 methylation of *Drosophila* piRNAs is available, investigations into the methylation of Ago2-bound siRNAs in *Drosophila* (Ameres *et al.* 2010, 2011) have revealed a similar pathway to that observed in both *Arabidopsis* and *C. elegans* (Li *et al.* 2005; Kamminga *et al.* 2012). Disruption of the *hen1/pimet* locus resulted in loss of siRNA methylation *in vivo*, correlated with a marked increase in mono- or oligouridylylated siRNAs and increased 3'→5' exonucleolytic trimming (Ameres *et al.* 2010). Interestingly, these coincident uridylylation and exonucleolysis activities were dependent on the presence of RNA targets

containing sequences perfectly complementary to the small RNA, since both activities could be induced on otherwise stable Ago1-bound miRNAs by the expression of an artificial construct containing sites of perfect complementarity to miRNAs. A similar effect was seen in HeLa cells expressing RNAs containing sites of perfect miRNA complementarity (Ameres *et al.* 2010) and in *C. elegans* larval lysates (Baccarini *et al.* 2011). This perfect complementarity-driven degradation of Ago1-bound small RNAs was required for efficient *in vivo* partitioning of a set of hairpin-derived siRNAs with miRNA-like siR:siR* duplex structures to Ago2 – though the siRNAs loaded onto both Ago1 and Ago2 *in vitro*, Ago1-specific degradation ensured they were functionally restricted to Ago2 *in vivo* (Ameres *et al.* 2011).

1.2.1.4 HEN1 and uridylylation in the zebrafish Danio rerio

Observations in zebrafish provide further support for a conserved role for HEN1 methylation as an antagonist of uridylyltransferase activity. Zebrafish Hen1 was shown *in vitro* to be a ssRNA-specific 3'-terminal methyltransferase which, upon depletion *in vivo*, caused a substantial decline in piRNA abundance (Kamminga *et al.* 2010) caused by loss of the constitutive piRNA methylation previously observed (Houwing *et al.* 2007). Deep sequencing of piRNAs in zebrafish *hen1* mutants showed that their decline was correlated with an increase in both 3' uridylylation and exonucleolytic 3'→5' trimming. The extent of uridylylation and severity of downregulation of piRNAs upon *hen1* mutation was correlated with the level of expression of transcripts from their target locus, suggesting that the uridylylation and degradation activity may be dependent on the presence of target transcripts (Kamminga *et al.* 2010) as observed in *Drosophila* (Ameres *et al.* 2010). The production of viable and fertile, but entirely male, offspring in zebrafish *hen1* mutants is reminiscent of phenotypes observed in mutants of *ziwi* and *zili*, the two zebrafish PIWI-clade Ago proteins that bind piRNAs and degrade their targets (Houwing *et al.* 2007; Houwing *et al.* 2008), and observations that this phenotype is correlated with delayed germline development and slower sex determination reinforce the

hypothesis of a biologically important role for Hen1 in the regulation of zebrafish piRNA stability (Kamminga *et al.* 2010).

1.2.1.5 The MUT68 nucleotidyltransferase in Chlamydomonas reinhardtii

In *Chlamydomonas reinhardtii*, a mutagenesis screen for loci involved in transgenic inverted repeat (IR)-mediated RNAi (Ibrahim *et al.* 2006) identified Mut68, a mutation that results in the accumulation of siRNAs and miRNAs *in vivo* (Ibrahim *et al.* 2006, 2010). Deep-sequencing of small RNAs from a wild-type strain showed the presence of uridylylation on siRNAs and miRNAs which was significantly reduced (though not abolished entirely) in the Mut68 background. This decrease correlated with a reduction in 3'→5' exonucleolytic trimming of siRNAs (Ibrahim *et al.* 2010). The mapping of Mut68 to a gene (*MUT68*) encoding an ncPAP (Ibrahim *et al.* 2006) exhibiting PAP and PUP activity *in vitro* (Ibrahim *et al.* 2010) suggests that Mut68 may be responsible for destabilising small RNAs by post-transcriptional uridylylation. As in *Arabidopsis*, both siRNAs and miRNAs in *Chlamydomonas* are constitutively 3'-end modified, likely by a single-copy orthologue of *Arabidopsis* HEN1 (Molnár *et al.* 2007). These observations, coupled with the finding that MUT68 PAP/PUP activity is inhibited *in vitro* by 2'-O-methylation, suggests that MUT68 may be one of two or more small RNA uridylyltransferases whose activity is inhibited by HEN1 (Ibrahim *et al.* 2010).

Surprisingly, earlier work described a discrete enzymatic and biological role for MUT68 as a cytoplasmic PAP that adenylates 5' products of miRNA/siRNA-directed endonucleolytic cleavage of mature mRNAs (Ibrahim *et al.* 2006). This adenylation occurs simultaneously with 3'→5' exonucleolytic degradation, suggesting it may be a degradative mark akin to uridylylation of 5' cleavage products in *Arabidopsis* (Shen and Goodman 2004) and TRAMP complex adenylation of nuclear transcripts (LaCava *et al.* 2005; Vaňáčová *et al.* 2005; Wyers *et al.* 2005). Though an ability of MUT68 to modify different targets with different enzymatic mechanisms would be highly unexpected, it is known that several ncPAPs including MUT68 exhibit multiple PNP activities *in vitro* (Rissland *et al.* 2007; Ibrahim *et al.* 2010), and possibly *in vivo* (Trippe *et al.* 2006; Mellman *et al.*

2008), suggesting the possibility that cofactor-mediated modulation of MUT68 activity could facilitate this bifunctionality (Ibrahim *et al.* 2010).

1.2.1.6 HEN1-like RNA methyltransferases in other species

A number of organisms with genome-encoded HEN1 orthologues have further been shown to methylate small RNAs on their 2'-OH group, though this methylation has thus far not been demonstrated to be a uridylylation-inhibitory modification. piRNAs purified from mouse testes were shown by two independent papers to possess near-total 2'-OH methylation (Kirino and Mourelatos 2007; Ohara *et al.* 2007); this modification appeared to protect synthetic piRNAs from degradation in mouse testes extracts (Kirino and Mourelatos 2007). mHEN1, the murine HEN1 orthologue, was found to possess 2'-OH methyltransferase activity on ssRNAs of varying lengths *in vitro* when expressed in *E. coli* (Kirino and Mourelatos 2007). 2'-OH methylation of piRNAs has also been shown to occur in other mammals including cattle (Horwich *et al.* 2007) and rats (Houwing *et al.* 2007). The unicellular ciliate *Tetrahymena thermophila* expresses a Hen1 orthologue which specifically methylates conjugation-expressed scan RNAs (scnRNAs) (Kurth and Mochizuki 2009), a particular class of small RNA required for heterochromatin establishment and consequent DNA elimination in the developing macronucleus (for review, see Kataoka and Mochizuki 2011). Methylation prevents the degradation of scnRNAs, probably by 3'→5' exonucleolysis, and is required for efficient DNA elimination in the macronucleus, a process essential for offspring viability (Kurth and Mochizuki 2009). In vegetative cells, Hen1 specifically methylates 23-24 nt sRNAs bound only to Twi8p (*Tetrahymena* Piwi 8) but not those bound other vegetative growth-specific Twi proteins (Couvillion *et al.* 2009). *In vitro*, *Tetrahymena* Hen1 is specific for ssRNA and likely acts on the scnRNAs/sRNAs after loading onto their associated PIWI-class Agos, Twi1p/Twi11p/Twi8p (Kurth and Mochizuki 2009). Finally, analysis of piRNA maturation in cell lysates from the silkworm *Bombyx mori* has demonstrated that pre-piRNAs are loaded onto PIWI proteins as long precursors, which are then trimmed to mature length and 2'-O-methylated, probably by a HEN1 orthologue (Kawaoka *et al.*

2011). Interestingly, Hen1 homologues have even been reported in a variety of bacterial species where they reside in an operon with polynucleotide kinase-phosphatase (Pnkp), a key bacterial end-healing enzyme (Jain and Shuman 2010).

1.2.2 Guiding principles of the sRNA methylation/uridylylation pathway

Extensive biochemical (Yang *et al.* 2006) and structural (Huang *et al.* 2009) studies have demonstrated that *Arabidopsis* HEN1 selectively methylates double-stranded precursors of both miRNAs and siRNAs generated by Dicer cleavage of pre-miRNA or long dsRNA precursors. This selectivity is conferred by an extended N-terminus containing several dsRNA binding domains (dsRBDs) and a La motif-containing domain (LCD) which cooperate to restrict HEN1 activity to dsRNAs of 21-24 nt. By contrast, this long N-terminal extension is absent in all known HEN1 orthologues in animals (Park *et al.* 2002; Tkaczuk *et al.* 2006) and consequently animal HEN1 orthologues exhibit completely different substrate specificity to *Arabidopsis*, with *in vitro* assays in several species demonstrating ssRNA-specific methyltransferase activity (Horwich *et al.* 2007; Kirino and Mourelatos 2007; Saito *et al.* 2007; Kurth and Mochizuki 2009; Kamminga *et al.* 2012). Given that sRNAs are loaded onto Ago proteins as duplexes prior to degradation/release of the passenger strand (Klattenhoff and Theurkauf 2008; Sashital and Doudna 2010) animal HEN1 orthologues likely differ significantly from *Arabidopsis* HEN1 and specifically methylate sRNAs in their mature, Ago-bound forms. Indeed, experiments in both *Drosophila* (Horwich *et al.* 2007) and *C. elegans* (Billi *et al.* 2012) demonstrated that depletion of Ago2 or ERGO-1 respectively abolished siRNA methylation and that, *in vitro*, dsRNA loaded onto *Drosophila* Ago2 was only methylated by recombinant HEN1 after dissociation of the passenger strand to yield mature ssRNA forms (Horwich *et al.* 2007).

Given the necessity for Ago loading of sRNAs for methylation in animals, it appears likely that the sRNA-bound Ago plays a role in recruiting or promoting HEN1-mediated methylation. In *C. elegans*, HEN1 methylates 26G primary siRNAs bound to female germline-specific ERGO-1, but not to male-specific ALG-3/4; the concurrent methylation of piRNAs in both male and female germlines suggests

that this specificity is not an artefact of HEN1 activity or expression and is instead imparted by the differential Argonaute binding of 26G siRNAs (Ameres *et al.* 2010; Billi *et al.* 2012; Kamminga *et al.* 2012; Montgomery *et al.* 2012). Similarly, *Tetrahymena* Hen1p methylates vegetative 23-24 nt sRNAs bound to Twi8p but not to any of the other vegetatively expressed Twis (Couvillion *et al.* 2009), while recombinant DmHen1/Pimet was able to methylate piRNAs, but not miRNAs on their respective purified Ago complexes (Saito *et al.* 2007). Finally, DmHen1/Pimet was able to methylate atypical miRNAs that partitioned onto the siRNA-specific Ago2, but not miRNA-like hp-endo-siRNAs that partitioned onto the miRNA-specific Ago1 (Ameres *et al.* 2010, 2011), suggesting that DmHen1 specificity for Ago2- and PIWI-bound sRNAs is responsible for HEN1 substrate specificity in *Drosophila*.

The most parsimonious explanation of an Ago-mediated specificity of HEN1 would be a direct physical interaction, and indeed several papers have reported HEN1-Ago interactions (Ibrahim *et al.* 2006; Saito *et al.* 2007). However, others have failed to recapitulate these interactions (Horwich *et al.* 2007), whilst HEN1-Ago interactions could not be found in other systems (Billi *et al.* 2012; Kamminga *et al.* 2012). The observation that HEN1-mediated methylation in silkworm lysates requires simultaneous Siwi (Silkworm Piwi) binding and exonucleolytic trimming suggests a model of piRNA methylation whereby HEN1 is only recruited and/or activated once Siwi and the unidentified exonuclease come into close proximity. Such a model could explain the varying observations of HEN1-Piwi interactions, suggesting that these interactions may be transient/indirect or may require a third factor not available in some experiments (Kawaoka *et al.* 2011). *C. elegans* HENN1 (a 50 kDa protein) exists in a ~150 kDa complex *in vivo*, suggesting the possible presence of accessory factors that may govern recruitment or activity (Kamminga *et al.* 2012). However, the divergent maturation pathway of Dicer-dependent methylated small RNAs suggests that, unless HEN1 recruitment and/or activation could be analogously stimulated by components of these maturation pathways (such as non-Ago components of the RISC complex or accessory proteins such as loading complex subunits or passenger strand-degrading exonucleases), a different HEN1 mechanism must exist for these sRNAs.

An alternative hypothesis is suggested by the observation that HEN1 localises to cytoplasmic perinuclear granules, or 'nuages' in germline tissues in several organisms including *C. elegans* (Billi *et al.* 2012; Kamminga *et al.* 2012) and zebrafish (Kamminga *et al.* 2010). These structures are required for the efficacy of RNAi (Claycomb *et al.* 2009) and the biogenesis of piRNAs in the germline (Aravin *et al.* 2009) and contain a number of different methylated sRNA-carrying Ago proteins including Piwi proteins in zebrafish (Houwing *et al.* 2007; Houwing *et al.* 2008), mice (Aravin *et al.* 2009) and *C. elegans* (Batista *et al.* 2008; Wang and Reinke 2008) and the *C. elegans* siRNA Ago ERGO-1 (Claycomb *et al.* 2009; Billi *et al.* 2012). The greatly increased effective concentration of HEN1 mediated by the immediate colocalisation with target Agos in these cytoplasmic granules could be sufficient to drive efficient methylation of small RNAs bound to Agos, whilst distribution of non-target Agos throughout the cytoplasm would not bring them into juxtaposition with HEN1 frequently enough to efficiently methylate their bound sRNAs. The divergent C-terminal domain (CTD) of zebrafish HEN1 mediates its localisation to the nuage, and loss of this domain is sufficient to abolish piRNA methylation activity (Kamminga *et al.* 2010). Analogously, the restriction of *Tetrahymena* HEN1 to the macronucleus in early conjugation ensures its colocalisation with Twi1p, the Piwi protein carrying its target scnRNAs, and not any of the 11 other Piwi-clade Twi proteins expressed in *Tetrahymena* (Ibrahim *et al.* 2010). However, it is noted that other, non-HEN1 target Agos have been shown (Gu *et al.* 2009) or proposed (Billi *et al.* 2012) to localise to perinuclear granules as well as the cytoplasmic lumen, and whilst restriction of HEN1 to different granule subcompartments (Claycomb *et al.* 2009) or to different granule populations (Aravin *et al.* 2009) is possible, a much greater understanding of the dynamics and composition of these granules would be required to assess this hypothesis.

One striking observation arising from the varied examples of sRNA uridylylation above is that all the Ago proteins identified to date that permit HEN1 methylation (bar *Drosophila* Ago2) fall within the PIWI clade of Ago family proteins (Wei *et al.* 2012), a clade that typically binds piRNAs in the germline, though can bind other sRNAs as well. Structural analyses of the 3’-end binding PAZ

domains from Ago (Lingel *et al.* 2004; Ma *et al.* 2004) and PIWI (Simon *et al.* 2011; Tian *et al.* 2011) clade proteins shows that both selectively and tightly bind the two 3'-most nucleotides of loaded RNAs in a way that buries the terminal 2'- and 3'-OH residues deep within a pre-formed hydrophobic pocket. In PIWI clade PAZ domains, however, a short loop insertion between two strands of the β-sheet structure results in a slightly wider binding pocket that specifically and preferentially accommodates 2'-O-methylated sRNA ends; by contrast, the narrower pocket of Ago-clade PAZ domains selectively excludes 2'-OMe structures due to steric clashes with hydrophobic sidechains within the binding pocket. These observations suggest that methylation of PIWI-clade sRNAs is likely essential for the efficient binding of these sRNAs to their Ago proteins, and consequently a key part of their biological function.

Another key feature of the siRNA subpopulations that exhibit the opposed activities of methylation and uridylylation – miRNAs and siRNAs in plants, as well as piRNAs and certain siRNAs in animals – is their preference for binding their RNA targets with complete or near-complete complementarity. By contrast, animal miRNAs typically bind targets with complementarity limited to the 7-8 nt seed sequence at the 5' end and a short region in the middle of the miRNA (Millar and Waterhouse 2005; Ruby *et al.* 2006; Vagin *et al.* 2006; Houwing *et al.* 2007). Importantly, while the lack of 3'-end complementarity in animal miRNA target sites is likely to allow the 3' end of the miRNA to stay bound to the Ago PAZ domain throughout the target interaction cycle, perfect target complementarity among other sRNAs requires the release of the 3' end of the sRNA from the PAZ domain, exposing it to potential metabolism and/or degradative pathways. In this context, 2'-O-methylation may represent an alternative mechanism that has evolved to protect sRNAs that utilise perfect target complementarity from the degradative actions of uridylyltransferases that ubiquitously target the free 3' ends of unmodified sRNAs. Consistent with this hypothesis, uridylylation-mediated degradation can be imposed on miRNAs in both *Drosophila* (Ameres *et al.* 2010) and *C. elegans* lysates (Baccarini *et al.* 2011) by the ectopic expression of perfectly complementary miRNA targets; this effect is attenuated by inducing mispairing at the 3' end of the

miRNA (Ameres *et al.* 2010). Similarly, siRNAs that are aberrantly loaded onto the Ago-clade Ago1 in *Drosophila* are selectively degraded, suggesting their fully complementary binding on this methylation-incompetent Ago is a signal for rapid degradation (Ameres *et al.* 2011).

If this model is correct, the uridylylation of sRNAs may represent a key quality control mechanism for maintenance of the integrity and specificity of the varied sRNA pathways. By rapidly and specifically degrading fully complementary siRNAs/piRNAs loaded onto miRNA-specific Agos unable to recruit methylation activity, uridylyltransferases would ensure that these sRNAs do not compete with miRNA for Ago occupancy and/or redirect their targets into inappropriate degradation/regulation pathways (Ameres *et al.* 2011). Alternatively, and not necessarily exclusively, uridylyltransferases may induce degradation of PIWI-clade sRNAs that have lost 2'-O-methylation through damage or inhibition of methylation activity, providing a mechanism for the maintenance of a pool of mature, active siRNAs/piRNAs on PIWI clade Agos. Though several key mechanistic points are missing from this model, most noticeably the mechanism by which uridylyltransferases are selectively able to recognise free sRNA 3' ends and the mechanism(s) by which this uridylylation is able to signal for degradation of sRNAs, the evolutionarily conserved nature of this pathway suggests that it may represent an important and hitherto underappreciated mechanism of sRNA homeostasis.

1.2.3 CDE-1 uridylylation of CSR-1 pathway 22G siRNAs in C. elegans

In addition to the previously discussed 21U piRNAs and 26G primary RNAs, *C. elegans* express a diverse population of 22 nt, 5'-G-enriched siRNAs (22G siRNAs) which map to a wide variety of loci including repetitive elements, non-coding genes and the antisense strand of many protein-coding genes (Ruby *et al.* 2006). The vast majority of 22G siRNAs antisense to protein-coding genes are found in the germline, where they load onto the Ago CSR-1. Instead of cleaving target mRNAs, 22G siRNA-loaded CSR-1 is directed to chromatin containing target gene sequences as part of a mechanism proposed to define regions of euchromatin for exclusion from the *C. elegans* holocentric centromere structures (Claycomb *et al.* 2009; Gu *et al.* 2009; van Wolfswinkel *et al.* 2009). CSR-1

pathway siRNAs are of much lower abundance than those mediating degradation of repetitive element RNAs in the parallel WAGO-1 pathway (Claycomb *et al.* 2009; Gu *et al.* 2009) and, in *wild-type* germlines, are extensively oligo-uridylylated (Claycomb *et al.* 2009; van Wolfswinkel *et al.* 2009). This uridylylation is catalysed by CDE-1, a germline-specific ncPAP initially identified (along with several other CSR-1 pathway components including *csr-1* and *ekl-1*) in a screen for mutants defective in an RNAi-mediated germline phenomenon termed cosuppression (Robert *et al.* 2005). CDE-1 exhibits poorly processive PUP activity *in vitro* and, *in vivo*, coimmunoprecipitates with EGO1, an RNA-dependent RNA polymerase (RdRP) which, in complex with the helicase DRH-3 and Tudor-domain containing protein EKL-1, is essential for synthesis of CSR-1 pathway 22G siRNAs. *cde-1* ablation causes a marked increase in the abundance of 22G siRNAs of the CSR-1, but not WAGO-1, pathway, and correlates strongly with the loss of uridylylation of CSR-1 22G siRNAs in *wild-type* cells, suggesting that CDE-1-mediated uridylylation destabilises these siRNAs *in vivo* (van Wolfswinkel *et al.* 2009). Interestingly, *cde-1* mutants exhibit a number of phenotypes including repression of many CSR-1 target mRNAs as well as phenotypes symptomatic of decreased activity of other RNAi pathways including mobilisation of transposons and resistance to exogenous RNAi. These observations suggest that the hyperabundant CSR-1 class siRNAs in *cde-1* mutants may be able to aberrantly enter other RNAi pathways, thereby simultaneously causing degradation of CSR-1 target mRNAs and titrating out endogenous siRNAs of these RNAi pathways, decreasing their efficacy (van Wolfswinkel *et al.* 2009).

Surprisingly, another group have described alternative roles for CDE-1 (CID-1 in their paper) and several mitotic cell cycle checkpoint proteins in regulating thermotolerance and life-span in the soma of post-mitotic animals (Olsen *et al.* 2006). Whilst interactions with checkpoint-related machinery have been described for other ncPAPs including the RdRP-associated Cid12 (Win *et al.* 2006) and the cytoplasmic PUP Cid1 (Wang *et al.* 1999; Wang *et al.* 2000), these observations were all made in mitotic cells and relate to the canonical cell cycle functions of these pathways. In addition, van Wolfswinkel *et al.* (2009) were unable to recapitulate the observed somatic expression

of a CDE1::GFP transgene (Olsen *et al.* 2006) with either endogenous CDE-1 mRNA or protein. Though a novel role for CDE-1 and checkpoint proteins in non-dividing somatic tissues is not impossible, the data reviewed above suggest that these observations may be secondary effects of germline-specific activities of these proteins, and that such hypotheses should be treated with caution until a mechanistic description is available.

The proposed CSR-1 pathway in the *C. elegans* germline (Claycomb *et al.* 2009; van Wolfswinkel *et al.* 2009) is highly reminiscent of the RNA-induced transcriptional silencing/RNA-dependent RNA polymerase complex (RITS/RDRC) pathway of *S. pombe*, which establishes centromeric heterochromatin essential for mitosis and meiosis (Motamedi *et al.* 2004). In both pathways, siRNAs are synthesised by an RdRP (EGO-1 in *C. elegans*/Rdp1 in *S. pombe*) in complex with an ncPAP (CDE-1/Cid12) and a helicase (DRH-3/Hrr1) and loaded onto an Ago (CSR-1/Ago1) which is directed to target genomic sequences to regulate chromatin dynamics (Motamedi *et al.* 2004; Claycomb *et al.* 2009; van Wolfswinkel *et al.* 2009). Mutations in components of both pathways elicit similar defects in mitotic structures and chromosome dynamics and cause phenotypes associated with failure of efficient chromosome segregation (Win *et al.* 2006; Claycomb *et al.* 2009; van Wolfswinkel *et al.* 2009). Though the molecular output of the pathways is proposed to be different – the RITS/RDRC pathway establishes centromeric heterochromatin via recruitment of repressive histone modifying enzymes (Motamedi *et al.* 2004), whilst the CSR-1 pathway is proposed to interact with euchromatic regions which are excluded from heterochromatic regions comprising the *C. elegans* holocentric centromere (Claycomb *et al.* 2009) – the conservation of pathway architecture and component functionality suggests that similar ncPAP-containing pathways may regulate chromatin in other systems.

Interestingly, the other characterised 22G siRNA pathway in *C. elegans*, the WAGO-1 pathway, also involves an ncPAP, RDE-3 (Tabara *et al.* 1999; Chen *et al.* 2005), in addition to the partially redundant DRH-3/EKL-1/(EGO-1 or RRF-1) RdRP complexes and the putative exoribonuclease MUT-7

(Gu *et al.* 2009). Though the function of RDE-3 in this pathway is unclear, several *rde-3* mutations deplete WAGO-1 path siRNAs suggesting that, unlike CDE-1, it is required for stability of siRNAs (Gu *et al.* 2009). RDE-3 could stabilise siRNAs by adenylation, as has been observed for GLD-2 in mice and humans (Katoh *et al.* 2009). Alternatively, since RDE-3 was found to have no catalytic activity *in vitro* (Kwak and Wickens 2007), it could be a catalytically inactive competitor of destabilising ncPAPs, excluding them from interactions with components of the WAGO-1 pathway (van Wolfswinkel *et al.* 2009); however, the retrieval of several putative loss-of-function alleles of *rde-3* encoding point mutants in residues essential for catalysis (Chen *et al.* 2005) suggest that catalytic activity is important for RDE-3 function. The presence of ncPAPs in multiple 22G siRNA pathways in *C. elegans* and their interaction with RdRPs in other species including *S. pombe* (Motamedi *et al.* 2004) and *Tetrahymena* (Lee *et al.* 2009; Talsky and Collins 2010) suggests that ncPAPs may be conserved components of siRNA biogenesis and/or turnover pathways in many organisms.

1.2.4 Uridylylation of mature miRNAs

ZCCHC11 is a member of the pol β-like nucleotidyltransferase superfamily which uridylylates diverse cellular substrates including the pre-let-7 miRNA precursor (Hagan *et al.* 2009; Heo *et al.* 2009; Lehrbach *et al.* 2009) and replication-dependent histone mRNA (Schmidt *et al.* 2011), thereby driving their degradation (discussed above and below). Paradoxically, depletion of functional ZCCHC11 resulted in downregulation of several cytokines including IL-6, suggesting ZCCHC11 was required for full stability of these transcripts (Minoda *et al.* 2006; Jones *et al.* 2009); despite this, no ZCCHC11-dependent nucleotidyltransferase activity was observed on *IL-6* transcripts, which instead showed deadenylation upon *ZCCHC11* siRNA treatment (Jones *et al.* 2009). These contradictory observations were resolved by the finding that miR-26, a conserved miRNA that targets *IL-6* via its 3' untranslated region (UTR), was uridylylated *in vitro* and *in vivo* by ZCCHC11, and that this uridylylation was able to inhibit miR-26-mediated *IL-6* mRNA degradation, thereby stabilising *IL-6* mRNA in cells expressing ZCCHC11 (Jones *et al.* 2009). Importantly, other miRNAs targeting cytokines may be regulated by

this mechanism, such as *let-7*, which also targets *IL-6* mRNA (Iliopoulos *et al.* 2009) and is separately regulated by ZCCHC11 in concert with LIN28 (Hagan *et al.* 2009; Heo *et al.* 2009; Lehrbach *et al.* 2009). The large number of miRNAs exhibiting 3' nontemplated uridylylation in immune cells (Wu *et al.* 2007) and other tissues (Landgraf *et al.* 2007; Morin *et al.* 2008) suggests that miR-26 regulation may represent the first example of a widespread miRNA uridylylation pathway. Consistent with this, oligouridylylation has also been observed on a subset of virally-produced miRNAs upon lytic infection of murine fibroblasts with murine cytomegalovirus (MCMV); this uridylylation was correlated with subsequent underrepresentation of these miRNAs in the total sRNA pool, suggesting a degradative mechanism (Dölken *et al.* 2007). Whether this uridylylation is a virally-directed miRNA regulation pathway or a host cell antiviral defence is currently unclear.

Surprisingly, depletion of ZCCHC11 was found to have no effect on the levels of the miR-26b isoform *in vivo*, suggesting that the uridylylation may be inhibiting miR-26 activity by a non-degradative mechanism (Jones *et al.* 2009). However, deep sequencing of miR-26 isoforms has shown remarkably different uridylylation patterns, with miR-26b showing a near-complete absence of 3'-terminal uridylylation in both *wt* and *ZCCHC11(RNAi)* cell lines, whilst the miR-26a isoform, which is ~30-fold more abundant than miR-26b, shows extensive uridylylation in *wt* cells which is abolished in *ZCCHC11(RNAi)* (Jones *et al.* 2009). Though it should be noted that uridylylation has been reported on miR-26b in naïve $CD8^+$ T cells (Wu *et al.* 2007), and that a putative mechanism for isoform-specific uridylylation of miR-26 is not available, these observations suggest that, at least in the cell line tested (Jones *et al.* 2009), isoform-specific uridylylation may promote the degradation of miR-26a, abolishing its repressive effect on *IL-6* mRNA whilst leaving miR-26b levels unchanged. Such a mechanism would bring miR-26 regulation in line with other observations of uridylylation-mediated degradation of mature small RNAs (Li *et al.* 2005; van Wolfswinkel *et al.* 2009; Kamminga *et al.* 2010; Kamminga *et al.* 2012) and would negate the requirement for a new, degradation-independent pathway of uridylylation repression of small RNA activity.

1.2.5 Uridylylation of pre-let-7 precursor miRNA

Lin28 (or the closely-related orthologues Lin28/Lin28a and Lin28b in mammals) is a highly conserved RNA binding protein first identified through study of a heterochronic mutant in *C. elegans* (Ambros and Horvitz 1984). Lin28 is essential for the regulation of developmental timing, maintenance of pluripotency and oncogenesis in several species (reviewed in Thornton and Gregory 2012), and its concurrent expression with three other factors (Sox2, Oct4 and Nanog) is sufficient to transform adult cells into induced pluripotent stem cells (iPSs; Yu *et al.* 2007). Many of the activities of Lin28 are mediated by its ability to suppress the biogenesis of the highly conserved *let-7* miRNA and consequently its ability to target multiple proliferative and developmental genes (Büssing *et al.* 2008). Whilst *let-7* is also tightly regulated transcriptionally (Van Wynsberghe *et al.* 2011), Lin28 specifically and directly interacts with intermediates of *let-7* biogenesis early in embryogenesis/differentiation and inhibits their processing, providing a second, post-transcriptional layer of *let-7* regulation.

Early observations (Heo *et al.* 2008) of the Lin28/*let-7* axis demonstrated that Lin28 binding to stem-loop precursor *let-7* (*pre-let-7*) was able both to inhibit its cleavage into mature miRNA and to promote its destabilisation. These dual activities coincided with, and were at least partly dependent on, the Lin28-dependent oligouridylylation of *pre-let-7* (Heo *et al.* 2008). ZCCHC11, a conserved mammalian ncPAP that also uridylylates mature miR-26a (Jones *et al.* 2009) and replication-dependent histone mRNAs (Schmidt *et al.* 2011), was found to interact specifically with the Lin28/*pre-let-7* binary complex via an N-terminal C2H2-type zinc finger (Heo *et al.* 2009; Thornton *et al.* 2012) and to uridylylate Lin28-bound *pre-let-7 in vitro* and *in vivo* (Heo *et al.* 2008; Hagan *et al.* 2009). This uridylylation was sufficient to repress mature *let-7* accumulation *in vivo* (Heo *et al.* 2008). A parallel pathway has been proposed in *C. elegans* wherein Lin28-dependent uridylylation of ectopically over-expressed *pre-let-7* is mediated by the ncPAP PUP-2 (Lehrbach *et al.* 2009), which has been previously shown to possess poly(U) polymerase activity *in vitro* (Kwak and Wickens 2007).

However, another group has been unable to detect a role for PUP-2 in the regulation of endogenous *pre-let-7*, demonstrating instead a dominant role for Lin28 as an inhibitor of processing of primary *let-7* transcripts (*pri-let-7*) into *pre-let*-7 by the endonuclease Drosha (Van Wynsberghe *et al.* 2011). Importantly, these differing activities of Lin28 are not mutually exclusive, and PUP-2-mediated *pre-let-7* uridylylation may represent a secondary pathway of Lin28 regulation acting on *pre-let-7* generated by residual Drosha activity. In mice, Lin28 is able to recruit both Zcchc11 and a closely-related ncPAP, Zcchc6, to uridylylate *pre-let*-7, suggesting redundancy of uridylyltransferases may be a conserved feature of this pathway (Thornton *et al.* 2012); similarly, the recent description of discrete *let-7* regulation activities for the two murine Lin28 orthologues (Piskounova *et al.* 2011) suggests that the regulatory landscape of the Lin28/*let-7* pathway may be more complicated than initially suspected.

In addition to the ZCCHC11/PUP-2-dependent pathway of *let-7* regulation, Lin28 has been demonstrated in other studies to exhibit additional, uridylylation-independent mechanisms of action in the *let-7* pathway. Interaction of Lin28 with *pri-let-7* at the stem-loop structure is able to inhibit Drosha cleavage in several organisms (Newman *et al.* 2008; Viswanathan *et al.* 2008; Piskounova *et al.* 2011; Van Wynsberghe *et al.* 2011), potentially by sequestering *pri-let-7* transcripts within the nucleolus (Piskounova *et al.* 2011). Furthermore, Lin28 is able to directly inhibit Dicer processing of *pre-let-7* in a mechanism independent of *pre-let-7* uridylylation (Rybak *et al.* 2008; Lightfoot *et al.* 2011). Structural (Nam *et al.* 2011; Mayr *et al.* 2012) and biochemical (Lightfoot *et al.* 2011; Desjardins *et al.* 2012) studies have suggested that this mechanism involves remodelling of the *pre-let-7* stem loop upon Lin28 binding, melting the double-stranded structure of the *pre-let-7* stem in the region of the Dicer cleavage site and thereby sterically inhibiting the binding of Dicer to *pre-let-7* and/or its ability to endonucleolytically cleave the *pre-let-7* substrate. Finally, a number of recent papers (Polesskaya *et al.* 2007; Xu *et al.* 2009; Xu and Huang 2009; Balzer *et al.* 2010; Qiu *et al.* 2010; Peng *et al.* 2011; Feng *et al.* 2012; Vadla *et al.* 2012) have demonstrated that Lin28 is able to interact directly with a wide variety of mRNAs via a *let-7*-independent Lin28-responsive element

frequently found in the coding region (Lei *et al.* 2012) and to thereby promote their translation directly via a pathway involving interaction with RNA helicase A (Qiu *et al.* 2010; Jin *et al.* 2011; Peng *et al.* 2011). These mRNAs include several known *let-7* targets (Lei *et al.* 2012; Vadla *et al.* 2012) as well as a wide variety of pluripotency, proliferation and metabolic factors, suggesting that Lin28-mediated translation of these factors may act as further method of counteracting the pro-differentiation and anti-proliferative activities of *let-7*. Though the full extent of Lin28-targeted mRNAs is not known, the recent identification of over 1,200 mRNAs in a deep sequencing study of Lin28 RNA immunoprecipitates suggests it may be an extensive mechanism by which Lin28 regulates proliferative and pluripotent properties of cells (Peng *et al.* 2011).

The detailed characterisation of a uridylylation-dependent pathway of *pre-let-7* processing inhibition and degradation suggests that the biogenesis of other miRNAs may be regulated similarly. Though a small number of non-*let-7* family pre-miRNAs have been identified which also exhibit Lin28-dependent destabilisation (Heo *et al.* 2009), this particular pathway seems largely specific to the regulation of *let-7* and its conserved roles in differentiation and pluripotency. However, a recent deep sequencing study of pre-miRNAs observed uridylylation on the majority of the pre-miRNAs tested, many of which were modified in a developmentally, temporally or spatially specific fashion (Newman *et al.* 2011). Furthermore, recent analyses of mature miRNA deep sequencing data sets have revealed a significant enrichment of uridylylation on miRNAs derived from the 3’ strand of their pre-miRNAs (whose 3’ ends would bear uridyl residues added to pre-miRNAs) over those derived from the 5’ strand, suggesting that some uridylylation initially believed to be specific to mature miRNA may in fact represent short uridylylation events on pre-miRNAs which were not sufficient to inhibit Dicer processing (Chiang *et al.* 2010). Together, these observations suggest that uridylylation of pre-miRNAs may be much more widespread than previously realised and may represent an hitherto unappreciated mechanism of regulation of small RNAs.

1.3 Non-degradative uridylylation

1.3.1 U insertion/deletion editing in Trypanosomal mitochondria

Though PUP activity had been observed previously in other organisms (Zabel *et al.* 1981; Andrews and Baltimore 1986), the first characterisation of an enzyme possessing PUP activity arose through study of the RNA editing machinery in the mitochondria of trypanosomatids. In this highly unusual system, termed U-insertion/deletion editing, mitochondrial mRNAs are transcribed in a non-functional, multi-frameshifted form which is processed into mature mRNA by the insertion and/or deletion of UMP residues at multiple points within the coding sequence (Aphasizhev and Aphasizheva 2008). The location and nature of these editing events is determined by separate mitochondrial transcripts, termed guide RNAs (gRNAs), which are complementary to the mature mRNA sequence (Blum *et al.* 1990; Blum and Simpson 1990). These gRNAs base-pair with the unedited mRNAs via both the encoded sequence and a 3' non-templated oligo(U) tail (Blum and Simpson 1990; McManus *et al.* 2000) and facilitate the activity of a large multiprotein complex, termed the 20S editosome, which catalyses sequential endonucleolysis, uridylyltransferase activity/exonucleolysis and ligation steps at sites of mismatch in the gRNA:mRNA duplex (Aphasizhev and Aphasizheva 2008).

Detailed proteomic analysis of editosome subunits in *Trypanosoma brucei* and *Leishmania tarentolae* identified a conserved 57 kDa protein exhibiting domain architecture typical of Pol β family nucleotidyltransferases (Aphasizhev *et al.* 2003; Panigrahi *et al.* 2003; Schnaufer *et al.* 2003). This protein, subsequently named TbMP57 or RET2 (RNA editing TUTase 2), possesses highly specific but poorly processive TUTase activity *in vitro* and *in vivo*, is capable of inserting a defined number of Us into a model editosome substrate containing a synthetic gRNA duplexed with a gapped mRNA and is essential for U-insertion editing (Aphasizhev *et al.* 2003; Aphasizhev *et al.* 2003; Ernst *et al.* 2003), supporting its role as the core uridylyltransferase of the editosome.

Recently, a more complex landscape of RNA editing has emerged with the identification of multiple editosome-like complexes in trypanosomal mitochondria (Panigrahi *et al.* 2006; Carnes *et al.* 2008). Coincident with these discoveries was the description of a second mitochondrial TUTase in *Trypanosoma brucei*, MEAT1 (mitochondrial editosome-like complex associated TUTase 1; Aphasizheva *et al.* 2009). Though MEAT1 possesses similar catalytic activity and substrate specificity to RET2 and is able to act on model editing substrates *in vitro*, it is found to reside in a novel editosome-like complex from which the U-insertion editing subcomplex containing RET2 is selectively excluded (Aphasizheva *et al.* 2009). The finding that *MEAT1(RNAi)* treatment does not abrogate RNA editing but results in a general stabilisation of mRNAs argues against a role for the MEAT1-containing complex as a secondary editosome; instead, MEAT1 may represent the catalytic component of a novel uridylylation-dependent degradative pathway. The observation of heteropolymeric poly(A/U) tails on numerous mitochondrial mRNAs in trypanosomes (Campbell *et al.* 1989; Etheridge *et al.* 2008) suggests an attractive model in which editing-like insertion of UMP residues by MEAT1 within the poly(A) tail of mRNAs acts as a degradative signal (Aphasizheva *et al.* 2009; Aphasizheva and Aphasizhev 2010). This process has been shown to require RET1, suggesting that this is the TUTase moiety of the poly(A/U) synthesis pathway (Aphasizheva and Aphasizhev 2010); however, it is possible that this effect is indirect, like the effect of RET1 on U-insertion editing by RET2, leaving open the possibility that MEAT1 may participate in this pathway.

Purification of TUTase activity observed in early studies in fractionated mitochondria (Bakalara *et al.* 1989) led to the identification of RET1 (RNA editing TUTase 1), a processive and highly specific TUTase (Aphasizhev *et al.* 2002; Aphasizhev *et al.* 2004). Early observations that RET1 was essential for U insertion/deletion editing led to the suggestion that it was the editosome uridylyltransferase (Aphasizhev *et al.* 2002). However several points of evidence including the exclusion of RET1 from the 20S editosome in several proteomic studies (Aphasizhev *et al.* 2003; Panigrahi *et al.* 2003; Schnaufer *et al.* 2003), the inhibitory effect of *RET1(RNAi)* on U-deletion as well as U-insertion editing (Aphasizhev *et al.* 2002) and the description of RET2 (Aphasizhev *et al.* 2003; Aphasizhev *et*

al. 2003; Ernst *et al.* 2003) suggested that the requirement of RET1 for RNA editing was likely to be indirect. Indeed, subsequent evidence that RET1 is required for the uridylylation of mature gRNAs (Aphasizhev *et al.* 2003; Ernst *et al.* 2003; Panigrahi *et al.* 2003), coupled with the previously described requirement for the gRNA oligo(U) tail for gRNA:mRNA duplex formation (Blum and Simpson 1990; McManus *et al.* 2000) confirmed that loss of RET1-mediated gRNA uridylylation was the likely mechanism of *RET1(RNAi)* RNA editing inhibition.

In addition to uridylylation of mature gRNA, a recent paper has suggested that RET1 represents a functionally diverse TUTase able to uridylylate a large number of different mitochondrial transcripts. Coordinated overexpression and RNAi studies demonstrated that RET1 was capable of uridylylating mature rRNA in addition to mature gRNAs (Adler *et al.* 1991; Aphasizheva and Aphasizhev 2010). However, despite observations that these RNAs exhibited greatly decreased accumulation upon depletion of *RET1*, this uridylylation had no stabilising effect on these RNAs; instead, the decreased accumulation surprisingly resulted from the inhibition of processing of these transcripts (and those of multiple mRNAs) from their polycistronic precursors (Aphasizheva and Aphasizhev 2010). Intriguingly, this paper identified long antisense transcripts exhibiting limited 5' end complementarity to gRNA precursor transcripts, and found that these antisense transcripts were stabilised in the presence of RET1, suggesting an hypothesis in which RET1 uridylylates and stabilises antisense transcripts required for the endonucleolytic processing of pre-gRNAs, and thereby indirectly facilitates gRNA maturation (Aphasizheva and Aphasizhev 2010). In sharp contrast to the results presented above, this paper demonstrated that the addition of homopolymeric oligo(U) tails to a subset of never-edited mitochondrial mRNAs resulted in the rapid degradation of these mRNAs by 3'→5' exonucleolysis (Aphasizheva and Aphasizhev 2010). These observations, whilst surprising, are supported by earlier biophysical analyses which described a UTP- and RET1-dependent rapid poly(A)$^+$ mRNA degradation pathway in fractionated trypanosomal mitochondria (Militello and Read 2000; Ryan and Read 2005) and suggest that the near-ubiquitous uridylylation of RNAs in trypanosomal mitochondria can be interpreted in multiple different fashions – stabilising,

destabilising or functional – depending on the nature of the tail and the transcript to which it is added.

In addition to the well-characterised TUTases of the trypanosomatid mitochondria, homology searches within the *Trypanosoma brucei* genome have identified a number of other candidate TUTases (Aphasizhev *et al.* 2004; Aphasizhev 2005). Though some of these enzymes have been subsequently assigned other functions as mitochondrial poly(A) polymerases (Kao and Read 2007; Etheridge *et al.* 2008), two, TUT3 and TUT4, have been shown to be specific and processive TUTases that reside in the cytoplasm rather than the mitochondria of trypanosomes (Aphasizhev *et al.* 2004; Stagno *et al.* 2007), reinforcing early observations of uridylyltransferase activity in whole-cell trypanosomal extracts (White and Borst 1987). Though the physiological function(s) of these enzymes is currently unknown, the widespread uridylylation in trypanosomal mitochondria and extensive uridylylation-based regulatory pathways in the cytoplasm of other species (reviewed above) suggests that these proteins may represent the catalytic components of hitherto-uncharacterised RNA metabolism pathways for trypanosomal nuclear-encoded RNAs.

1.3.2 mRNA polyuridylylation in dinoflagellate chloroplasts

Photosynthetic dinoflagellates, unicellular algae which represent major aquatic primary producers and the cause of toxic 'red tides', possess an unusual and evolutionarily highly divergent chloroplast. In addition to structural and physiological differences, dinoflagellate chloroplasts possess a unique genome organisation in which a highly reduced number of genes are encoded on unique, mostly uni- or digenic circular plasmids only a few kilobases in size (Zhang *et al.* 1999). These plasmids are transcribed by rolling-circle transcription generating long, polycistronic transcripts from which mature mRNAs are processed (Dang and Green 2010; Barbrook *et al.* 2012). Investigations into the complement of chloroplast transcripts led to the serendipitous discovery of ubiquitous poly(U) tails of 25-40 nt on chloroplast mRNAs in the dinoflagellate genera *Lingulodinium* and *Amphidinium* (Barbrook *et al.* 2012) and the species *Heterocapsa triquetra* (Nelson *et al.* 2007; Dang and Green

2009). In *Amphidinium*, these tails were added to all mature monocistronic and some precursor polycistronic mRNAs and were correlated with an intact 3' end of the mRNA (Barbrook *et al.* 2012), suggesting they may play a stabilising/protective role akin to that of polyadenylation on eukaryotic nuclear mRNAs (Proudfoot *et al.* 2002). Polyuridylylation was further restricted to a subset of plastid RNAs, being absent from both 23S rRNA (Dang and Green 2009) and plastid-encoded tRNAs (Nelson *et al.* 2007) and preceded substitutional editing of mRNA/rRNA sequences in *H. triquetra* (Dang and Green 2009), suggesting it may represent a key marker for maturation and downstream processing. Recent observation of a similar polyuridylylation mechanism in the distantly related photosynthetic apicomplexan *Chromera velia* suggests that mRNA polyuridylylation may represent an evolutionarily ancient mechanism of plastid RNA regulation (Janouškovec *et al.* 2010).

The identity of the uridylylating enzyme in dinoflagellates and *Chromera velia* is currently unknown. The chloroplasts of higher plants utilise oligoadenylation to promote degradation of mature RNAs and/or degradation intermediates (Lange *et al.* 2009). This modification is not added by a poly(A) polymerase type enzyme, but instead requires the reversible activity of the polynucleotide phosphorylase (PNPase) in a mechanism apparently conserved in the putative endosymbiont precursors of chloroplasts, cyanobacteria (Rott *et al.* 2003; Slomovic *et al.* 2008). Though organellar uridylylation is well described in trypanosomatids where it is mediated by multiple ncPAP family members (see **1.3.1** above), dinoflagellates lack uridylyl insertion/deletion editing. Furthermore, bioinformatic interrogation of the dinoflagellate genome with diverse ncPAP family members fails to identify any homologues (BLAST and PSI-BLAST interrogation of *Amphidinium* and *Heterocapsa* genera using *S. pombe*, *S. cerevisiae*, *T. brucei* and *A. thaliana* ncPAP family members; personal observations) , though it is noted that ORF annotation in these genomes is far from complete. It is thus concluded that, while the possibility of a ncPAP family member in dinoflagellates remains, it appears equally possible that the poly(U) polymerase activity of dinoflagellate chloroplasts is mediated by a divergent PNPase or prokaryotic-type PAP gene with selective PUP rather than PAP activity (Wang and Morse 2006).

1.3.3 Maintenance of the U6 snRNA 3' end

U6 snRNA is a highly conserved non-coding RNA (ncRNA) which, with four other snRNAs (U1, U2, U4 and U5) and approximately 60 proteins comprises the spliceosome (reviewed most recently in Will and Lührmann 2011). U6 snRNA exhibits a number of features that distinguish it from other U-class snRNAs, including transcription by RNA Pol III rather than Pol II and modification with a different methylated triphosphate cap structure (Kunkel *et al.* 1986; Reddy *et al.* 1987). U6 snRNA is also modified at its 3' end, with the predominant mature form possessing exactly five terminal UMP residues (one more than the templated U_4 tail) and ending with an unusual 2',3' cyclic phosphate (Lund and Dahlberg 1992; Terns *et al.* 1992). However, a minor population exhibits heterogeneous 3'-OH ends containing oligouridylylate tails both shorter and longer than the templated U_4 (Rinke and Steitz 1985; Reddy *et al.* 1987; Hirai *et al.* 1988; Terns *et al.* 1992; Tazi *et al.* 1993). Though several studies have attributed this 3' heterogeneity to the opposing actions of a U6 snRNA-specific uridylyltransferase(s) (Reddy *et al.* 1987; Hirai *et al.* 1988; Tazi *et al.* 1993) and exonuclease (Booth and Pugh 1997), these studies were complicated by the overlapping activity of two separate poly(U) polymerase activities in the cell lysates tested, one of which non-specifically monouridylylated a wide range of cellular RNAs whilst the other exhibited highly specific mono-/oligo-uridylylation activity on truncated U6 snRNAs (Trippe *et al.* 1998). Purification and characterisation of this U6-specific uridylyltransferase activity identified a predominantly nucleolar ncPAP-family enzyme now named U6 TUTase (Trippe *et al.* 2006). This enzyme is highly specific both for its template, being able to distinguish between U6 snRNA and the highly-related minor spliceosomal component U6atac snRNA (Trippe *et al.* 2006), and for the reaction catalysed, which adds 1, 2 or 3 nts to different 3' terminally truncated U6 snRNAs in such a way as to specifically restore the templated U_4 tail (Trippe *et al.* 2003). This highly specific restorative activity of U6 TUTase suggests that the elongated oligo(U) tails on U6 snRNA described in early studies (Reddy *et al.* 1987; Tazi *et al.* 1993) may be due to the coincident activity of the non-specific TUTase separately identified by Trippe *et al*. (1998).

The fact that the highly-specific U6 TUTase is an essential gene suggests that this restorative uridylylation of U6 snRNA is vital for cell viability, and may represent a mechanism for the formation of mature U6 snRNA, or rescue and recycling of 3'-degraded forms. Alternatively, an intact 3' terminal may be essential for catalytic activity and/or structural features of an active spliceosome (Trippe *et al.* 2006). Uridylylation of snRNA has been reported in other species including mice and *Xenopus* (Lund and Dahlberg 1992; Terns *et al.* 1992), and U6 snRNAs from diverse species (including *C. elegans* and *Drosophila*) are efficient uridylylation targets of human U6 TUTase (Trippe *et al.* 1998). However, U6 TUTase orthologues have only been found in vertebrate species (Trippe *et al.* 2006; Figure 1.1, Table 1.1), suggesting that other species such as *C. elegans*, *Drosophila* and *S. pombe* either utilise non-uridylylation methods to protect and/or recycle 3' ends of U6 snRNAs or, alternatively, may utilise a evolutionarily divergent uridylyltransferase to recapitulate U6 snRNA uridylylation (Trippe *et al.* 1998; Trippe *et al.* 2006).

A possible alternative explanation for the essentiality of U6 TUTase was provided by another group, who described an entirely independent role for U6 TUTase (in this context alternatively named Star-PAP, nuclear speckle targeted PIPKIα-regulated poly(A) polymerase) wherein Star-PAP was required for the cotranscriptional cleavage and polyadenylation of a large set of pre-mRNAs in place of the canonical PAP, PAPα (Mellman *et al.* 2008). Subsequent work proposed that, unlike the well-characterised mechanism of canonical 3' end processing (Chan *et al.* 2011), Star-PAP interacts with a *cis*-regulatory element in target mRNA 3' UTRs in order to recruit CPSF subunits to the cleavage site, allowing cleavage and subsequent polyadenylation (Laishram and Anderson 2010). This activity was further suggested to be regulated by Star-PAP phosphorylation by PIPKIα-directed phosphatidylinositol signalling via a range of different cellular kinases (Gonzales *et al.* 2008; Laishram *et al.* 2011; Li *et al.* 2012). Given the well characterised and highly specific activity of U6 TUTase on U6 snRNA, it would be extremely surprising for a role so divergent in target, activity, partners and regulation to be conserved in the same protein. Furthermore, the functional interplay and direct physical interaction between splicing and 3' maturation machinery (Martinson 2011)

suggests that at least some of the observations made in support of the Star-PAP pathway may be attributable to secondary effects of a failure of U6 snRNA maturation/recycling and splicing. However, given the proposed dual adenylation/uridylylation activity of Mut68 in Chlamydomonas (Ibrahim *et al.* 2006; Ibrahim *et al.* 2010) and the dual polyadenylation/polyuridylylation activity of *S. pombe* Cid1 *in vitro* (Rissland *et al.* 2007), it is conceivable that U6 TUTase could simultaneously possess these two highly divergent activities.

1.3.4 Intramolecular loop-mediated priming of Tetrahymena *RdRP*

The unicellular ciliate *Tetrahymena thermophila* expresses a variety of different sRNA populations including 23-24 nt sRNAs derived from pseudogenes, repeat loci, protein-coding ORFs and hairpin loci in vegetative growth as well as a conjugation-specific population of 27-30 nt scan RNAs (scnRNAs) that direct essential heterochromatin formation and DNA elimination in the developing macronucleus (Lee and Collins 2006; Couvillion *et al.* 2009; Kataoka and Mochizuki 2011). These sRNAs are loaded specifically onto one or more of the eight functional Twi proteins (Twi1-2,7-12p) encoded by the *Tetrahymena* macronuclear genome (Couvillion *et al.* 2009) and most are synthesised by the single essential RdRP in the *Tetrahymena* genome, Rdr1p (Lee and Collins 2007). Rdr1p achieves this broad target activity through different RDRCs containing one of two mutually exclusive ncPAPs, Rdn1p or Rdn2p (Rdr1-associated nucleotidyltransferase), both of which possess specific and processive poly(U) polymerase activity *in vitro* (Lee *et al.* 2009). Rdn1p RDRCs additionally contain one of two mutually exclusive proteins of unknown function, Rdf1p or Rdf2p (Rdr1-associated factor; Lee and Collins 2007; Lee *et al.* 2009). The RDRCs further associate with vegetatively expressed Dcr2p in a manner that facilitates its endonucleolytic activity (Lee and Collins 2007). Rdn1/2p and Rdf1/2p show different expression profiles and mutants show phenotypic differences, with Rdn1 in particular being essential for viability. Deletion of each of the non-essential RDRC components depletes the levels of particular subsets of the 23-24 nt sRNA population, confirming functional specification of the different RDRCs (Lee *et al.* 2009).

Surprisingly, *Tetrahymena* Rdr1p mediates dsRNA synthesis by a novel mechanism in which the 3' end of the ssRNA template forms an intramolecular loop which primes the second-strand synthesis of dsRNA (Talsky and Collins 2010). In this context, Rdn1/2p were found to synthesise a short oligo(U) tail on the 3' end of the ssRNA template which subsequently forms the intramolecular loop, allowing priming of dsRNA synthesis from the extreme 3' end of the template rather than from an internal residue. This uridylylation-mediated looping alters the sRNA repertoire after Dcr2p endonucleolysis, providing additional species from the 3' end of the template sequence (Talsky and Collins 2010). Though it seems surprising that the lethality of Rdn1p deletion could be explained by the relatively modest fluctuation in sRNA repertoire proposed by this model, Rdn1p may act coordinately with Rdr1p (Talsky and Collins 2010) or may otherwise modulate Rdr1p substrates/activity in such a way as to make its activity essential.

Alternatively, Rdn1p may have additional functions outside of its role in the RDRC. Indeed, a recent paper has demonstrated that Rdn1p, in complex with Rdf1/2p, is able to form an RDRC-independent complex (RSPC) with an uncharacterised protein named Rsp1 (RNA silencing protein 1; Talsky and Collins 2012). In this complex the activity of Rdn1p on substrates efficiently uridylylated in the context of the RDRCs is approximately 1000-fold reduced, suggesting that RSPCs may have poor uridylyltransferase activity (Talsky and Collins 2012), or alternatively may act on different substrate RNAs to the RDRCs. A deletion mutant of *RSP1* is viable for vegetative growth but, surprisingly, shows near-complete loss of all sRNAs derived from RDRC activity (Talsky and Collins 2012). Given the observation that all classes of sRNA in *Tetrahymena* not protected by Hen1 methylation show frequent mono- or oligo-uridylylation (Lee and Collins 2006; Couvillion *et al.* 2009), it is tempting to speculate that RSPCs act on sRNA species downstream of Dcr2p; however, the apparent stabilising role that uridylylation would play in this context would be contrary to numerous other observations of sRNA uridylylation (see above). Additionally, other possible roles for the RSPC, such as in mRNA substrate selection or Rdn1p/Rdf complex assembly, cannot be discounted (Talsky and Collins 2012).

1.3.5 nucleotidyl-protein priming of viral RNA genome synthesis

Poliovirus, as well as other member of the *picornaviridae* family, have an monopartite, ssRNA, plus-strand genome with a 3' poly(A) tail (Yogo and Wimmer 1972) and, at its 5' end, an unusual covalent linkage to a small virally-encoded protein, VPg (virion protein genome-linked; Flanegan *et al.* 1977; Lee *et al.* 1977). Early observations suggested that a host terminal uridylyltransferase could uridylylate the poly(A)$^{+}$ 3' end of the genome, generating a hairpin capable of priming RNA synthesis (Andrews and Baltimore 1986). However, observation of unusual nucleotidyl-protein forms of VPg carrying diuridylylation of a conserved tyrosine side-chain (Crawford and Baltimore 1983; Takegami *et al.* 1983) led to the discovery that uridylylated VPg is able to base-pair with the 3' ends of both plus- and minus-strand genomes and prime genome synthesis by the viral RNA polymerase, 3D^{pol} (Paul *et al.* 1998; Sharma *et al.* 2005; Steil and Barton 2008). The uridylylation of VPg is catalysed by 3D^{pol} and is templated by a conserved stem-loop structure in the coding region of the poliovirus genome, the *cis* regulatory element (*cre*, also termed *oril*; Goodfellow *et al.* 2000; Paul *et al.* 2000; Paul *et al.* 2003). Uridylylation of VPg orthologues and other small proteins prior to protein-dependent genome replication priming has been observed in other pathologically important members of the *picornaviridae* family including foot-and-mouth disease virus and human rhinovirus 2 (Gerber *et al.* 2001; Nayak *et al.* 2005) as well as in the *caliciviridae* (Machín *et al.* 2001; Rohayem *et al.* 2006) and *potyviridae* families (Puustinen and Mäkinen 2004; Anindya *et al.* 2005).

1.4 The non-canonical PAP family

Despite their myriad functional and biochemical activities described above, phylogenetic analysis has identified conserved structural and primary sequence features that place all currently identified eukaryotic uridylyltransferases in a common evolutionary family, termed the non-canonical poly(A) polymerases (ncPAPs; Stevenson and Norbury 2006). The ncPAP family falls within the larger DNA polymerase β superfamily of nucleotidyltransferases, a diverse clade encompassing genes from all three kingdoms of life, capable of modifying targets including DNA, RNA, proteins and

aminoglycoside antibiotics (Holm and Sander 1995; Aravind and Koonin 1999). In addition to the pol β-type nucleotidyltransferase domain (NTD) containing the defining core motif ΦG[GS]-X_{9-13}-DΦ[DE]Φ (where Φ represents a hydrophobic residue) common to all members of the pol β superfamily, ncPAPs are defined by the presence of a second poly(A) polymerase/2'-5' oligo(A) synthetase (PAP/25A) domain shared with 2'-5' oligo(A) synthetases and archaeal CCA-adding enzymes (Aravind and Koonin 1999; Stevenson and Norbury 2006). Though some members of the ncPAP family possess motifs capable of binding nucleic acids, some of which are essential for activity (Aphasizhev *et al.* 2004; Aphasizheva *et al.* 2004; Thornton *et al.* 2012), it is notable that most ncPAPs possess no definable nucleic acid recognition motif. Many achieve target recognition through functional partnering with professional nucleic acid recognition proteins (Wang *et al.* 2002; Aphasizhev *et al.* 2003; Ernst *et al.* 2003; LaCava *et al.* 2005; Vaňáčová *et al.* 2005; Wyers *et al.* 2005), forming heterooligomeric functional complexes; however, the recent description of direct RNA binding by Cid1 through a positively-charged surface (Yates *et al.* 2012) suggests that monomeric members of this family lacking readily identifiable interaction domains may still be able to bind RNA targets.

The first characterised member of the ncPAP family was *S. cerevisiae* Trf4p (topoisomerase I-related function 4), identified in 1995 via a genetic screen for synthetic-lethal genetic interactors of a topoisomerase I mutant (Sadoff *et al.* 1995) and subsequently shown to regulate the condensation of chromosomes and their efficient segregation during mitosis (Castaño *et al.* 1996a; Castaño *et al.* 1996b). Early reports suggested that Trf4p and its paralogue Trf5p, and by extension the ncPAP family, may represent Pol β-like DNA polymerases (Aravind and Koonin 1999; Wang *et al.* 2000). However, the parallel identification of *C. elegans* GLD-2 and *S. pombe* Cid13 as cytoplasmic RNA PAPs (Read *et al.* 2002; Wang *et al.* 2002) and trypanosomal RET1 as a mitochondrial RNA PUP (Aphasizhev *et al.* 2002), as well as the subsequent redefinition of Trf4p as a nuclear RNA PAP (Saitoh *et al.* 2002; LaCava *et al.* 2005; Vaňáčová *et al.* 2005; Wyers *et al.* 2005) defined the ncPAPs

as a widely conserved and functionally varied family of post-transcriptional RNA nucleotidyltransferases (see Figure 1.1, Table 1.1).

Though the number of ncPAP family members encoded in the genome varies widely between species, from two in *S. cerevisiae* to at least 12 in *C. elegans* and *T. brucei* (see Table 1.1), most species investigated possess several family members with differing cellular functions; indeed, *S. cerevisiae*, in which the only two family members encode highly overlapping but non-identical functions (San Paolo *et al.* 2009), is something of an anomaly. The fission yeast *Schizosaccharomyces pombe*, upon which this study is based, encodes six different family members, numbered *cid1, cid11-14* and *cid16*, with varying catalytic activities and functions (Aravind and Koonin 1999; Wang *et al.* 2000). Though the functions of *cid1* have been extensively characterised (discussed in **1.1.2** above), the state of knowledge regarding other ncPAP family members in *S. pombe* varies widely.

1.4.1 Cid14, a Tr4p/Trf5p homologue

Despite early identification as a DNA polymerase (Wang *et al.* 2000), *S. cerevisiae* Trf4p, and later its paralogue Trf5p, were soon shown to possess no DNA polymerase activity *in vitro*, instead exhibiting robust, RNA-specific adenylyltransferase activity (Saitoh *et al.* 2002; Kadaba *et al.* 2004; LaCava *et al.* 2005; Vaňáčová *et al.* 2005; Wyers *et al.* 2005; Houseley and Tollervey 2006). Trf4p and Trf5p represent the catalytic subunits of conserved nuclear post-transcriptional polyadenylation complexes, TRAMP and TRAMP5 respectively, which also contain an RNA helicase, Mtr4p, and one of two zinc finger-containing proteins, Air1p/Air2p (LaCava *et al.* 2005; Vaňáčová *et al.* 2005; Wyers *et al.* 2005; Houseley and Tollervey 2006). As part of these complexes, Trf4p/Trf5p oligoadenylate large but non-overlapping populations of aberrant, unprocessed or nonfunctional nuclear RNAs including hypomethylated $tRNAi_{Met}$ (Kadaba *et al.* 2004; Vaňáčová *et al.* 2005), cryptic Pol II transcripts (Wyers *et al.* 2005), rRNA and snoRNA precursors (LaCava *et al.* 2005), spliced-out introns

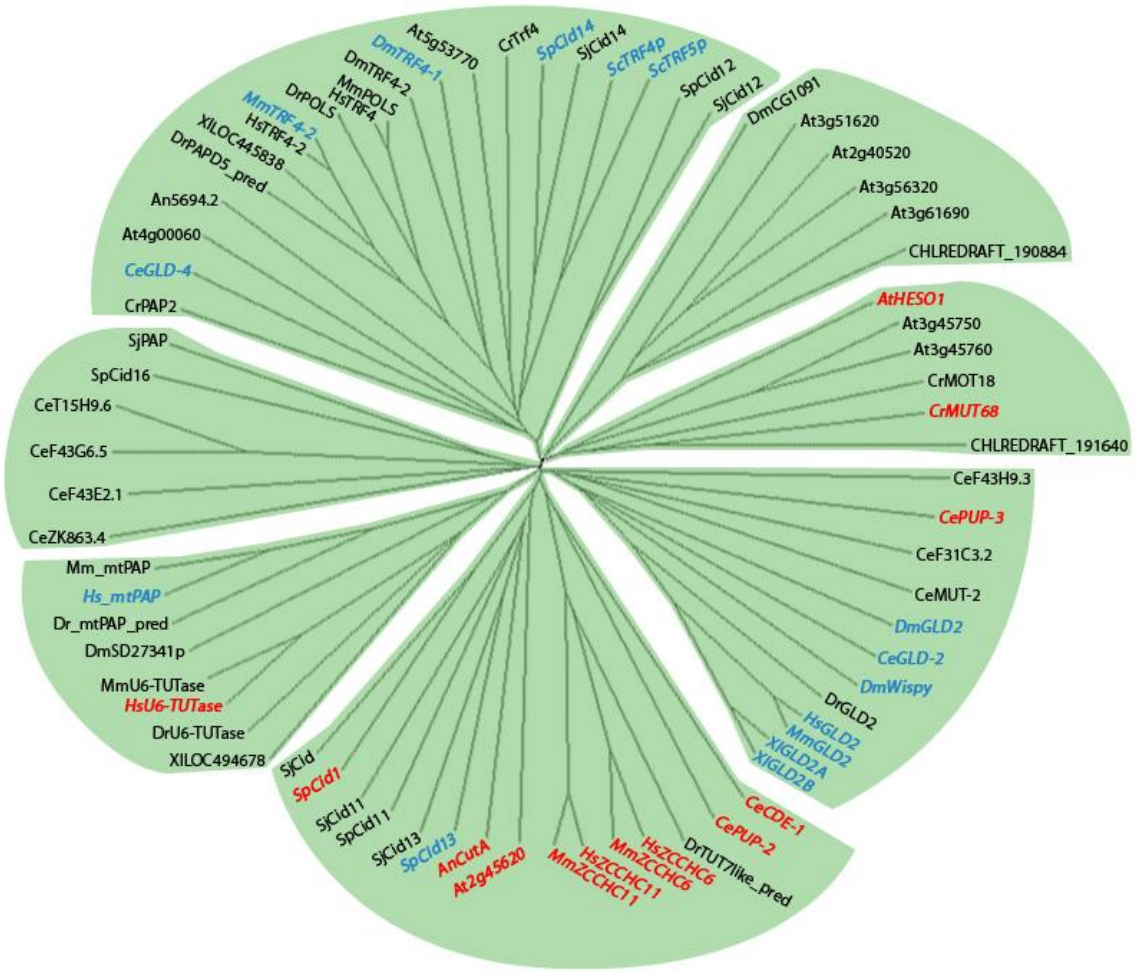


Figure 1.1: Phylogenetic tree of ncPAPs from species with known degradative uridylylation activity. ncPAP family members were identified by BLAST searches using SpCid1, HsZCCHC11, AtHESO1 and ScTrf4p as baits with a significance threshold of p<0.0001 or by annotation in the literature. Full-length protein sequences were aligned with CLUSTALΩ (Sievers *et al.* 2011) and phylogenetic tree compiled with the CLUSTALW2 Phylogeny tool (Larkin *et al.* 2007). Tree was visualised using TREEVIEW (Page 1996). DrTUT4 and CHLREDRAFT_206054 were excluded from the tree due to the lack of the canonical Pol β superfamily motif ΦG[G/S]-$X_{9\text{-}13}$-DΦ[D/E]Φ (where Φ = hydrophobic residue). Red titles represent species with known in vitro and/or in vivo PUP activity, whilst Blue titles represent species with PAP activity. Though *S. cerevisiae* does not encode uridylyltransferases, ScTrf4p and ScTrf5p are included for comparison. Species from which ncPAP sequences are retrieved are indicated by a two-letter prefix as follows: An: *Aspergillus nidulans*; At: *Arabidopsis thaliana*; Ce: *Caenorhabditis elegans*; Cr/CHLRE: *Chlamydomonas reinhardtii*; Dm: *Drosophila melanogaster*; Dr: *Danio rerio*; Hs: *Homo sapiens*; Mm: *Mus musculus*; Sc: *Saccharomyces cerevisiae*; Sj: *Schizosccharomyces japonicus*; *Sp*: *Schizosaccharomyces pombe*; Xl: *Xenopus laevis*.

Table 1.1: Compiled List of ncPAPs from species with characterised ncPAP activity. Species were identified as described in Figure 1.1 and properties determined from the literature where applicable. Italicised and emboldened text indicates ncPAP family members in which the canonical Pol β superfamily motif ΦG[G/S]-$X_{9\text{-}13}$-DΦ[D/E]Φ (where Φ = hydrophobic residue) is mutated or absent.

Name	Alternative Name	Activity	Effect	Target	References
Human					
GLD2	TUTase2, PAPD4, Hs1	adenylation	stabilisation, translation	mRNA	(Kwak *et al.* 2004; Rouhana *et al.* 2005; Kwak and Wickens 2007; Katoh *et al.* 2009; Glahder and Norrild 2011)
mtPAP	TUTase1, PAPD1, Hs4	adenylation	stabilisation	mitochondrial mRNAs	(Tomecki *et al.* 2004; Nagaike *et al.* 2005)
TRF4-2	TUTase3, PAPD5	adenylation	destabilising	rRNA, snoRNA precursors	(Shcherbik *et al.* 2010; Rammelt *et al.* 2011; Berndt *et al.* 2012)
U6 TUTase	TUTase6, PAPD2, Hs5, TUT1, RBM21	uridylylation	stabilising	U6 snRNA	(Trippe *et al.* 2003; Trippe *et al.* 2006)
	Star-PAP	adenylation?	maturation?	pre-mRNA (cotranscriptional)	(Mellman *et al.* 2008; Mellman and Anderson 2009)
ZCCHC11	TUTase4, PAPD3, Hs3	uridylylation	destabilising	miRNA, pre-miRNA, histone mRNA	(Kwak and Wickens 2007; Heo *et al.* 2009; Jones *et al.* 2009; Schmidt *et al.* 2011)
ZCCHC6	TUTase7, PAPD6, Hs2	uridylylation	?	?	(Kwak and Wickens 2007; Rissland *et al.* 2007)
TRF4	TUTase5, PAPD7, POLσ	?	?	?	N/A
Mouse					
GLD2	TUTase2, PAPD4, Mm1	adenylation	stabilisation, translation	mRNA	(Kwak *et al.* 2004; Shin *et al.* 2004; Rouhana *et al.* 2005; Nakanishi *et al.* 2006; Kwak and Wickens 2007; Rouhana and Wickens 2007; Kashiwabara *et al.* 2008; Katoh *et al.* 2009)
mtPAP	TUTase1, PAPD1, Mm3	?	?	?	N/A
TRF4-2	TUTase3, PAPD5	?	?	?	N/A
U6 TUTase	TUTase6, PAPD2, TUT1, Star-PAP	?	?	?	N/A
ZCCHC11	TUTase4, PAPD3	uridylylation	destabilising	miRNA, pre-miRNA	(Hagan *et al.* 2009; Heo *et al.* 2009; Jones *et al.* 2009; Thornton *et al.* 2012)
ZCCHC6	TUTase7, PAPD6, Mm2	uridylylation	destabilising	pre-miRNA	(Thornton *et al.* 2012)
POLσ	TUTase5, PAPD7, TRF4	?	?	?	N/A
Zebrafish (Danio rerio)					
GLD2	TUTase2, PAPD4	?	?	?	N/A
mtPAP	TUTase1, PAPD1	?	?	?	N/A
TRF4-2	TUTase3, PAPD5	?	?	?	N/A
U6 TUTase	TUTase6, PAPD2, TUT1, Star-PAP	?	?	?	N/A
ZCCHC11	***TUTase4, PAPD3***	***?***	***?***	***?***	***N/A***
ZCCHC6	TUTase7, PAPD6	?	?	?	N/A
POLσ	TUTase5, PAPD7	?	?	?	N/A
Xenopus laevis					
GLD2A		adenylation	translation	developmental mRNAs	(Barnard *et al.* 2004; Rouhana *et al.* 2005; Kim and Richter 2006; Radford *et al.* 2008)
GLD2B		adenylation	translation	developmental mRNAs	(Rouhana *et al.* 2005)
LOC445838		?	?	?	N/A
LOC494678		?	?	?	N/A

Name	Alternative Name	Activity	Effect	Target	References
Drosophila melanogaster					
TRF4-2		none?	?	?	(Nakamura *et al.* 2008)
SD27341p		?	?	?	N/A
GLD2	Dm1	adenylation	translation	neuronal mRNAs	(Kwak *et al.* 2004; Kwak *et al.* 2008)
Wispy	Dm2	adenylation	translation	oocyte mRNAs	(Kwak *et al.* 2004; Benoit *et al.* 2008)
TRF4-1		adenylation	destabilising	snRNA	(Nakamura *et al.* 2008)
CG1091		?	?	?	N/A
Caenorhabditis elegans					
CDE-1	CID-1, PUP-1, K10D2.3, Ce5	uridylylation	destabilising	CSR-1 pathway siRNAs	(Olsen *et al.* 2006; Kwak and Wickens 2007; van Wolfswinkel *et al.* 2009)
PUP-2	K10D2.2, Ce3	uridylylation	destabilising	pre-miRNA	(Kwak and Wickens 2007; Lehrbach *et al.* 2009; Van Wynsberghe *et al.* 2011)
PUP-3	CDE-2, Ce4, F59A3.9	uridylylation	?	?	(Kwak and Wickens 2007)
MUT-2	RDE-3, Ce6, K04F10.6	?	?	WAGO-1 pathway siRNAs?	(Kwak and Wickens 2007; Gu *et al.* 2009)
GLD-4	ZK858.1	adenylation	translation, stabilisation	oocyte mRNAs	(Schmid *et al.* 2009)
GLD-2	Ce1, ZC308.1	adenylation	translation, stabilisation	oocyte mRNAs	(Wang *et al.* 2002; Kwak *et al.* 2004; Suh *et al.* 2006; Kwak and Wickens 2007; Kim *et al.* 2010)
F31C3.2		none?	?	?	(Kwak and Wickens 2007)
ZK863.4		?	?	?	N/A
F43H9.3		?	?	?	N/A
F43E2.1		?	?	?	N/A
F43G6.5		?	?	?	N/A
T15H9.6		?	?	?	N/A
Aspergillus nidulans					
CutA	AN7748.2	cytidylation/ uridylylation	destabilising	mRNA	(Morozov *et al.* 2010; Morozov *et al.* 2010)
AN5964.2		?	?	?	N/A
Schizosaccharomyces pombe/Schizosaccharomyces japonicus					
Cid1	Sp2, Cid (*S. japonicus*)	uridylylation	destabilising	mRNA	(Wang *et al.* 2000; Rissland *et al.* 2007; Rissland and Norbury 2009)
Cid11	Sp1	?	?	?	N/A
Cid12		?	?	siRNAs/siRNA targets?	(Motamedi *et al.* 2004; Win *et al.* 2006)
Cid13	Sp3	adenylation	stabilising?	suc22 mRNA?	(Read *et al.* 2002; Saitoh *et al.* 2002)
Cid14		adenylation	destabilising	rRNA, pre-mRNA, heterochromatic transcripts, mRNA	(Win *et al.* 2006; Buhler *et al.* 2007; Buhler *et al.* 2008; Wang *et al.* 2008; Lemieux *et al.* 2011)
Cid16	PAP (*S. japonicus*)	uridylylation	destabilising	mRNA	This study
Saccharomyces cerevisiae					
Trf4p	POLσ, POLκ	adenylation	destabilising	aberrant tRNA, pre-snoRNA, pre-mRNA, intergenic RNA, CUTs	(Kadaba *et al.* 2004; LaCava *et al.* 2005; Vaňáčová *et al.* 2005; Wyers *et al.* 2005; Houseley *et al.* 2007; Kwak and Wickens 2007; Grzechnik and Kufel 2008; Roth *et al.* 2009; Dickinson *et al.* 2012)
Trf5p		adenylation	destabilising	pre-rRNA	(Kwak *et al.* 2004; Houseley and Tollervey 2006; Kwak and Wickens 2007; San Paolo *et al.* 2009; Wery *et al.* 2009)

Name	Alternative Name	Activity	Effect	Target	References
Arabidopsis thaliana					
HESO1	At2, At2g39740	uridylylation	destabilising	miRNAs, siRNAs	(Kwak and Wickens 2007; Ren *et al.* 2012; Zhao *et al.* 2012)
At2g40520		?	?	?	N/A
At2g45620	At1	uridylylation	?	?	(Kwak and Wickens 2007)
At3g45750		?	?	?	N/A
At3g45760		?	?	?	N/A
At3g51620		?	?	?	N/A
At3g56320		?	?	?	N/A
At3g61690		?	?	?	N/A
At4g00060		?	?	?	N/A
At5g53770		?	?	?	N/A
Chlamydomonas reinhardtii					
Trf4	Chlredraft_17066	?	?	?	N/A
PAP2	Chlredraft_143498	?	?	?	N/A
MOT18	Chlredraft_177575	?	?	?	N/A
MUT68	pol β NTase, Chlredraft_149294	uridylylation adenylation	destabilising destabilising	miRNAs/siRNAs miRNA cleavage 5' products	(Ibrahim *et al.* 2010) (Ibrahim *et al.* 2006)
Chlredraft_206054		***?***	***?***	***?***	***N/A***
Chlredraft_191640		?	?	?	N/A
Chlredraft_190884		?	?	?	N/A
Tetrahymena thermophila					
Rdn1p	TTHERM_00094000	uridylylation	sRNA biogenesis	RDRC, RSPC template mRNAs;	(Lee *et al.* 2009; Talsky and Collins 2010; Talsky and Collins 2012)
Rdn2p	TTHERM_00094040	uridylylation	sRNA biogenesis	RDRC template mRNAs	(Lee *et al.* 2009; Talsky and Collins 2010)
TTHERM_00194830		?	?	?	N/A
TTHERM_00569420		?	?	?	N/A
TTHERM_01050360		?	?	?	N/A
TTHERM_01251260		?	?	?	N/A
TTHERM_01394310		?	?	?	N/A
Trypanosoma brucei/Leishmania tarentolae					
mtPAP1	KPAP1, TUT5	adenylation	stabilising?	mitochondrial mRNAs	(Etheridge *et al.* 2008)
mtPAP2 (put)	KPAP2, TUT6	adenylation?	stabilising?	mitochondrial mRNAs	(Kao and Read 2007)
RET1	KRET1, TUTase, TUT1	uridylylation	stabilising, destabilising	guide RNAs, antisense RNA, mitochondrial mRNA, rRNA	(Aphasizhev *et al.* 2002; Aphasizhev *et al.* 2003; Aphasizheva *et al.* 2004; Aphasizheva and Aphasizhev 2010)
RET2	TbMP57, TUT2	uridylylation	RNA editing	mRNA insertion sites	(Aphasizhev *et al.* 2003; Aphasizhev *et al.* 2003; Ernst *et al.* 2003; Panigrahi *et al.* 2003; Schnaufer *et al.* 2003)
MEAT1	TUT7	uridylylation	?	mRNA insertion sites?	(Aphasizheva *et al.* 2009; Stagno *et al.* 2010)
TUT3		uridylylation	?	?	(Aphasizhev *et al.* 2004)
TUT4		uridylylation	?	?	(Stagno *et al.* 2007)
ncPAP1	POLS	adenylation	destabilising?	nuclear RNAs?	(Etheridge *et al.* 2009)
ncPAP2	ncPAP1	adenylation	?	nuclear RNAs?	(Etheridge *et al.* 2009)
CBH14343		?	?	?	N/A
CBH14926		?	?	?	N/A
CBH18636		?	?	?	N/A

and antisense transcripts (San Paolo *et al.* 2009). This oligoadenylation provides a free 3' end that facilitates the rapid degradation of these sometimes highly structured transcripts by the nuclear exosome (LaCava *et al.* 2005; Vaňáčová *et al.* 2005; Wyers *et al.* 2005), demonstrating that Trf4p/Trf5p represent the key catalytic PAPs of a general nuclear RNA surveillance pathway.

Cid14 represents the closest relative of *S. cerevisiae* Trf4p/Trf5p in *S. pombe* and exhibits similar poly(A) polymerase activity as a recombinant protein *in vitro* (Win *et al.* 2006; Buhler *et al.* 2007). Cid14 resides in a TRAMP-like complex *in vivo* with orthologues of *S. cerevisiae* TRAMP components Air1p and Mtr4p and is able to rescue the lethality of a *trf4Δ trf5Δ* double mutant, confirming its status as a true functional homologue of these genes (Win *et al.* 2006; Buhler *et al.* 2007). Candidate and genome-wide expression studies identify a range of non-coding RNAs as substrates of a Cid14-mediated nuclear RNA surveillance pathway including rRNA precursors, euchromatic antisense transcripts and transcripts from heterochromatically silenced loci (Win *et al.* 2006; Buhler *et al.* 2007; Wang *et al.* 2008; Reyes-Turcu *et al.* 2011; Zhang *et al.* 2011), placing Cid14 at the centre of a nuclear RNA surveillance pathway closely resembling that of Trf4p/Trf5p. Recently, new connections have emerged between the nuclear RNA surveillance activity of Cid14-TRAMP and the regulation of heterochromatin formation and gene silencing, with *cid14* being required both for efficient heterochromatic gene silencing and, paradoxically, the suppression of an RNAi-independent pathway of heterochromatin nucleation at the centromeres (Buhler *et al.* 2007; Buhler *et al.* 2008; Wang *et al.* 2008; Reyes-Turcu *et al.* 2011; Zhang *et al.* 2011). The ability of Cid14-TRAMP to interact with Mlo3, a mRNP assembly and export protein and functional partner of Clr4 and the RITS complex (Zhang *et al.* 2011), has led to the proposal that Mlo3 may represent a 'gate-keeper' for ncRNA metabolism, diverting transcripts either into the RNAi pathway via RITS or the exosome-mediated degradation pathway via Cid14 (Zhang *et al.* 2011). Consistent with this hypothesis, deletion of *cid14* led to a marked decrease in levels of centromere-targeted siRNAs in several studies (Buhler *et al.* 2007; Reyes-Turcu *et al.* 2011; Zhang *et al.* 2011) and a coincident increase

in Ago-bound siRNAs derived from normally degraded structural RNAs including rRNAs and tRNAs (Buhler *et al.* 2008).

1.4.2 Cid12 and the RDRC

cid12 was initially identified simultaneously with *cid1* as a gene exhibiting increased HU/caffeine sensitivity when deleted, suggestive of a role similar to that of Cid1 in the regulation of S-phase arrest in response to DNA replication inhibition (Wang *et al.* 2000; Win *et al.* 2006). However, unlike in *cid1Δ*, *cid12Δ* strains exhibit additional defects in heterochromatic gene silencing as well as in mitotic and meiotic chromosome segregation (Win *et al.* 2006), suggesting roles for Cid12 in the formation and maintenance of centromere integrity. Purification analyses demonstrated that Cid12 resides in a heterotrimeric complex with a helicase, Hrr1, and the *S. pombe* RNA-directed RNA polymerase (RdRP), Rdp1 (Motamedi *et al.* 2004). This complex, termed the RNA-directed RNA polymerase complex (RDRC), is recruited to ncRNAs transcribed from the centromeric repeats in a manner that requires interaction with the RNA-induced transcriptional silencing (RITS) complex and is responsible for the secondary synthesis of siRNAs via the sequential activities of Rdp1 and the Dcr1 endonuclease (Motamedi *et al.* 2004). Deletion of any component of the RDRC results in catastrophic loss of siRNA production and abolishes RITS targeting to, and heterochromatic silencing of, centromeric repeats (Motamedi *et al.* 2004; Win *et al.* 2006). While no catalytic activity or targets of Cid12 has been thus far described, it has been observed that mutations in the catalytic aspartate residues of Cid12 are sufficient to induce chromosome segregation and heterochromatic gene silencing defects resembling those of the *cid12Δ* allele, suggesting that the catalytic activity of Cid12 is essential to the efficient synthesis of siRNAs (Win *et al.* 2006). Intriguingly, whilst inactivation/deletion of any one component of the RDRC is sufficient to induce defects in siRNA synthesis and centromeric heterochromatin, no *cid12Δ*-like checkpoint defects

could be found in *rdp1Δ* strains, suggesting that *cid12* may achieve its checkpoint functions through a secondary pathway independent of its role in the RDRC (Win *et al.* 2006).

ncPAP family members have been reported as key components of sRNA generation pathways in other species. As discussed above, the ncPAPs Rdn1p and Rdn2p are central members of the RDRCs in *Tetrahymena* and support siRNA synthesis through a template-looping mechanism (Lee *et al.* 2009; Talsky and Collins 2010; Talsky and Collins 2012). Mutants of the *C. elegans* ncPAP *rde-3/mut-2* exhibit defects in chromosome segregation and RNAi-mediated cosuppression of endogenous sequences by a multicopy transgene (Chen *et al.* 2005; Robert *et al.* 2005). These defects correlate with a marked depletion of siRNAs and a failure to propagate RNAi induced by feeding of dsRNA (Chen *et al.* 2005). *In vitro* tethering experiments with RDE-3 failed to report any catalytic activity (Kwak *et al.* 2004); however it is noted that one null allele of *rde-3* retrieved from an early screen for defects in RNAi-mediated cosuppression corresponded to a mutation in the catalytic core of the protein, suggesting that, as with Cid12, the catalytic activity of RDE-3 is essential to its ability to facilitate siRNA synthesis. Another *C. elegans* ncPAP, CDE-1, has been described as an interactor of the RdRP EGO-1 (van Wolfswinkel *et al.* 2009), though in this context the ncPAP appears to be responsible for the degradation, rather than the synthesis, of siRNAs. Characterisation of the polymerase activity of Cid12 and the role this activity plays in the synthesis of dsRNA by RDRC remains a major outstanding topic in research into the RNAi pathways of *S. pombe* (Goto and Nakayama 2011).

1.4.3 Putative cytoplasmic PAPs, Cid11 and Cid13

Cid11 and Cid13 are two highly similar PAPs in *S. pombe*, exhibiting 34% identity across their length and 45% identity within the catalytic domains (see Figure 3.6). *cid11Δ* was shown alongside *cid1Δ* and *cid12Δ* to elicit sensitivity to the combination of HU and caffeine, suggesting a possible role for Cid11 in the S phase checkpoint; however, subsequent genome-wide expression analyses have shown *cid11* to

be a highly specific meiotic gene with a strong expression peak during meiosis I and II but no detectable expression throughout the vegetative cell cycle (Rustici *et al.* 2004; Mata and Bähler 2006). Cid13 by contrast is expressed constitutively throughout the cell cycle and localises to granular structures in the cytoplasm (Saitoh *et al.* 2002). Cid13 was shown by two separate groups to be a processive PAP *in vitro*, and to be a multi-copy suppressor of the HU sensitivity of DNA replication checkpoint mutants (Read *et al.* 2002; Saitoh *et al.* 2002). While this appears to identify *cid13* as a paralogue of *cid1*, it was noted in these papers that *cid13* appears to act by bypassing the checkpoint pathway rather than rescuing it, given that *cid13Δ* shows sensitivity to HU alone rather than a combination of HU and caffeine, and is able to rescue the HU sensitivity of a *cdc2* mutant largely independent of checkpoint signalling (Read *et al.* 2002; Saitoh *et al.* 2002). Consistent with these observations, *cid13* acts by enhancing the pool of dNTPs available for DNA synthesis, though the mechanism of this is controversial – while Saitoh *et al*. (2002) observed a *cid13*-dependent elongation of the mRNA poly(A) tail of the *suc22* ribonucleotide reductase subunit and consequent stabilisation of this transcript, Read *et al*. (2002) found no such regulation. In an independent study, Cid13 was reported to interact with a putative nuclear import regulator, Mog1, and to be able to rescue a *mog1*^{ts} mutant when overexpressed (Oki *et al.* 2007). Phylogenetically, *cid11* and *cid13* lie in a clade of ncPAPs containing many of the characterised PUPs including those from *S. pombe*, *C. elegans* and mammals (see Figure 1.1), suggesting that they may have developed from an ancestral PUP that reverted to PAP activity and was subsequently duplicated.

Cytoplasmic polyadenylation represents a conserved and widely employed function for members of the ncPAP family. Cytoplasmic polyadenylation was first reported for the *C. elegans* ncPAP GLD-2, which forms a functional heterodimer with the RNA-binding developmental regulator GLD-3 (Eckmann *et al.* 2002; Wang *et al.* 2002; Eckmann *et al.* 2004). GLD-2 and GLD-3 cooperate to induce entry into the meiotic cell cycle during gametogenesis (Wang *et al.* 2002), at least in part by promoting the expression of the translational repressor *gld-1* (Suh *et al.* 2006). In *Xenopus* oocytes, GLD-2 resides in a mRNA-

bound complex containing CPEB and the deadenylase PARN; the superior activity of PARN maintains bound mRNAs in a deadenylated, translationally inactive state (Barnard *et al.* 2004; Kim and Richter 2006). Upon maturation of the oocyte, CPEB phosphorylation causes PARN to be ejected from the complex, allowing GLD-2 to polyadenylate mRNA and elicit rapid induction of translation independently of transcription (Kim and Richter 2006; Kim and Richter 2007). Similar pathways involving GLD-2-mediated regulation of oogenesis and early embryonic development have subsequently been reported in other organisms including *Drosophila* and mice (Nakanishi *et al.* 2006; Benoit *et al.* 2008; Kashiwabara *et al.* 2008). Additionally, GLD-2-mediated cytoplasmic polyadenylation is important for long-term neuronal plasticity in *Drosophila* and vertebrates; the mechanism, though not fully characterised, is thought to resemble that observed in oocytes (Wu *et al.* 1998; Shin *et al.* 2004; Du and Richter 2005; Rouhana *et al.* 2005; Kwak *et al.* 2008).

Interestingly, multiple cytoplasmic polymerases are known to exist in *Xenopus laevis* and *C. elegans* (Rouhana *et al.* 2005; Schmid *et al.* 2009) and have been proposed in mice (Nakanishi *et al.* 2007), suggesting that functional and/or temporal/spatial diversification may be a feature of cytoplasmic PAPs. The temporal separation of *cid13* and *cid11* expression into vegetative and meiotic life cycles respectively (Rustici *et al.* 2004; Mata and Bähler 2006) is reminiscent of the divergent functions of vertebrate GLD-2 orthologues in gametogenesis and neuronal plasticity; however, without extensive further characterisation of the activity of *cid11* and mechanisms/targets by which both ncPAPs act, no further conclusions can be drawn.

1.4.4 The unknown - Cid16

The final ncPAP in *S. pombe*, Cid16, remains the least characterised of the family members. Deletions of *cid16* have been shown to have no effect on the expression of heterochromatic repeat transcripts or markers integrated into regions of heterochromatin (Buhler *et al.* 2007; Wang *et al.* 2008) and to be

indistinguishable from *wt* under HU/caffeine treatment (Wang *et al.* 2000). Interestingly, however, whilst phylogenetic analyses of the ncPAP family place full-length Cid16 in an individual clade distantly related to a number of *C. elegans* ncPAPs (Figure 1.1), a parallel analysis using a novel, *in silico* evolution-based algorithm identified Cid16 as a member of the Cid1 clade and the closest *S. pombe* relative of Cid1 (S. Kelly, pers. comm.).

1.5 RNA and *S. pombe* heterochromatin formation

S. pombe represents an ideal model organism for the study of chromatin regulation due to the fact that the structure of and chromatin modifications on its centromeres are significantly more complex than those of *S. cerevisiae*, resembling much more closely those of multicellular organisms; additionally, its small genome size and presence of only three chromosomes, along with the presence of single copy genes encoding many key chromatin-regulatory proteins allow detailed genetic and biochemical dissection of heterochromatin formation (reviewed most recently in Cam 2010; Grewal 2010; Goto and Nakayama 2011; Lejeune and Allshire 2011). The *S. pombe* genome contains three clearly defined regions of heterochromatin: at the centromeres, telomeres and the mating-type locus (Cam *et al.* 2005; Hansen *et al.* 2005); however, the bulk of research on the establishment of heterochromatin and the roles of RNA has focused on the centromeres (Goto and Nakayama 2011).

While early mutagenesis studies were able to characterise many of the histone modifying enzymes (HMEs) and chromodomain proteins required for the establishment and maintenance of heterochromatin in *S. pombe* (Ekwall *et al.* 1996; Thon and Verhein-Hansen 2000; Yamada *et al.* 2005) and identify the hallmark chromatin modifications diagnostic of heterochromatic and euchromatic regions (Cam *et al.* 2005; Hansen *et al.* 2005), a paradigm-shifting breakthrough came with the discovery that integrity of the *S. pombe* RNAi pathway is essential for the formation of centromeric heterochromatin, silencing of reporter genes integrated at the centromere and efficient chromosome

segregation (Volpe *et al.* 2002). Similar findings have subsequently been reported in other species, supporting a widely conserved role for RNAi-directed mechanisms of heterochromatin specification (Huisinga and Elgin 2009; Matzke *et al.* 2009; Bourc'his and Voinnet 2010).

The first mechanistic data regarding this pathway came with the description of a stable heterotrimeric complex, subsequently named the RNA-induced transcriptional silencing (RITS) complex, containing the *S. pombe* argonaute protein Ago1, a heterochromatin-associated chromodomain protein Chp1 (Thon and Verhein-Hansen 2000) and Tas3 (Verdel *et al.* 2004). The RITS complex localised specifically to the heterochromatic centromere and was essential for the silencing of a centromere-integrated *ura4*$^{+}$ marker and for efficient H3K9 methylation of the centromeric repeats, a modification essential to the formation of heterochromatin (Verdel *et al.* 2004). Centromeric localisation required RITS to be loaded with mature siRNAs, since deletion of siRNA biogenesis factors (Motamedi *et al.* 2004; Sadaie *et al.* 2004; Verdel *et al.* 2004) and abolition of the Ago1 slicer catalytic activity (Irvine *et al.* 2006; Buker *et al.* 2007) both resulted in failure of RITS-mediated silencing. The requirement for RNA Pol II-mediated transcription of the heterochromatic repeats (Kato *et al.* 2005; Irvine *et al.* 2006) and the observed coimmunoprecipitation of RITS with centromeric RNA (Motamedi *et al.* 2004) suggested that the RITS complex is targeted to heterochromatic regions via interaction with nascent RNA Pol II transcripts from these heterochromatic regions. Interestingly, mutations in the *rpb2* subunit of RNA Pol II were sufficient to abolish RNAi-mediated heterochromatin formation without altering RNA Pol II occupancy of the centromeric repeats, suggesting that RNA Pol II may play a direct role in the nucleation of RNAi machinery (Kato *et al.* 2005).

Once localised to the centromeres, the primary function of the RITS complex is to recruit the histone modifying activity of the Clr4 H3K9 methyltransferase. Clr4 resides in the Clr4 complex (CLRC) which also contains Rik1, Dos1, Dos2 and the cullin-dependent ubiquitin E3 ligase Pcu4 (Hong *et al.* 2005; Horn

et al. 2005; Jia *et al.* 2005; Li *et al.* 2005; Thon *et al.* 2005); its recruitment is mediated via the LIM domain protein Stc1, which crosslinks the Dos1/Dos2/Rik1 components of CLRC to the Ago1 component of the RITS complex (Bayne *et al.* 2010). Loss of RITS localisation to centromeres or mutations in the predicted Stc1-Ago1 interaction site were sufficient to severely attenuate Clr4-mediated H3K9 methylation at the centromeric repeats (Noma *et al.* 2004; Sadaie *et al.* 2004; Bayne *et al.* 2010). Conversely, tethering of either Clr4 (Kagansky *et al.* 2009) or Stc1 (Bayne *et al.* 2010) to DNA was sufficient to induce fully functional heterochromatin formation independent of the RNAi pathway and could rescue chromosome segregation defects in centromere-defective strains (Kagansky *et al.* 2009). However, loss of *clr4* was found to abolish siRNA synthesis and prevent RITS localisation to the centromeric repeats (Volpe *et al.* 2002; Noma *et al.* 2004; Sadaie *et al.* 2004). Consistent with this, the localisation of RITS to centromeric repeats required the Chp1 chromodomain (Verdel *et al.* 2004), suggesting that Clr4-mediated H3K9 methylation generated sites for the binding of RITS to heterochromatin (Verdel *et al.* 2004). In total, these observations led to the formulation of a model in which CLRC and RITS act in a positive-feedback loop involving successive rounds of siRNA targeting of RITS to heterochromatin and consequent Clr4-mediated H3K9 methylation (Noma *et al.* 2004; Goto and Nakayama 2011).

Interestingly, while endogenous heterochromatic repeats retained a low level of H3K9 methylation even in RNAi-compromised strains, this modification was completely absent from *ura4*$^{+}$ cassettes inserted into the centromeres (Volpe *et al.* 2002; Sadaie *et al.* 2004), suggesting that the RNAi machinery may be important for the spreading of heterochromatin from the centromeric repeats into adjacent, nonrepetitive sequence. While low levels of siRNAs are likely generated from the centromeric repeats though bidirectional transcription by RNA Pol II and direct endonucleolytic processing of the dsRNA products (Volpe *et al.* 2002; Kato *et al.* 2005; Iida *et al.* 2008), the propagation and spreading of heterochromatin requires secondary synthesis of siRNAs both within the heterochromatic repeats and in

neighbouring unidirectionally-transcribed, nonrepetitive sequences. This secondary siRNA synthesis is mediated by the sole *S. pombe* RNA-dependent RNA polymerase *rdp1* (Volpe *et al.* 2002; Motamedi *et al.* 2004). *In vivo*, Rdp1 resides in the RDRC with an ATP-dependent helicase Hrr1 and the ncPAP Cid12; RDRC interacts directly with the RITS complex and is recruited by this interaction to nascent centromeric transcripts where it can synthesise long dsRNA products which are processed into siRNAs by the endonuclease Dcr1 and loaded onto RITS complexes to provide new siRNA repertoires able to target an expanded population of transcripts (Motamedi *et al.* 2004). The synthesis of dsRNA by RDRC and its digestion by Dcr1 into siRNAs is likely to be a physically coupled process, given that Dcr1 and RDRC are known to associate physically (Motamedi *et al.* 2004; Colmenares *et al.* 2007) and that this physical interaction is essential for siRNA synthesis (Colmenares *et al.* 2007). In support of this mechanism, RNAi-mediated heterochromatic silencing of centromeric *ura4*$^{+}$ cassettes was shown to require readthrough transcription from *ura4*$^{+}$ into the centromeric repeats, allowing *ura4-cen* transcripts to be targeted by centromere-specific siRNAs and subsequently act as templates for RDRC, allowing synthesis of *ura4*$^{+}$-specific siRNAs (Irvine *et al.* 2006).

While siRNAs are synthesised as short dsRNA fragments, they reside on the RITS complex almost exclusively as ssRNA forms, having ejected the complementary passenger strand. That this process also represents a point of regulation of the RNAi cascade was revealed by the discovery of a second Ago1-containing complex in *S. pombe*, the Argonaute siRNA chaperone (ARC) complex (Buker *et al.* 2007). The ARC contains Ago1 in association with two conserved proteins, Arb1 and Arb2, and is loaded with siRNA which is maintained in the dsRNA form by the inhibitory activity of Arb1 on Ago1 slicing activity (Buker *et al.* 2007). It appears likely that Ago1-bound siRNA is maintained in this inactive state until assembly into the mature RITS complex when the slicer activity of Ago1 is able to cleave the passenger strand, allowing its degradation and the consequent formation of an active RITS complex (Irvine *et al.* 2006; Buker *et al.* 2007).

Collectively, these observations have allowed the construction of a detailed model for the RNAi-mediated establishment of heterochromatin in which siRNAs are loaded onto the RITS complex and target it to the centromeric repeats via interaction with nascent centromeric RNA Pol II transcripts; this association results in the recruitment of both CLRC, which methylates H3K9, and RDRC, which synthesises dsRNA from the nascent transcript. This dsRNA is digested by Dcr1 into siRNA which is loaded onto the ARC and subsequently matured into new RITS complexes capable of targeting a larger range of nascent transcripts, allowing the cycle to repeat and siRNA-targeted, H3K9 methylation-marked heterochromatin to spread along the chromosome. This model is summarised in Figure 1.2.

One considerable feature missing from this model is the mechanism by which the initial heterochromatic mark which allows subsequent reinforcement by this positive feedback loop is established. Residual H3K9 methylation is observed on centromeric repeats in the absence of RNAi factors, suggesting that this modification may represent the *de novo* mark required for initiation of heterochromatin (Volpe *et al.* 2002; Sadaie *et al.* 2004); however, the source of these initiating marks is unclear. While several mechanisms have been proposed, including dsRNA synthesis through bidirectional transcription at the centromeric repeats (Volpe *et al.* 2002; Kato *et al.* 2005; Iida *et al.* 2008) or convergent genes (Gullerova *et al.* 2011) and/or the action of novel, Dcr1- and Rdp1-independent 'primal RNAs', which bind Ago1 and target centromeric transcripts (Halic and Moazed 2010), the synthesis of these initial marks in strains lacking Ago1 (Shanker *et al.* 2010) suggests a different mechanism must be in action. Recently, two papers (Li *et al.* 2011; Zaratiegui *et al.* 2011) identified an unexpected physical interaction between components of the CLRC complex (Dos2, Rik1) and Cdc20, the catalytic subunit of the leading-strand DNA polymerase ε. This complex progresses with

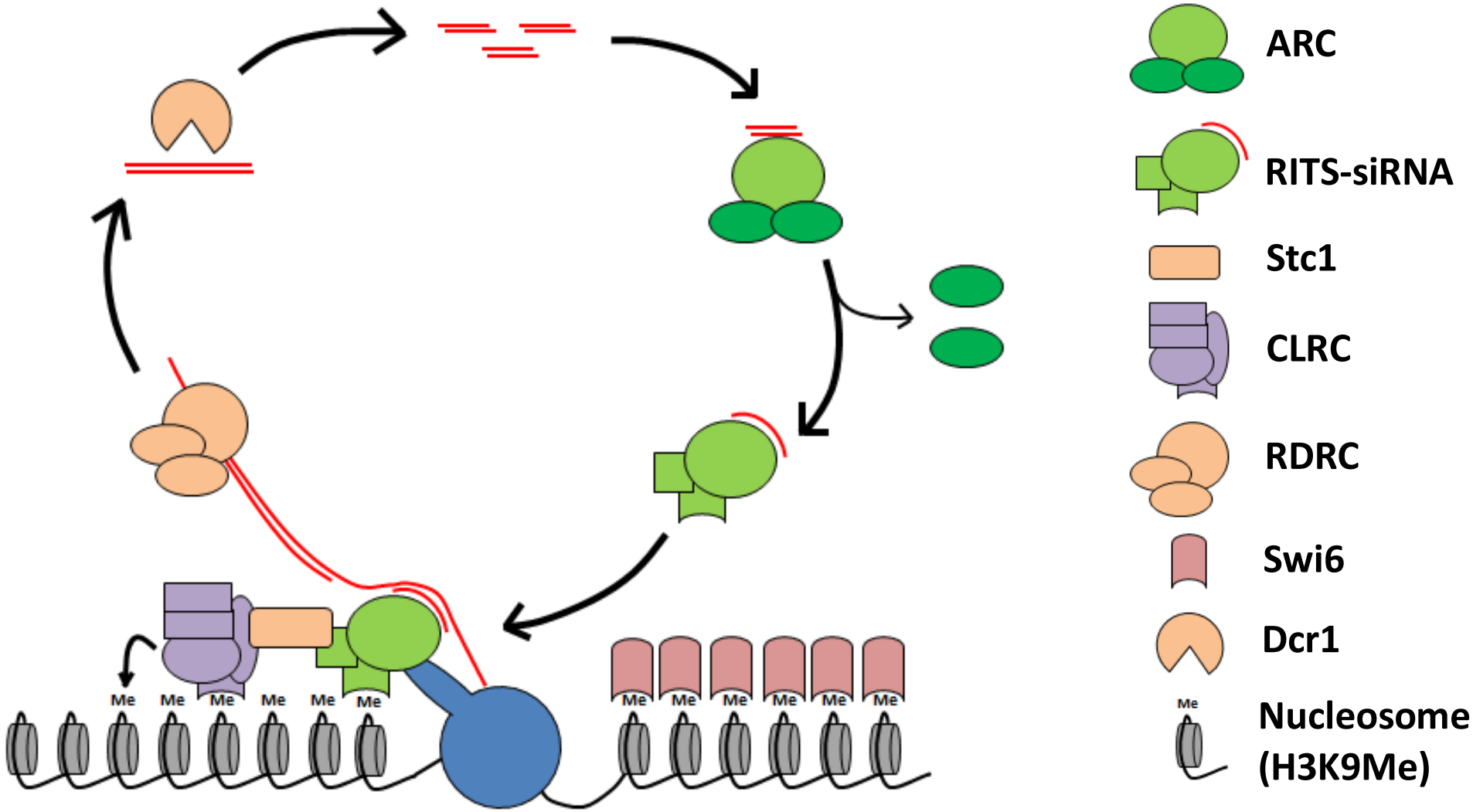


Figure 1.2: Schematic of the RITS complex/CLRC positive feedback loop and its role in the formation of H3K9-methylated heterochromatin. Figure adapted from Goto and Nakayama (Goto and Nakayama 2011) and described in the text (**1.5**)

the replication fork and is required for formation of heterochromatin on newly-assembled chromatin (Li *et al.* 2011; Zaratiegui *et al.* 2011), suggesting that functional coupling of the heterochromatic and DNA replication machinery may provide a mechanism for the RNAi-independent establishment of heterochromatin.

A second question arising from the model above is its generality to all sites of heterochromatin formation in *S. pombe*. It has been noted that the mating-type locus and the subtelomeres both contain repeat structures highly similar to those at the centromeres (Grewal and Klar 1997; Hansen *et al.* 2006) and that these repeat structures are targeted by siRNAs (Hansen *et al.* 2005) and the RITS complex *in vivo* (Noma *et al.* 2004). Deletion of these repeat structures results in reduction of telomeric and mating-type locus heterochromatin formation (Jia *et al.* 2004; Kanoh *et al.* 2005), supporting a conserved role for repeat-targeted RNAi in heterochromatin formation (Noma *et al.* 2004; Sadaie *et al.* 2004). However, mechanistic differences do exist between the heterochromatic regions. While deletion of RITS components resulted in derepression of a *ura4*$^{+}$ cassette inserted at the centromeres, an identical cassette inserted at the mating-type locus remained repressed (Volpe *et al.* 2002; Sadaie *et al.* 2004; Verdel *et al.* 2004) and retained near wild-type levels of H3K9 methylation (Sadaie *et al.* 2004). Additionally, the chromodomain proteins Chp2 and Swi6 rely on the presence of Chp1 for centromeric repeat localisation but can be efficiently localised to the mating-type locus and subtelomeres even in *chp1Δ* strains (Thon and Verhein-Hansen 2000; Sadaie *et al.* 2004), while deletions of RNAi factors that result in robust upregulation of centromeric transcripts cause only modest upregulation of the subtelomeric *tlh1-4* mRNAs, which contain the telomeric repeats (Hansen *et al.* 2006). These observations collectively suggest that, while RNAi plays an important role in the silencing of all three heterochromatic regions, additional mechanisms act at the subtelomeres and the mating-type locus that facilitate maintenance of heterochromatin in an RNAi-independent fashion (Goto and Nakayama 2011). Subsequent work in the Grewal laboratory (Jia *et al.* 2004) confirmed this hypothesis by showing that

two ATF/CREB family DNA-binding proteins, Atf1 and Pcr1, are able to interact with the mating-type region and recruit H3K9 methylation and Swi6 independently of the RNAi pathway. Simultaneous deletion of these genes and RNAi factors resulted in a failure to initiate heterochromatin assembly at the mating-type locus and derepression of an integrated *ura4*$^{+}$ marker. In a similar mechanism the Taz1 protein, a TRF-class telomere binding protein which localises to the telomeric repeats, is capable of recruiting Swi6 to telomeric and subtelomeric sequences independently of the RNAi machinery (Kanoh *et al.* 2005) and its deletion results in an additive hyperabundance of subtelomeric *tlh1-4* transcripts when combined with *dcr1Δ* (Hansen *et al.* 2006). It is thus concluded that, while the RNAi machinery is important for the formation of heterochromatin at all heterochromatic regions of the *S. pombe* genome, telomeric and mating-type loci utilise additional, RNAi-independent pathways to reinforce heterochromatin and provide multiple redundant pathways of heterochromatin initiation.

1.6 Overview of this thesis

In this work, a combination of *in vivo* genetic and *in vitro* biochemical assays of uridylyltransferase activity in native and recombinant proteins will be presented which cumulatively demonstrate that the hitherto-uncharacterised *S. pombe* ncPAP Cid16 is a second mRNA poly(U) polymerase. Characterisation of the biochemical activity and deletion phenotype of *cid16* is subsequently presented, demonstrating Cid16 to be a more processive and RNA substrate-selective PUP than Cid1 but to regulate the expression of very few mRNAs during exponential growth.

Next, a detailed analysis of the transcriptome-wide RNA stability in mutants of the *cid1* PUP demonstrates that, while this enzyme appears to have little effect on the degradation of mRNAs during exponential growth, it is surprisingly responsible for the regulation of transcription of multiple transcripts clustered in extended subtelomeric regions on chromosomes I and II. Epistasis analysis demonstrates that, while this activity is achieved independently of the RNAi machinery, it involves the

activity of a known dsRNA-targeted and/or siRNA exoribonuclease, Eri1, and thus may represent the first evidence of a previously unknown role for *cid1* in the maintenance of subtelomeric heterochromatin. Finally, the ability of *cid1* to regulate the transcription of these targets is shown to vary in response to DNA replication arrest in an allele- and gene-specific fashion, suggesting that Cid1-mediated regulation of subtelomeric heterochromatin could represent the mechanism by which *cid1* regulates S phase arrest.

Taken together, these results demonstrate that the process of RNA uridylylation in *S. pombe* is significantly more complicated than previously suspected, and that RNA uridylylation, as in other species, may play a role in the regulation of RNA-dependent heterochromatin formation in *S. pombe*.

Chapter 2
Materials and Methods

2.1 *Schizosaccharomyces pombe* Methods

2.1.1 Growth and maintenance of Schizosaccharomyces pombe

Standard methods for fission yeast growth were followed (Moreno *et al.* 1991; see also http://www.sanger.ac.uk/PostGenomics/S_pombe/docs/nurse_lab_manual.pdf). *S. pombe* was maintained on YE5S agar or minimal agar containing appropriate supplements; liquid cultures were grown in YE5S medium or supplemented Edinburgh Minimal Medium (EMM2). Strains were stored in YE5S containing 20% glycerol at -80°C. Strains used in this study are shown in Appendix A.

2.1.2 Determination of Cell Concentration

The concentration of *S. pombe* cultures was determined by measuring absorbance at 600 nm (A_{600}) with an Eppendorf Biophotometer. An A_{600} of 1.0 was taken to be equivalent to $2x10^7$ cells per mL.

2.1.3 Genomic DNA extraction

$5x10^7$ cells grown to mid-exponential phase were harvested by centrifugation (4000 x *g*, 2', RT), washed with 1 mL of dH_2O and resuspended in 200 µL cold DNA extraction buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl (pH8.0), 1 mM EDTA). 200 µL phenol:chloroform:isoamyl alcohol (25:24:1, pH 8.0) and 100 µL acid-washed glass beads were added in a screw-cap tube. Samples were vortexed for 2.5' in 30" bursts with 30" incubation on ice in between. Samples were then centrifuged at 16,000 x *g* for 5' at 4°C and the aqueous layer was moved to a separate tube with 3 volumes of 100% ice-cold

ethanol. Samples were centrifuged at 16,000 x *g* for 10' at 4°C and pellets were washed once with 500 µL ice-cold 70% ethanol, allowed to air dry and then resuspended in 50 µL TE and stored at -20°C.

2.1.4 Crossing of Schizosaccharomyces pombe

One colony each of two *S. pombe* strains of opposite mating types (h^+ and h^-) grown for 3 days on YE5S agar were collected with a bacterial loop, mixed in 15 µL of dH_2O on a ME4S plate and incubated at RT for 48-72 hr. Successful formation of asci was checked by light microscopy.

2.1.4.1 Selection of strains by tetrad dissection

A small scoop of crossed cell mixture was collected with a bacterial loop and streaked onto the edge of a YE5S agar plate. A Singer Instruments MSM System tetrad dissector was used to pick individual four-spore asci and place them in a line across the top surface of the plate. The plate was incubated for 4-8 hr at 37°C to induce ascus disintegration, then the tetrad dissector was used to separate individual spores from each ascus systematically down the plate. Plates were grown for 3-5 days and spore genotypes were determined by replica plating to selective agar. Genotypes were confirmed by *S. pombe* colony PCR.

2.1.4.2 Selection of strains by random spore analysis

A small scoop of crossed cell mixture was collected with a bacterial loop and resuspended in 1 mL of dH_2O containing 2 µL of *Helix pomatia* juice to lyse unconjugated cells and break down ascus walls, releasing spores. Samples were incubated overnight at RT or for 6 hr at 30°C and remaining spores were quantified by microscopy. 200 spores were inoculated directly onto selective medium and allowed to grow at 37°C for 3-5 days. Strain genotypes were confirmed by back-crossing against a *wild-type* strain and *S. pombe* colony PCR.

2.1.5 Transformation by the Lithium Acetate procedure

Transformations were conducted by the one-step transformation protocol using kanMX6 or hphMX6 antibiotic resistance cassettes (Bahler *et al.* 1998). 1 x 10^9 cells grown to mid-exponential phase were harvested by centrifugation (4000 x *g*, 5', RT). Cells were washed once with 1 volume of dH_2O, once with 1 mL LiAc/TE (100 mM LiAc, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA) and resuspended in 500 µL LiAc/TE. For each transformation, 100 µL cell suspension (2x10^8 cells total) were mixed with 2 µL boiled salmon sperm DNA and ~100 ng transforming DNA and incubated at RT for 10'. 260 µL of PEG/LiAc/TE (50% PEG-4000, 100 mM LiAc, 10 mM Tris-HCl (pH7.5), 1 mM EDTA) was added to each transformation followed by incubation at 30°C for 45'. Following the incubation, 43 µL DMSO was added and cells were heat-shocked for 5' at 42°C. Cells were harvested by centrifugation (1500 x *g*, 2', RT), washed once with 1 mL dH_2O and resuspended in 500 µL dH_2O. 300 µL of each transformation was plated onto non-selective YE5S for 18 hr then replica plated to selective media and grown at 30°C for 3-5 days. Successful transformation was confirmed by *S. pombe* colony PCR.

2.1.6 nmt1*-mediated overexpression of NTAP-tagged proteins*

Proteins were over-expressed in the appropriate *S. pombe* strains using the pREP1-NTAP vector, which drives expression of N-terminally TAP-tagged proteins from the thiamine-repressible *nmt1* promoter and contains the *LEU2* selectable marker (Maundrell 1993). Plasmids were transformed into *S. pombe* strains carrying the *leu1-32* auxotrophic mutant as described (2.1.5) and selected by plating onto EMM2 agar lacking leucine. Expression was repressed by adding 30 µM thiamine to the medium. When required, expression was induced by removing thiamine from overnight liquid precultures by washing four times in dH_2O and resuspending cells to an A_{600} of 0.1 in EMM2 medium lacking both leucine and thiamine. After 18 hr growth in this inductive medium, full expression was achieved.

2.1.8 Growth and induction of temperature-sensitive cdc mutants

Strains carrying a temperature-sensitive allele were inoculated into a 5 mL culture and grown overnight in YE5S medium at permissive temperature (25°C for all temperature-sensitive alleles used in this study) with shaking. Two identical cultures of each strain were diluted to an A_{600} of 0.1 and grown at 25°C to an A_{600} of 0.3-0.5. One of the two cultures was then moved to restrictive temperature (36°C for all temperature-sensitive alleles used in this study) whilst the other was left at 25°C and both were incubated with shaking for 4 hr. Restrictive cultures were checked for phenotypic evidence of induction of the temperature-sensitive mutant. Both cultures were then treated for RNA extraction as described in 2.4.1.2.

2.2 *Escherichia coli* Methods

XL-1 Blue (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI^{q}ZΔM15 Tn10* (Tet^R)]) were used to select and amplify plasmid DNA. BL21 DE3 pLysS (F' *ompT hsdS$_B$* (r_B^- m_B^-) *gal dcm* (DE3) pLysS (Cam^R)) were used to over-express GST-tagged recombinant protein.

2.2.1 Growth and Maintenance of E. coli

Standard methods were followed. *E. coli* was maintained on LB agar with appropriate selective antibiotics at 4°C; liquid cultures were grown in LB medium with shaking at 37°C. Strains were stored in LB containing 20% glycerol at -80°C.

2.2.2 Generation of Chemically Competent E. coli

2.2.2.1 $RbCl_2$ Method

5 mL cells grown to saturation overnight at 37°C were diluted into 200 mL LB and grown at 37°C until A_{600} reached 0.4-0.8. Cells were harvested in chilled 50 mL Falcon tubes (4000 x *g*, 15', 4°C) and

resuspended in 80 mL cold TFBI (30 mM KAc, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% glycerol (pH 5.8)). After incubation on ice for 20', cells were harvested by centrifugation as above and resuspended in 8 mL cold TFBII (10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% glycerol (pH 6.5)). After a further 20' incubation on ice, 200 µL aliquots were placed in chilled 1.5mL Eppendorf tubes and snap-frozen in liquid N_2 before long-term storage at -80°C.

2.2.2.2 KCM Method

5 mL cells grown to saturation overnight at 37°C were diluted into 50 mL fresh SOB (20 g/L Bacto-tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.7 mM KCl, 20 mM Mg^{2+}) and grown at 37°C until A_{600} reached 0.3-0.6. The culture was chilled on ice for 10-30' with occasional mixing then cells were harvested by centrifugation (4000 x *g*, 10', 4°C). Cells were resuspended in 5 mL cold TSB (LB medium containing 10% PEG-6000, 5% DMSO, 20 mM Mg^{2+}) and incubated on ice for 10'. 100 µL aliquots were placed in chilled 1.5 mL Eppendorf tubes and snap-frozen in liquid N_2 before long-term storage at -80°C.

2.2.3 Transformation of E. coli

2.2.3.1 $RbCl_2$ Method

An aliquot of competent cells (prepared as in 2.2.2.1) was mixed on ice with 100 ng plasmid DNA or 10 µL of a ligation reaction. The transformations were incubated on ice for 30' then heat-shocked at 42°C for 45" and returned to ice for 2'. 250 µL of RT LB medium was added to each transformation and cells were incubated at 37°C with shaking for 45'. 50-200 µL of each transformation was spread onto pre-warmed LB agar containing the necessary selective antibiotics and incubated overnight at 37°C.

2.2.3.2 KCM Method

100 ng of plasmid DNA or 10 µL of a ligation reaction was mixed with 20 µL of KCM buffer on ice (0.5 M KCl, 0.15 M $CaCl_2$, 0.25 M $MgCl_2$) and made up to 100 µL with cold dH_2O. This mixture was added to an

aliquot of competent cells (prepared as in 2.2.2.2) and incubated on ice for 30'. Cells were then heat-shocked at 42°C for 90" then returned to ice for 2' before addition of 700 µL RT LB medium and incubation at 37°C with shaking for 1 hr. Cells were harvested by centrifugation (1500 x *g*, 2', RT) and resuspended in 500 µL LB medium, 50-250 µL of which was spread onto LB agar containing the necessary selective antibiotics. Plates were incubated overnight at 37°C.

2.2.4 Purification of Plasmid DNA

2.2.4.1 Miniprep by Qiagen kit

Small-scale preparations of plasmid DNA were carried out using a Plasmid Miniprep kit (QIAgen) according to the manufacturer's instructions. Briefly, 1.5 mL of an overnight bacterial culture was harvested by centrifugation (1500 x *g*, 2', RT). The cell pellet was resuspended in 250 µL cold buffer P1, lysed with 250 µL buffer P2 and neutralised with 350 µL buffer N3. Samples were clarified by centrifugation (16,000 x *g*, 10', RT) and the supernatant was bound to a silica-gel column then washed with 750 µL buffer PE. DNA was eluted with 30 µL dH_2O and stored at -20°C.

2.2.4.2 Phenol-Chloroform method

1.5 mL of an overnight bacterial culture was harvested by centrifugation (1500 x *g*, 2', RT). The cell pellet was resuspended in 200 µL Resuspension Buffer (4% Triton X-100, 2.5 M LiCl, 50 mM Tris-HCl (pH7.5), 62.5 mM EDTA). 200 µL of phenol: chloroform: isoamyl alcohol (25:24:1, pH 8.0) was added to each suspension and layers were mixed by vortexing for 15" then separated by centrifugation (16,000 x *g*, 2', RT). The aqueous layer was removed to a separate tube containing 400 µL ice-cold 100% ethanol and vortexed. DNA was precipitated at -20°C for 5' and collected by centrifugation (16,000 x *g*, 5', RT). DNA pellets were washed once with 200 µL ice-cold 70% ethanol, allowed to air-dry and resuspended in 12 µL 10 mM TE containing 100 µg/mL RNAse A.

2.3 DNA Methods

2.3.1 Determination of DNA concentration

DNA concentrations were determined by measuring the absorbance at 260 nm (A_{260}). An A_{260} of 1.0 was taken as the equivalent of 50 µg/mL double-stranded DNA.

2.3.2 PCR from a DNA template

2.3.2.1 PCR with Taq polymerase

Reactions were performed on template DNA diluted to approximately 0.1 mg/mL in dH_2O or Tris/EDTA. Reagents were added as follows (to final concentration in brackets) to a final concentration of either 50 µL or 25 µL: 10x *Taq* buffer (1:10 dilution); dATP/dCTP/dGTP/dTTP mix (200 µM each); $MgCl_2$ (500 µM); primers forward and reverse (200 µM each); and red*Taq* polymerase (50 U/mL). Thermal cycling was carried out as follows: 94°C for 3', 25-35 cycles of (94°C for 30", 55°C for 30", 72°C for 1' per kb of product), 6' at 72°C then chilled to 8°C for storage. Primers used in this study are listed in Appendix C.

2.3.2.2 Proof-reading PCR with Expand *polymerase*

Reactions were performed on template DNA diluted to approximately 0.1 mg/mL in dH_2O or Tris/EDTA. Reagents were added as follows (to final concentration in brackets) to a final concentration of either 50 µL or 25 µL: 10x Expand buffer (1:10 dilution); dATP/dCTP/dGTP/dTTP mix (200 µM each); primers forward and reverse (200 µM each); and Expand polymerase (70 U/mL). Thermal cycling was carried out as follows: 94°C for 3', 35 cycles of (94°C for 45", 55°C for 45", 72°C for 1' per kb of product), 6' at 72°C then chilled to 4°C for storage. Primers used in this study are listed in Appendix C.

2.3.3 S. pombe *colony PCR*

A fresh colony of *S. pombe* was collected with a 200 µL pipette tip and resuspended in 25 µL dH_2O. Cells were lysed at 99°C for 7' then cooled to 25°C. Immediately before use a further 25 µL of dH_2O was added and the pellet resuspended. 5 µL of the cell lysate was made up to 50 µL with red*Taq* reaction mixture containing gene-specific primers (listed in Appendix C) as described in 2.3.2.1 and thermal cycling was carried out as follows: 94°C for 3', 35 cycles of (94°C for 40", 55°C for 40", 72°C for 1' per kb of product), 6' at 72°C then chilled to 8°C for storage.

2.3.4 E. coli *colony PCR*

A single colony of bacteria was briefly resuspended directly in 25 µL red*Taq* reaction mixture containing gene-specific primers (listed in Appendix C), and thermally cycled as described in 2.3.2.1. If parallel amplification of the bacteria was desired, the tip used to collect the pellet was (subsequent to PCR mixture resuspension) placed in 2 mL LB supplemented with Ampicillin (100 µg/mL) and grown overnight with shaking at 37°C.

2.3.5 Agarose gel electrophoresis

0.2 volumes of 6x DNA loading dye (30% (v/v) glycerol containing 0.25% (w/v) each of Bromophenol Blue and Xylene Cyanol FF) were added to each sample prior to loading on a 1-2% agarose gel (high-quality agarose dissolved in 1x TBE buffer) containing 0.5 µg/mL ethidium bromide. Electrophoresis was carried out at 2-10 V/cm in 1x TBE buffer. DNA was visualised using a UV transilluminator. Hyperladder I or II (Bioline) was used as a size marker.

2.3.6 Gel extraction of DNA

2.3.6.1 Extraction by QIAgen kit

DNA was purified from agarose gel slices using the QIAquick gel extraction kit (QIAgen) according to the manufacturer's instructions. Briefly, gel slices were melted in high-salt buffer and DNA was bound to a silica-gel column. Impurities were removed with wash buffer containing ethanol and the purified DNA was eluted with 30 µL dH_2O.

2.3.6.2 Extraction by filter-tip centrifugation

Gel slices were loaded into the top of a 1000 µL Gilson pipette filter tip which had had the tip trimmed to just below the filter. The trimmed filter was loaded into a 1.5mL Eppendorf tip and centrifuged (16,000 x *g*, 1', RT) to elute the buffer from the agarose. DNA was then precipitated by addition of NaAc (pH 5.2) to 0.3 M and 3 volumes of chilled 100% ethanol followed by incubation at -20°C for 20'. DNA was pelleted by centrifugation (16,000 x *g*, 10', RT), washed once with chilled 70% ethanol and air-dried, then resuspended in 10-30 µL dH_2O.

2.3.7 Restriction enzyme digestion

Digest reaction mixtures were performed on 0.2-1.5 µg DNA under the following conditions: 1x restriction buffer (as specified by enzyme(s) manufacturer); 10 µg/mL bovine serum albumin; and 5-10 U restriction enzyme(s), incubated at optimum temperature (in most cases, 37°C) for 1-4 hr then (for thermolabile enzymes) at inactivating temperature for 20'. Digestion was verified by agarose gel electrophoresis as described in 2.3.5.

2.3.8 DNA ligation

Digested and purified insert and linearised vector DNA were combined in a 4:1 molar ratio in a reaction medium containing 1x T4 DNA ligase buffer and 10 U T4 DNA ligase. Ligations were incubated at 16°C overnight and transformed into bacteria as described in 2.2.3.

PCR products for T/A cloning were ligated directly into pCR2.1 using the TA Cloning Kit (Invitrogen). Briefly, 0.5 µL pCR2.1 and 2 µL fresh PCR product were combined in the presence of 1x T4 DNA ligase and 10U T4 DNA ligase. Reactions were incubated at RT overnight and transformed into bacteria as described In 2.2.3.

2.3.9 DNA sequencing

Dideoxy-DNA sequencing was carried out using the BigDye Terminator sequencing mixture v2.0 (Source Bioscience). ~200 ng DNA was combined with 3.2 pmol oligonucleotide primer (primers OR114/OR115 for cRACE sequencing; primers SF31/SF32 for pGEX sequencing; primers DS180/DS181/DS243 and various gene-specific primers for pREP1 sequencing; primer sequences can be found in Appendix C), 1x BigDye reaction mixture and 0.75x reaction buffer in a total volume of 10 µL. Reactions were thermally cycled as follows: 96°C for 4' then 25 cycles of (96°C for 10", 50°C for 5", 60°C for 4') and chilled to 4°C. Reactions were halted with 0.25 volumes EDTA (125 mM) and dideoxy-terminated PCR products were ethanol precipitated at -20°C for 20' then pelleted by centrifugation (16,000 x *g,* 20', 4°C). Pellets were washed once with chilled 70% ethanol, air-dried and submitted to Source Bioscience at the Department of Zoology, University of Oxford for analysis. The sequencing data was analysed using open-source ApE software.

2.3.10 Site-directed mutagenesis of vectors

Plasmids were mutagenised using primers designed according to the parameters described in the QuikChange Site-Directed Mutagenesis kit (Agilent Technologies). 10 ng of template plasmid was combined in 50 µL total reaction with 0.25 µM mutagenic primers fwd and rev, 1x *Pfu* buffer, dATP/dCTP/dGTP/dTTP mix (200 µM each) and *Pfu* Turbo enzyme (50 U/mL) (Stratagene). Mutagenesis was performed by thermal cycling as follows: 95°C for 1'; 18 cycles of (95°C for 50", 55°C for 50", 68°C for 1' per kb template plasmid), 68°C for 7' then chilling to 4°C. 10 µL of the PCR product was resolved by agarose gel electrophoresis as described in 2.3.5 to confirm productive PCR and the remaining 40 µL reaction mixture was spiked with 20 U DpnI (New England Biolabs) for 1 hr at 37°C to digest methylated parental DNA. Mutagenised plasmid was then transformed into bacteria as described in 2.2.3. Positive transformants were selected and plasmid DNA purified as described in 2.2.4.1. Successful mutagenesis was confirmed by sequencing of retrieved plasmids as described in 2.3.9. A list of mutant plasmids created in this study is provided in Appendix B; the mutagenic primers used can be found in Appendix C.

2.3.11 Synthesis of α-^{32}P-dCTP-labelled probes for northern blotting

A 300-500 bp region of the gene of interest was amplified from *S. pombe* genomic DNA using PCR as described in 2.3.1, quantified and diluted to 25 µg/µL. 100 ng of this PCR product was then labelled with 5 µL α-^{32}P-dCTP (6000 Ci/mmol) (Perkin-Elmer) using the Rediprime II Random Prime Labelling kit (Roche). Briefly, DNA was denatured at 98°C for 5' in 45 µL TE, cooled on ice and added to the kit-provided mixture of dATP/dGTP/dTTP, exonuclease-free Klenow enzyme and random hexameric primers. 5 µL α-^{32}P-dCTP was added to this reaction, thoroughly mixed and incubated at 37°C for 30' to synthesise ^{32}P-CMP-containing second strand DNA. Reactions were quenched by the addition of 5 µL 0.2 M EDTA (pH 8.0). Residual α-^{32}P-dCTP was removed by centrifugation (1,000 x *g*, 4', RT) using miniQuickSpin DNA Columns (Roche). Eluate was stored at -20°C until hybridisation.

2.4 RNA Methods

2.4.1 S. pombe RNA purification

2.4.1.1 High-yield method

$1x10^8$ cells grown to early exponential phase were harvested by centrifugation (4,000 x *g*, 2', 4°C), washed once with one volume of dH_2O and stored at -80°C until use. Cells were resuspended in 750 µL TES (10 mM Tris-HCl (pH 7.5), 10 mM EDTA, 0.5% SDS) and immediately vortexed for 10" with 750 µL acidic phenol: chloroform (5:1, pH 4.7). Samples were incubated at 65°C for 1 hr with vortexing every 15'. After cooling on ice for 2', samples were vortexed for 20" and centrifuged (16,000 x *g*, 15', 4°C) in heavy phase-lock tubes to separate layers. The aqueous (upper) phase was added to 700 µL chilled chloroform:isoamyl alcohol (25:1) in another phase-lock tube and mixed by inversion. Samples were again centrifuged (16,000 x *g*, 5', 4°C) and 400 µL of the aqueous phases were combined with 10 µL 7 M NH_4Ac and 1.2 mL chilled 100% ethanol. Samples were incubated at -20°C for >1 hr or -80°C for >20' and centrifuged (16,000 x *g*, 15', 4°C). The pellet was washed once with 500 µL chilled 70% ethanol, air-dried and resuspended in 100 µL dH_2O.

2.4.1.2 High-stringency method

$1x10^8$ cells grown to early exponential phase were collected by centrifugation (4000 x *g*, 2', 4°C), washed once with one volume of dH_2O and stored at -80°C until use. Cells were resuspended in 400 µL chilled AE buffer (50 mM NaAc (pH 5.0), 10 mM EDTA) and added to 50 µL 10% SDS and 500 µL acidic phenol: chloroform (5:1, pH 4.7). Samples were vortexed for 15" and incubated at 65°C for 9', then returned to ice for 5' to cool. Samples were again vortexed for 10" then centrifuged (16,000 x *g*, 15', 4°C) to separate layers. Aqueous layers were transferred to a second 500 µL of acidic phenol: chloroform (5:1, pH 4.7), vortexed for 10" and centrifuged (16,000 x *g*, 5, 4°C) a second time. Aqueous layers were combined in a new tube with 10 µL 7 M NH_4Ac and 1 mL chilled 100% ethanol and incubated at -20°C for

>30'. Pellets were collected by centrifugation (16,000 x *g*, 15', 4°C), washed once with chilled 70% ethanol, air-dried and resuspended in 100 µL dH_2O.

2.4.2 Determination of RNA concentration

The concentration of RNA preparations was determined by measuring the absorbance at 260 nm using an Eppendorf Biophotometer. An A_{260} of 1.0 was taken as the equivalent of 40 µg/mL RNA. Pure RNA had an A_{260}:A_{280} ratio of approximately 1.9.

2.4.3 Phenol-chloroform extraction

An equal volume of acidic phenol:chloroform (5:1, pH 4.7) was added to RNA preparations. Samples were mixed by vortexing for 10" and centrifuged (16,000 x *g*, 10', 4°C). The aqueous layer was transferred to a new tube and an equal volume of chloroform: isoamyl alcohol (25:1) was added to each sample. Samples were then vortexed for 10" again and re-centrifuged (16,000 x *g*, 5', 4°C). The aqueous layer was moved to a new tube and the RNA was ethanol precipitated as described in 2.4.4.

2.4.4 Ethanol precipitation

2.5 volumes of chilled 100% ethanol and 0.025 volumes 7 M NH_4Ac were added to RNA preparations, which were then vortexed for 10" and incubated at 20°C for >30' or -80°C for >10'. Samples were centrifuged (16,000 x *g*, 15', 4°C) and washed once with 100-500 µL chilled 70% ethanol. Pellets were air-dried and resuspended in dH_2O.

2.4.5 circularised Rapid Amplification of cDNA Ends (cRACE)

This method was initially described in Couttet *et al.* (1997); for a diagrammatic representation of the cRACE technique, see Figure 2.1. RNA samples prepared as in 2.4.1.1 were quantified. To analyse decapped messages, 12 µg RNA was immediately combined with 1x T4 RNA ligase I buffer and 1 U T4

RNA ligase I in a large volume (400 µL) to favour intra-molecular over inter-molecular ligation and incubated at RT overnight. To analyse capped RNAs, RNA was digested with shrimp alkaline phosphatase (SAP) to dephosphorylate (and hence exclude from ligation reactions) any non-methylguanosine capped degradation intermediates. SAP reactions were performed as follows: 12 ug of RNA was combined with 1x SAP buffer and 1 U SAP and incubated at 37°C for 30'. RNA was then phenol-chloroform extracted to remove all SAP as described in 2.4.3. 5'-methylguanosine-capped RNAs were then digested with tobacco acid pyrophosphatase (TAP) to expose a ligatable 5' monophosphate as follows: SAP-treated RNA was combined with 1x TAP buffer and 2.5 U TAP and incubated at 37°C for 1 hr prior to ethanol precipitation as described in 2.4.4. Finally, 5'-monophosphate RNAs were intramolecularly ligated as described for decapped RNAs above.

After intramolecular ligation, decapped and capped circularised RNA products (cRNA) were ethanol precipitated as described in 2.4.4 and resuspended in 20 µL dH_2O. cDNA synthesis was performed on 4 µL of this cRNA with a gene-specific primer as described in 2.4.9. 4 µL of the resulting cDNA was then amplified with divergent primers (i.e. converging across the intramolecular ligation site) as described in 2.3.2.1, except with 23 cycles and a 72°C elongation time of 50". Gene-specific products of this 1st PCR were then amplified with a nested PCR using a second set of divergent primers (also 23 cycles and 50" elongation). Nested PCR products were cloned directly in pCR2.1 as described in 2.3.8. Transformants were screened for inserts by *S. pombe* colony PCR using nested PCR divergent primers as described in 2.3.3 and insert-containing plasmids were purified as described in 2.2.4.2. Inserts were sequenced as described in 2.3.9.

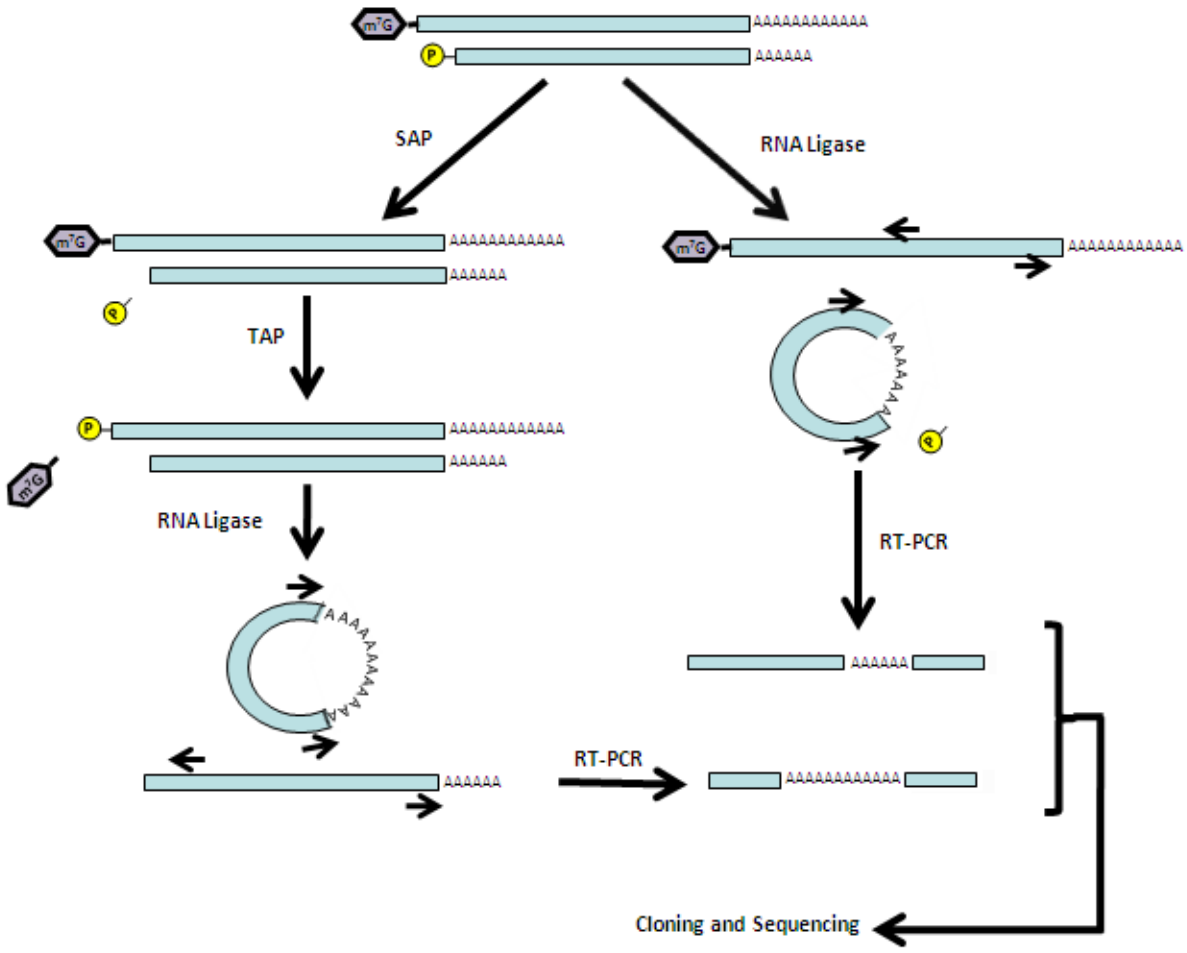


Figure 2.1: Schematic of the circularised Rapid Amplification of cDNA Ends (cRACE) technique. Total RNA purified from *S. pombe* cells is separated into two pools, "decapped" (right arm) and "capped" (left arm). The "decapped" sample is immediately treated with T4 RNA ligase in a large volume to favour intramolecular ligations and the formation of 5'-3'-ligated circular RNA (cRNA) molecules specifically from the decapped transcripts but not the 5'-blocked capped transcripts. The "capped" sample is treated sequentially with SAP to remove ligatable phosphates from the decapped transcripts, and TAP to cleave the pyrophosphate bond of the capped mRNAs, liberating ligatable 5'-monophosphates. The formerly capped RNAs are then intramolecularly ligated as above and both "decapped" and "capped" 5'-3' ligation sites are amplified by reverse transcription and nested PCR. These PCR products are cloned and selected in *E. coli* and the ligation sites sequenced.

2.4.6 Poly(N) Polymerase assay

2.4.6.1 5'-^{32}P labelling of RNA primers

RNA oligonucleotides were labelled on the 5' end with ^{32}P-phosphorylation as follows: 16 pmol of oligonucleotide was combined with 16 pmol γ-^{32}P-ATP (6000 Ci/mmol) (Perkin-Elmer) in the presence of 1x polynucleotide kinase buffer and 20 U polynucleotide kinase in a total reaction volume of 20 µL. The reaction was incubated for 45' at 37°C and made up to 50 µL with 0.3 M NaAc (pH 5.2). Residual γ-^{32}P-ATP was removed by size-exclusion chromatography using miniQuickSpin Oligo Columns (Roche) and the purified radiolabelled oligonucleotide was ethanol precipitated on dry ice for 15'. The pellets was washed once with chilled 80% ethanol, air-dried and resuspended in 50 µL dH_2O.

2.4.6.2 Poly(N) polymerisation reaction

GST- or NTAP-tagged purified protein was tested for poly(N) polymerisation activity as follows: 43 ng of GST-purified recombinant protein or 5 µL of NTAP-tagged protein-bound IgG-Sepharose beads was combined with 0.5 µL per reaction of 5'-^{32}P-labelled RNA oligonucleotide prepared as in 2.4.8.1 in the presence of 1x PAP buffer (20 mM Tris-HCl (pH 7.8), 150 mM KCl, 2 mM $MgCl_2$, 5% glycerol, 100 µg/mL BSA, 500 µM DTT), 4 U RNAseOUT™ Recombinant Ribonuclease Inhibitor and 500 µM ATP, CTP, GTP or UTP (as appropriate). For physiological conditions, a ribonucleotide triphosphate mixture (500 µM ATP, 50 µM CTP, 50 µM GTP, 50 µM UTP) was used. Final reaction volumes were 10 µL for recombinant GST-tagged protein and 15 µL for NTAP-tagged protein-bound IgG-Sepharose beads. Reactions were incubated for 15' at 37°C (GST-tagged protein reactions) or 30' at 30°C (NTAP-tagged protein reactions). Reactions were stopped by addition of 20 µL Stop Buffer (20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM EDTA) and RNA oligonucleotides were ethanol precipitated by the addition of 100 µL of chilled 100% ethanol and incubation at -20°C overnight or -80°C for >15'. Reaction products were then resolved by 7 M Urea/10% PAGE analysis as described in 2.4.6.3.

2.4.6.3 RNA polyacrylamide gel electrophoresis

Ethanol-precipitated RNA samples were collected by centrifugation (16,000 x *g*, 20', 4°C) and pellets washed once with 100 µL chilled 80% ethanol, air dried and checked for successful pellet retrieval with a Geiger counter. RNA pellets were resuspended in 12 µL urea loading buffer (7 M Urea containing 0.25% (w/v) Bromophenol Blue and Xylene Cyanole FF) and denatured at 98°C for 1' then cooled on ice for 2' prior to loading.

A 1.5 mm-thick 20x40 cm 7 M Urea/10% polyacrylamide sequencing gel was poured according to the following recipe: 25.6 mL 7 M Urea; 10 mL 40% sequencing mix acrylamide (19:1 acrylamide:bisacrylamide); 4 mL 10x TBE; 400 µL 10% ammonium persulphate; and 40 µL N,N,N',N'-tetramethylethylene-1,2-diamine (TEMED). The wells of the gel were washed free from urea and pre-run in 1x TBE at 40 W for 1 hr prior to sample loading. 6 µL of each RNA sample was loaded and the gel was run at 40 W for a further 70'. Subsequent to running, gel apparatus was dismantled, the gel on one glass plate was wrapped in Saran plastic wrap and exposed to a phosphor-screen (Molecular Dynamics) for 15-30'. Screens were scanned using a FLA5000 Fuji PhosphoImager and Images were analysed with AIDA™ analysis software (Raytest GmbH) and Adobe Photoshop™ CS3.

2.4.6.4 Densitometry analysis

For analysis of band density, AIDA™ analysis software was used to define lanes for the no protein, NTAP-Cid1/UTP and NTAP-Cid16/UTP samples from A_{13} to $A_{15}U_{15}$ products. Densitometry was binned into signal for each individual nt extension product and bins were applied equally to all three samples to eliminate running variation. Individual band signals were then normalised for total densitometry in each lane, and background signal (no protein sample) for each individual band was deducted. Finally, the percentage of total lane densitometry represented by each individual band was calculated and compared band-by-band between NTAP-Cid1/UTP and NTAP-Cid16/UTP samples.

2.4.7 UTP selectivity testing

2.4.7.1 RNAse A-RNAse T1 digestion assay

NTAP-Cid1- and NTAP-Cid16-bearing IgG-Sepharose beads were used to conduct four identical poly(N) polymerase tests on 5'-^{32}P-A_{15} in the presence of UTP alone or physiological concentrations of nucleotides as described in 2.4.8.2 (except with the omission of the RNAseOut™ recombinant ribonuclease inhibitor). However, after the 30' incubation at 30°C, each of the four identical samples were spiked with 0.5 µL of dH_2O, RNAse A, RNAse T1 or RNAse A/RNAse T1 mixture and incubated for a further 15' at 30°C. Samples were then carried through the rest of the protocol as described in 2.4.8.2 and resolved on a 7 M Urea/10% polyacrylamide gel as described in 2.4.8.3.

2.4.7.2 RNAse III digestion assay

NTAP-Cid1- and NTAP-Cid16-bearing IgG-Sepharose beads were used to conduct seven identical poly(N) polymerase reactions on 5'-^{32}P-A_{15} in the presence of physiological concentrations of nucleotides. Reactions were prepared as described in 2.4.8.2; however, samples were then incubated for 45' rather than 30' at 30°C. After this incubation, samples were boiled at 98°C for 5' then placed immediately on ice to degrade active polymerases and dissociate excess polymerised 5'-^{32}P-A_{15} from any nascent oligo(U) tails. Each sample was supplemented with 20 µL of RNAse III Adjustment Buffer (40 mM Tris-HCl (pH 7.9), 15 mM MgAc, 0.5 mM DTT) and either dH_2O, RNAse A, RNAse III (2 U) alone or RNAse III in the presence of excess cold A_{15}, C_{15}, G_{15} or U_{15} (120 pmol each). Samples were then incubated for a further 45' at 30°C. Reactions were stopped with 70 µL Stop Solution as described in 2.4.8.2 and ethanol precipitated with 200 µL chilled 100% ethanol at -20°C overnight of -80°C for >15'. Samples were resolved on a 7 M urea/10% polyacrylamide gel as described in 2.4.8.3

2.4.8 DNAse treatment of total RNA

10 µg of RNA was diluted to 50 µL in 1x RQ1 DNAse buffer and 2.5 U RQ1 DNAse (Promega) and incubated at 37°C for 1 hr. After 1 hr, a further 2.5 U of RQ1 DNAse was added and the reaction was incubated for a further 1 hr at 37°C. After the 2 hr incubation, samples were incubated at 70°C for 15' to inactivate DNAse.

2.4.9 cDNA synthesis

cDNA was synthesised using ImProm-II™ Reverse Transcriptase (Promega). Briefly, an RNA sample (usually 1 µg) was combined with 20 pmol primer (oligo$(dT)_{20}$ unless noted otherwise) in a volume of 9.1 µL and incubated at 70°C for 5' before immediate cooling to 4°C on ice for 5'. After this incubation, remaining reagents were added to the following final concentrations in a total volume of 20 µL: ImProm-II Reaction Buffer (1x); $MgCl_2$ (3 mM); dATP/dCTP/dGTP/dTTP mix (0.5 mM each) and ImProm-II Reverse Transcriptase (4 U/µL). Reactions were then incubated as follows: RT for 5'; 42°C for 1 hr; 70°C for 15'; then chilled to 4°C. Samples were stored at -20°C until use. Primer OR243 was used for cDNA syntheses from circularised *act1* sequences in cRACE; oligo$(dT)_{20}$ was used for poly$(A)^+$ RNA-dependent cDNA synthesis. A list of primer sequences is provided in Appendix C.

2.4.10 Quantitative RT-PCR (qRT-PCR)

1 µg DNAse-treated RNA was used for oligo-$(dT)_{20}$ cDNA synthesis (+RT reaction) as described in 2.4.9; in parallel, cDNA syntheses lacking reverse transcriptase enzyme (-RT reaction) were prepared for each sample to provide a control for DNA contamination and/or non-specific priming. +RT and –RT cDNAs were then diluted to 40 µL with dH_2O. qRT-PCR primers for each gene of interest and an endogenous house-keeping gene (in this study, *pik1*, a phosphatidylinositol-4-kinase that shows no expression fluctuations over the course of the cell cycle (Rustici *et al.* 2004) or in response to varying environmental

stressors; Chen *et al.* 2003) were designed using the Real-Time PCR Primer Design Tool (GenScript). For each sample, qRT-PCR reactions were prepared for both +RT and -RT material with primer pairs for each gene of interest and for the endogenous house-keeping control gene. For each primer pair/sample combination, three technical replicates of the qRT-PCR reaction were prepared for the +RT cDNA and one technical replicate of the qRT-PCR reaction was prepared for the -RT control material. To prepare the qPCR reactions, 1 µL of a +RT/-RT reaction were combined with 500 µM final concentration of fwd and rev qPCR primers and 1x SYBR Green master mix (Bioline) and quantified using the RotorGene RG-3000 (Corbett Research) running the following thermal cycle: 95°C for 15'; 40 cycles of (95°C for 10", 55°C for 10", 72°C for 10"), 0.3°C/s ramping from 72°C→95°C (melt curve) and finally chilling to 4°C. Data were analysed using the Rotor-Gene 6 software.

2.4.11 Northern blotting

2.4.11.1 RNA agarose gel electrophoresis

10-20 µg RNA were combined with 3 volumes of Loading Buffer (50% formamide, 6% formaldehyde, 1x MOPS, 10% glycerol, 0.25% (w/v) Bromophenol Blue, 20 µg/mL ethidium bromide) and incubated at 65°C for 10' then cooled on ice prior to loading. A 1% agarose/0.66 M formaldehyde gel was poured and samples were loaded. The gel was run in 1x MOPS at 5 V/cm (75 V for the 15 cm gel used here) for 4 hr and ribosomal RNA (rRNA) migration was visualised on an UV transilluminator.

2.4.11.2 Transfer to nitrocellulose membrane

The gel was rinsed in water for 10' then rinsed twice in 10x SSC (1.5 M NaCl, 0.15 M sodium citrate (pH 7.0)) for 10'. The gel was then placed on a Whatmann 3MM paper wick with ends immersed in excess 20x SSC buffer (3 M NaCl, 0.3 M sodium citrate (pH 7.0)). A Hybond-N^+ nylon membrane (Amersham) was placed in contact with the gel and 2 layers of Whatmann 3MM paper, a handful of paper towels and a weight were placed on top. This set-up was left overnight to allow RNA to be transferred from the gel

onto the nylon membrane by capillary action. RNA was then cross-linked to the membrane using a UVC 500 Crosslinker (Stratagene) at 700 mJ/cm^2 for 30". Membranes were visualised under UV transillumination to confirm successful RNA transfer and stored at 4°C until hybridisation.

2.4.11.3 Blocking and hybridisation of the membrane

α-^{32}P-dCMP-labelled DNA probes for the gene(s) of interest and an endogenous house-keeping gene (*pik1*, as used for qPCR analyses) were prepared as described in 2.3.11. Membranes were placed in a 150 mL hybridisation bottle with 7.5 mL hybridisation buffer (5x SSC (0.75 M NaCl, 75 mM sodium citrate (pH 7.0)), 50% formamide, 1% SDS, 1x Denhardt's solution (0.02% (w/v) BSA Fraction V, 0.02% (w/v) Ficoll 400, 0.02% (w/v) polyvinylpyrrolidone), 100 µg/mL boiled herring sperm DNA, 10% dextran sulphate) and pre-hybridised at 42°C for 1-2 hr. The ^{32}P-dCMP-containing DNA probe was denatured at 98°C for 5' then chilled on ice before being added to the hybridisation buffer and hybridised to the membrane at 42°C overnight. After the overnight incubation, the hybridisation solution was discarded and the membrane was washed as follows: 15' at 42°C in 1x SSC (150 mM NaCl, 15 mM sodium citrate (pH 7.0) containing 0.1% SDS; 15' at 65°C in 1x SSC containing 0.1% SDS; and twice for 10' at 65°C in 0.1x SSC (15 mM NaCl, 1.5 mM sodium citrate (pH 7.0)) containing 0.1% SDS. The membrane was then wrapped in plastic film and exposed to a phosphor-screen (Molecular Dynamics) for approximately 2 hr. Screens were scanned using a FLA5000 Fuji PhosphoImager. Images were analysed with AIDA^{TM} analysis software (Raytest GmbH) and Adobe Photoshop^{TM} CS3. If necessary, membranes were stripped with boiling 0.5% SDS and re-probed with a different probe.

2.5 Protein Methods

2.5.1 Determination of protein concentration

The concentration of protein in a sample was determined by Bradford assay. BSA protein standards between 0.05-10 ug/mL were used to generate a standard curve. 1 mL 20% Bradford Solution (Biorad) was mixed with two different dilutions of each sample and incubated at RT for 5'. The absorbance at 595 nm (A_{595}) was measured using an Eppendorf Biophotometer and concentration determined by comparison to the BSA standard curve.

2.5.2 SDS-PAGE Electrophoresis

Protein samples were added to 1 volume (or 15 µL, if total volume of sample is <15 µL) SDS-PAGE loading buffer (5% SDS, 10% glycerol, 100 mM Tris-HCl (pH 6.8), 0.025% Bromophenol Blue, supplemented immediately before use with 100 mM DTT) and boiled at 98°C for 5'. Samples were then loaded onto mini SDS-PAGE gels with 10% (20-100 kDa proteins) or 8% (100-200 kDa proteins) resolving gel layers and 6% stacking gel layers prepared according to the recipes in Table 2.1. Gels were run at 135 V in 1x SDS running buffer (25 mM Tris, 190 mM glycine, 0.1% SDS).

Table 2.1: SDS-PAGE gel recipes

Ingredient	10% Resolving Gel	8% Resolving Gel	6% Stacking Gel
1 M Tris (pH 8.8)	3.75mL	3.75mL	
1 M Tris (pH 6.8)			375µL
acrylamide:bisacrylamide (37.5:1), 30%	3.35mL	2.7mL	500µL
dH_2O	2.75mL	3.4mL	2.075mL
10% SDS	100µL	100µL	30µL
10% APS	50µL	50µL	15µL
TEMED	16µL	16µL	3µL

2.5.3 Total protein staining

SDS-PAGE gels were stained for total protein with Coomassie stain (0.1% Coomassie brilliant blue, 10% acetic acid, 20% methanol), removed by destaining solution (10% acetic acid, 40% methanol).

2.5.4 Western blotting

SDS-PAGE gels were soaked for 15' in transfer buffer (25 mM Tris (pH 8.0), 192 mM glycine, 20% ethanol, 0.0375% SDS) along with an equal size piece of Hybond™ nitrocellulose membrane (Amersham). After soaking, gel and membrane were placed between two pieces of moistened filter paper and transferred with a semi-dry blotter (Hoefer) at 20 V for 30' (small proteins) or 45' (large proteins). Transfer efficiency was checked by Ponceau S staining. Membranes were blocked with 1x TBS-T (50 mM Tris-HCl (pH 7.6), 150 mM NaCl, 0.5% Tween-20) containing 5% milk for 1 hr at RT with gentle shaking. To detect the NTAP-tag, 1 µL of HRP-conjugated rabbit α-mouse polyclonal antibody (diluted 1:1000) was added to 10 mL TBS-T containing 5% milk and incubated on the membrane for 1 hr at RT with gentle shaking. The membrane was then washed three times with 1x TBS-T for 5' each time at RT with gentle shaking. The washed membrane was then incubated with ECL solution (Amersham) for 1'; the HRP-antibody conjugate oxidised the ECL solution luminol to emit chemiluminescence, which was detected by exposure to Hyperfilm™ X-ray film (Amersham).

2.5.5 GST-Tagged purification of proteins from E. coli

E. coli BL21 (DE3) (pLysS) cells, transformed with pGEX-6P-1 containing the appropriate insert (for a full list of vectors used in this study, see Appendix C), were grown overnight in 5 mL LB containing ampicillin (100 ug/mL) and chloramphenicol (34 µg/mL). 2 mL of this saturated culture was diluted into 200 mL of LB containing ampicillin (100 µg/mL) and grown at 37°C with shaking to an A_{600} of ~0.5. This culture was equilibrated to the induction temperature of 24°C for 45' with shaking and then induced with 100 µM

IPTG. After 16 hr incubation, cells were harvested by centrifugation (4,000 x *g*, 15', 4°C) and pellets were stored at -20°C for a minimum of 2 hr. After thawing on ice, pellets were resuspended in 5 mL per gram of cells BugBuster® Protein Extraction Reagent (Novagen) supplemented with 0.5 µL/mL benzonase nuclease (Novagen). Samples were fully resuspended by vigorous pipetting then incubated at RT for 15' to complete lysis. The lysates were then clarified by centrifugation (16,000 x *g*, 20', 4°C). The supernatants were transferred to a fresh 15 mL tube and incubated with 120 µL pre-washed and BSA-blocked glutathione-Sepharose beads (Amersham) for 2 hr on a rotating wheel. The beads were washed three times with 900 µL Washing Buffer (50 mM Tris (pH 7.8), 150 mM KCl, 1 mM EDTA) containing 1 mM DTT then incubated with 1 mL of Washing Buffer containing 1 mM DTT, 5 mM ATP and 20 mM $MgCl_2$ for 10' at 4°C on a rotating wheel to elute bacterial DnaK bound nonspecifically to the beads. Beads were washed once more with 900 µL washing buffer containing 1 mM DTT then washed once with 900 µL 3C Protease Buffer (Washing Buffer containing 2 mM DTT and 0.05% Triton X-100). Beads were then resuspended in 120 µL 3C Protease Buffer containing 2 µL 3C Protease and incubated overnight at 4°C on a rotating wheel to cleave the recombinant protein from the bead-bound GST tag. Eluate was collected and incubated for 15' at 4°C on a rotating wheel with a further 30 µL pre-washed (not BSA-blocked) glutathione-sepharose beads to remove residual GST and GST-protein contaminants. Eluate was collected again and purified protein was quantified as described in 2.5.1. Several aliquots of protein of 5-30 µL were frozen on dry ice for 5' then moved to -80°C for storage. A list of plasmids used to express GST-tagged proteins in *E. coli* is provided in Appendix B.

2.5.6 NTAP-tagged purification of proteins from S. pombe

S. pombe strains transformed with the pREP1-NTAP vector containing the appropriate insert (for list of vectors used in this study, see Appendix C) were grown overnight at 30°C with shaking in EMM lacking leucine, supplemented with thiamine (30 µM) to repress transcription from the vector *nmt1* promoter.

This saturated culture collected by centrifugation (4,000 x *g*, 2', 4°C) and washed four times with 1 mL dH_2O to remove residual repressive thiamine. The washed cells were diluted to an A_{600} of ~0.1 in 100 mL EMM lacking leucine and grown at 30°C with shaking for 18 hr. Cells were harvested by centrifugation (4,000 x *g*, 2', 4°C) and washed twice with 50 mL of dH_2O. Cells were resuspended in chilled 2 mL TAP Purification Lysis Buffer (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.5% Igepal CA-630, 1 mM DTT, 1 mM PMSF, 1x Sigma Protein Inhibitor Cocktail) and split into four chilled screw-cap tubes containing 250 µL RNAse-free Glass Beads (Sigma). Cells were lysed in the FastPrep Homogenizer and Isolation System (GMI) with 5 cycles of shaking at 7000 rpm for 30", with 3' incubation on ice between cycles. Efficient lysis (>80%) was confirmed by light microscopy. Cell lysates was clarified by centrifugation (16,000 x *g*, 20', 4°C) and pooled with 100 µL pre-washed Ig-Sepharose beads. Beads were incubated with lysate for 2-3 hr at 4°C on a rotating wheel. Beads were collected and washed six times with 900 µL TAP Purification Washing Buffer (50 mM Tris-HCl (pH 7.8), 150 mM KCl, 1 mM EDTA, 1 mM DTT). Beads were resuspended in 40 µL TAP Purification Washing Buffer and successful NTAP-tagged protein purification was confirmed by western blotting. For smaller purified proteins, beads could be stored for <72 hr at 4°C but beads carrying larger proteins (especially NTAP-Cid16) were used immediately for down-stream experiments. A list of plasmids used to express NTAP-tagged proteins in *S. pombe* is provided in Appendix B.

2.6 Microarray Methods

All microarray analyses in this thesis were conducted in collaboration with Dr. Juan Mata and his lab in the Department of Biochemistry, University of Cambridge. Labelling, hybridisation and scanning of *cid16Δ* expression array samples was kindly conducted by Ayesha Hasan in the Mata Lab using RNA samples prepared at the University of Oxford.

2.6.1 Transcriptome-wide RNA half-life microarray analysis

A schematic representation of the protocol for 4-thiouridine-mediated labelling and purification of nascent RNA transcripts is provided in Figure 2.2. The protocol is used as previously described (Dölken *et al.* 2008) but with optimisation for use with *S. pombe* cells. In this protocol, total RNA and nascent RNA transcribed during a short incubation are separately purified from prototrophic *wt* and *cid1Δ* strains. These RNA samples are labelled with complementary Alexa fluorophores and ratios of nascent:total RNA content determined for each transcript in the transcriptome. These ratios could be used to determine the half-lives of transcripts using the formula:

$$t_{1/2} = -t_L \times (\ln(2) / \ln(1 - R))$$

where t_L is the duration of the 4-SU pulse (15' in these experiments) and R is the ratio of nascent:total RNA for each transcript in each strain (Dölken *et al.* 2008). Half-life changes were considered significant if (a) $t_{1/2}$ was >1.4x different in *cid1Δ* vs. *wt* strains in two separate biological replicates of nascent vs. total RNA microarrays; and (b) expression of the transcript was >2x different in *cid1Δ* vs. *wt* strains in two separate dye-reversed expression arrays run in parallel with the half-life arrays (see Table 2.2).

Arrays used in this experiment were designed by the Mata Lab, University of Cambridge and produced by Agilent. Arrays contain probes for all annotated ORFs in the *S. pombe* genome as well as a selection of probes for non-coding RNAs, antisense transcripts and intergenic regions; however, due to an error in production by Agilent, the non-ORF probes targeting sequences from chromosomes II and III were synthesised incorrectly and consequently were excluded from analysis (Dr. Juan Mata, pers. comm.).

2.6.1.1 Synthesis and purification of 4-thiouridine-containing nascent transcripts in S. pombe

cid1$^+$ and *cid1::kanMX* alleles were constructed in a prototrophic background to ensure the cells were able to grow in unsupplemented minimal media and take up amino acids from the growth medium.

Cells were grown overnight in unsupplemented EMM media then diluted into 50mL unsupplemented EMM media and grown at 32°C with shaking until OD_{600} = 1.0 (total $5x10^8$ cells). 4-thiouridine (4-SU) (Sigma-Aldrich) was spiked into the media to a final concentration 75 µg/mL and cells were grown for a further 15' to incorporate 4-SU into actively transcribing RNA. After the 4-SU pulse, cells were pelleted by centrifugation (4,000 x *g*, 2', 4°C) and immediately frozen at -80°C until further processing. To extract RNA, cells are thawed on ice and washed once with 750 µL dH_2O before purification of RNA as described in 2.4.1.1. RNA was quantified as described in 2.4.2 and 200 µg of total RNA (containing both 4-thiouridine labelled and unlabelled RNA) was purified using an RNeasy column (QIAgen) according to the manufacturer's protocol. Briefly, 200 µg of RNA was made up to 100 µL with dH_2O and added to 350 µL Buffer RLT and 250 µL 100% ethanol. The total 700 µL volume was loaded onto an RNeasy Mini spin column and centrifuged (16,000 x *g*, 30", RT) to bind RNA to the column. The column was washed twice with 500 µL Buffer RPE then centrifuged (16,000 x *g*, 1', RT) again to remove residual ethanol. RNA was eluted twice with 50 µL dH_2O and quantified as described in 2.4.2. Purified RNA was quantified using a Nanodrop machine and 20 µg aliquots were diluted into 12 µL and stored at -80°C until labelling.

100 µg total RNA was biotinylated to tag the sulfhydryl groups of the 4-SU-labelled RNAs with a biotin moiety. RNA was combined with 200 µL of 1 mg/mL EZ-Link® Biotin-HPDP (Thermo Scientific) dissolved in DMSO, 10 µL 1M Tris-HCl (pH 7.6) and 2 µL 0.5 M EDTA and made up to a total volume of 1 mL with dH_2O. The biotinylation reaction was incubated at RT for 90'. The reaction was then split into two aliquots of 500 µL each and each was combined with an equal volume of chloroform: isoamyl alcohol (24:1) in a phase-lock tube. The two aliquots were mixed by inversion and centrifuged (16,000 x *g*, 15', RT). The two upper (aqueous) layers were pooled in a 2 mL Eppendorf tube containing 0.1 volumes of 5 M NaCl and 1 volume of isopropanol. Samples were incubated at -80°C for 30' to precipitate RNA which was collected by centrifugation (16,000 x *g*, 15', RT). Pellets were washed with 500 µL 75% ethanol, air dried and resuspended in 20 µL dH_2O.

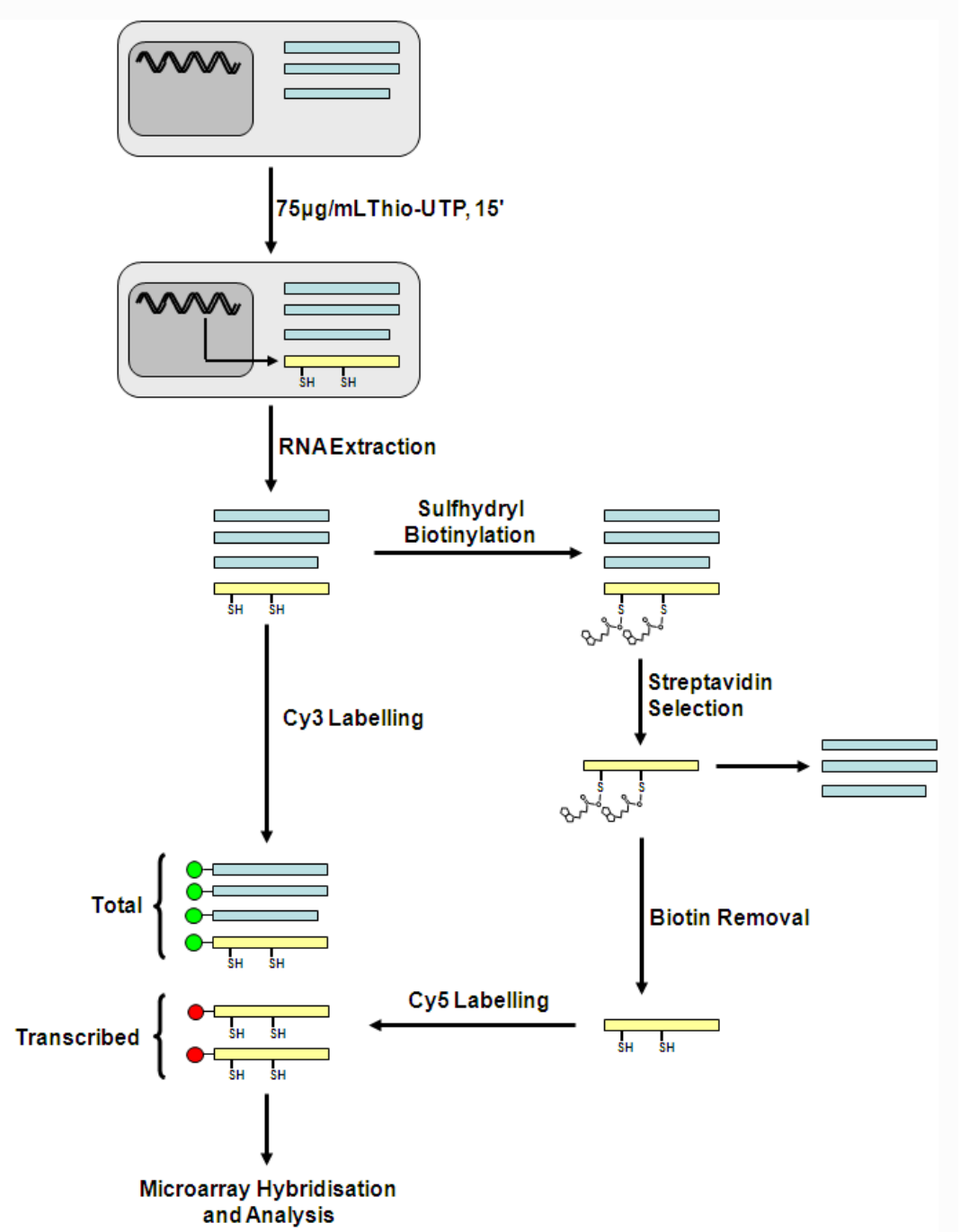


Figure 2.2: A Schematic representation of the 4-thiouridine labelling and whole-transcriptome RNA half-life assay protocol. Exponentially-growing prototrophic *S. pombe* are pulsed for 15' with 4-SU and total RNA extracted. This RNA is separated into "total" and "transcribed" pools. The nascently-transcribed, thiolated RNAs in the "transcribed" pool are cross-linked to biotin, purified by streptravidin selection and released from biotin. The "total" and "transcribed" RNAs are then labelled with opposing fluorescent dyes and the ratios of transcribed:total RNA for each transcript determined by competitive hybridisation on a custom microarray (Agilent). This work was performed in the laboratory of Dr. Juan Mata, Department of Biochemistry, University of Cambridge.

To purify biotinylated RNA, 100 µL M-280 Streptavidin Dynabeads® (Invitrogen) were washed three times with 500 µL MPG buffer (100 mM Tris-HCl (pH 7.6), 1 M NaCl, 10 mM EDTA) and resuspended in 400 µL MPG buffer. 20 µL of 10 mg/mL tRNA was added and beads were incubated at RT for 30' with gentle rocking to block non-specific interactions. Beads were washed a further three times with MPG buffer then resuspended in MPG buffer prior to use. RNA products from the biotinylation reaction were incubated at 65°C for 5' then cooled to RT for 1' and added to the tRNA-blocked streptavidin beads. Beads were incubated at RT for 40' with gentle rocking. After incubation, beads were washed three times with 500 µL MPG buffer pre-heated to 65°C then three times with 500 µL MPG buffer at RT. Beads were then incubated sequentially with two aliquots of 50 µL 100 mM DTT at RT for 5' to reverse the sulfhydryl-biotin reaction and elute 4-SU-labelled RNA. The two aliquots were pooled and RNA was purified using RNAqueous micro columns (Ambion). Briefly, 500 µL lysis buffer and 250 µL 100% ethanol were added to the 100 µL pooled RNA aliquots and mixed thoroughly. The samples were loaded onto an RNAqueous micro column and passed through the column by centrifugation (16,000 x *g*, 30", RT). The column was washed once with 180 µL Wash 1 Buffer then twice with 180 µL Wash 2/3 Buffer. 8 µL of dH_2O pre-heated to 85°C was loaded onto the column, incubated for 1' and eluted by centrifugation (16,000 x *g*, 1', RT). This elution was repeated with a further 8 µL of 85°C dH_2O and the two aliquots were pooled. 1 µL of the ~13 µL eluate was used to quantify the biotin-purified RNA and the remaining 12 µL (containing approximately 1-2 µg RNA) was stored at -80°C until labelling.

2.6.1.2 Cy3-/Cy5-labelled cDNA synthesis and purification

Two 20 µg aliquots each of *wt* and *cid1::kanMX* total RNA and the biotin-purified nascent RNA samples from each genotype prepared in 2.6.1.1 were used to synthesise cDNAs containing dCMP labelled with Alexa-647 or Alexa-555 fluorescent dyes using the SuperScript Plus Indirect Labelling Kit (Invitrogen) according to Table 2.2. Briefly, RNA samples are combined with 2 µL anchored oligo$(dT)_{20}$ primer

(2.5 µg/µL) and 1 µL Random Hexamers (0.5 µg/µL) and made up to a total volume of 18 µL. Samples were incubated at 70°C for 5' then immediately chilled on ice. Samples were supplemented with 1x First Strand buffer, 5 mM DTT, 1.3 U/µL RNAseOUT™ Ribonuclease Inhibitor, 26 U/µL SuperScript III™ Reverse Transcriptase and 0.1 volumes of dATP/dTTP/dGTP mix containing either 0.5 mM dCTP-(Alexa-555) or 0.5 mM dCTP-(Alexa-647) according to Table 2.2. Reactions were mixed gently and incubated at 46°C for 3 hr to synthesise Alexa fluorophore-labelled cDNA. As Alexa fluorophores are sensitive to photobleaching, this reaction and all following it were conducted in the dark wherever possible.

After cDNA synthesis, cDNA samples of opposing fluorophore labelling for comparison on microarrays (see Table 2.2) were pooled and template RNA was hydrolysed by addition of 0.5 volumes 0.1 M NaOH and incubation at 70°C for 30'. Following alkaline hydrolysis, samples were neutralised with a further 0.5 volumes of 0.1 M HCl. cDNAs were then purified by addition of 700 µL Binding Buffer and centrifugation (3,300 x *g*, 1', RT) through a Low-Elution Volume Spin Cartridge. Column-bound cDNAs were washed with 600 µL Wash Buffer and eluted with 20 µL of DEPC-treated dH_2O by centrifugation (16,000 x *g*, 1', RT). Synthesis of each Alexa fluorophore-labelled cDNA and their relative concentrations in each pooled sample was confirmed by Nanodrop quantitation of A_{555}, A_{650} and A_{555}:A_{650} ratio respectively.

Pooled purified cDNA was ethanol precipitated by addition of 0.1 volumes 3 M NaAc (pH 5.2) and 3 volumes chilled 100% ethanol and incubation at -80°C for 30'. Pellets were collected by centrifugation (16,000 x *g*, 15', RT) and washed with 500 µL chilled 70% ethanol, air dried and resuspended in 44 µL DEPC-treated dH_2O. RNA samples were mixed with 11 µL 10x Blocking Agent (Agilent) and 55 µL Hybridisation Buffer (Agilent) then incubated at 95°C for 5' and cooled to RT prior to microarray hybridisation.

Table 2.2: Alexa fluorophore labelling of cDNA samples in transcriptome-wide RNA half-life assays

Array *(sample)*	**Replicate 1**	**Replicate 2**
***wt* nascent vs. total**		
wt total RNA	Alexa-555	Alexa-555
wt 4-SU RNA	Alexa-647	Alexa-647
***cid1Δ* nascent vs. total**		
cid1Δ total RNA	Alexa-555	Alexa-555
cid1Δ 4-SU RNA	Alexa-647	Alexa-647
***wt* vs. *cid1Δ* total**		
wt total RNA	Alexa-555	Alexa-647
cid1Δ total RNA	Alexa-647	Alexa-555

2.6.1.3 Microarray hybridisation, scanning and analysis

110 µL of hybridisation-ready cDNA samples prepared in 2.6.1.2 were loaded onto microarrays such that each combination of cDNAs was in an individual array well. The arrays were sealed and incubated at 65°C for 17 hr with 10 rpm rotation to allow cDNAs to hybridise to the microarray probes. Hybridised microarrays were disassembled in RT Wash Buffer 1 (Agilent) then washed for 1' in RT Wash Buffer 1 and 1' in 37°C Wash Buffer 2 (Agilent). Slides were finally rinsed briefly in 100% acetonitrile to remove residual buffer and dried prior to scanning.

Microarrays were scanned at the Department of Pathology Microarray Facility, University of Cambridge. Alexa fluorophores were quantified with excitation at λ=555/650nm and measurement of emission at λ=565/665nm. Raw data were analysed and gene lists created with GeneSpring microarray analysis software (Agilent). Genomic distribution analyses were conducted manually using the gene annotations curated on http://www.pombase.org/ (Wood *et al.* 2012). Gene Ontology (GO) enrichment analyses were conducted with the g:Profiler internet-based GO analysis toolkit (Reimand *et al.* 2007; Reimand *et al.* 2011) using GO annotations curated on http://www.pombase.org/ (Aslett and Wood 2006).

2.6.2 Transcriptome-wide expression analysis

Transcriptome-wide expression analyses were conducted using prototrophic *wt* and *cid16::kanMX* strains as described for the control expression arrays in 2.6.1.1-2.6.1.3. Two biological replicates grown on different days were used for dye-swapped analysis and analysed as described in 2.6.1.3. Microarray labelling, hybridisation and scanning were kindly conducted by Ayesha Hasan in the Mata Lab, Department of Biochemistry, University of Cambridge.

Chapter 3
Cid16 is the second S. pombe mRNA 3' uridylyltransferase

3.1 Summary

In the course of investigations into the activity of Cid1, Rissland *et al*. (2007) previously observed the presence of four 3' terminal U tails on *act1* mRNA ($A_{30}U$, $A_{10}U$, U and U_3) in *cid1Δ* strains of *S. pombe*, suggesting that another enzyme, most likely a member of the Cid1 family, may be able to uridylylate mRNA at a low frequency (4.3% in that study). The presence of a second mRNA uridylyltransferase suggests that additional uridylylation-regulated pathways may exist in *S. pombe*; alternatively, Cid1 and the second uridylyltransferase may be partially or completely functionally redundant and thus deletion of either one may not reveal the full phenotypic effects of a loss of mRNA uridylylation.

Given this observation, a battery of double mutants of *cid1Δ* and other members of the Cid family in *S. pombe* (*cid11-cid14Δ* or *cid16Δ*) were here constructed, with or without the additional *lsm1Δ* deletion previously shown to cause accumulation of capped uridylylated mRNAs (Rissland and Norbury 2009). Additionally, NTAP-tagged Cid family members were overexpressed in a *cid1Δ* background. These strains were screened for loss of residual uridylylation activity (deletion mutants) or accumulation of uridylylated mRNAs (overexpression strains) mediated by the putative second uridylyltransferase. Residual uridylylation was lost in multiple Cid-family deletions in the *lsm1Δ cid1Δ* and *cid1Δ* backgrounds, suggesting the activity of the second uridylyltransferase may depend on multiple pathways, whilst overexpression of Cid-family members other than Cid1 itself did not increase the rate of uridylylation, suggesting multiple factors may be limiting for the activity of the

second uridylyltransferase. Though these methods limited the identity of the minor uridylyltransferase to Cid12 or Cid16, final identification was not possible.

An alternative approach, involving purification of NTAP-tagged Cid family members expressed in *S. pombe*, demonstrated that, whilst activity could not be determined for NTAP-Cid12, NTAP-Cid16 possessed highly specific and processive poly(U) polymerase activity. This observation was supported by a detailed consideration of catalytically active and conserved residues, which suggested that: (a) Cid16 active site architecture is supportive of PUP activity, while the Cid12 site is more selective for PAP activity; and (b) Cid12/Cid14 architecture probably differs significantly from those of other Cid family members, suggesting a different catalytic/structural mechanism. From these observations, it is concluded that Cid16 is a second *S. pombe* RNA 3' uridylyltransferase. Implications of the activity of Cid16 in different mutants are discussed, and the mutagenic studies are examined with respect to the recently described Cid1 structure (Lunde *et al.* 2012; Munoz-Tello *et al.* 2012; Yates *et al.* 2012).

3.2 A second mRNA poly(U) polymerase in *S. pombe*

Previous evidence for the existence of a second uridylyltransferase in *S. pombe* was limited to the observation of four 3'-terminal uridylyl tails in *cid1Δ* cells ($A_{30}U$, $A_{10}U$, U and U_3 respectively) from 92 sequenced samples (Rissland *et al.* 2007). To first confirm that a second uridylyltransferase was active in *S. pombe* cells, circularised *act1* transcript sequences collected below from all strains containing the *cid1Δ* allele in combination with a deletion of a second Cid family member were pooled and analysed cumulatively.

In all cases, RNA was purified from *S. pombe* strains in mid-log phase in rich media and 3'/5' ends of capped *act1* mRNA were sequenced using the circularised rapid amplification of cDNA ends (cRACE) technique (Couttet *et al.* 1997). 12 µg total RNA was sequentially treated with shrimp alkaline phosphatase (SAP) and tobacco acid pyrophosphatase (TAP) to dephosphorylate decapped RNA 5'

ends and cleave the pyrophosphate bond of the methyl-7-guanosine cap respectively, resulting in an RNA sample containing ligatable 5'-monophosphorylated ends only on transcripts derived from capped mRNAs. These were intramolecularly ligated 5'-to-3' to form circular mRNA molecules and ligation sites of capped *act1* mRNAs were amplified by reverse transcription and nested PCR with divergent primers. Following cloning and sequencing to determine the positions of the 5' and 3' ends, 3' non-templated additions were identified and analysed.

A total of 33 new 3'-terminal uridylylation events were observed among 770 *act1* mRNA sequences from *cid1Δ* strains, ranging in length from U_1 to U_{12} and occurring both on polyadenylated and on deadenylated transcripts (Table 3.1, Figure 4.3A; see Appendix D for full list of sequences retrieved). Three separate analyses converge to demonstrate that these events are the product of a second *S. pombe* uridylyltransferase:

Firstly, in the non-templated nucleotide population, UMP residues at the 3' terminus of transcripts occurred at a significantly higher frequency than those within non-templated poly(A) tails which were attributed to PCR-mediated mutagenesis or occasional non-AMP incorporation by nuclear poly(A) polymerase (Figure 3.1A). Furthermore, the 3'-terminal UMP residues were more abundant in the absence of *lsm1*, while internal UMP residues were not (Figure 3.1A), confirming the biogenesis of the observed oligo(U) tails must be differentially regulated in response to the genotype of the source culture. Importantly, since these analyses do not take into account the fact that the proposed second uridylyltransferase-specific tails are found only at the extreme 3'-terminal and not distributed through the poly(A) tail, the statistical tests used here are likely to underrepresent the significance of these observations.

Secondly, analysis of the relative proportions of non-AMP nucleosides internal or terminal to the poly(A) tail shows a highly significant (χ^2=393.1, df=2; $p < 10^{-50}$) enrichment of the UMP proportion in the 3'-terminal pool compared to the internal pool (Figure 3.1B), suggesting that the two non-AMP populations were generated by discrete mechanisms with different nucleotide preferences.

Table 3.1: Key statistics for cRACE sequencing datasets by genotype. Sequencing data collected in sections **3.3**, **3.4** and **3.5** below is here summarised. Raw sequences are presented in Appendix D.

Genotype	*N*	% oligo(U)*	Mean oligo(U) tail length (*nt*)	% poly(A)†	Mean poly(A) tail length (*nt*)
lsm1Δ cid1Δ	54	12.9	2.57 ± 1.62	70.3	11 ± 10.3
lsm1Δ cid1Δ cid11Δ	32	18.8	3.33 ± 2.58	59.4	8.9 ± 11.9
lsm1Δ cid1Δ cid12Δ	87	2.3	1 ± 0	56.3	10.4 ± 12.1
lsm1Δ cid1Δ cid13Δ	64	7.8	4.4 ± 4.51	71.8	11.4 ± 11.1
lsm1Δ cid1Δ cid14Δ	58	0	N/A	51.7	6.1 ± 7.8
lsm1Δ cid1Δ cid16Δ	51	3.9	1 ± 0	60.8	9.3 ± 10.5
TOTAL	346			61.7	9.6 ± 10.9
cid1Δ cid12Δ	105	2.9	1 ± 0	91.4	18.2 ± 10.8
cid1Δ cid13Δ	102	4.9	2.4 ± 1.67	90.1	19.8 ± 13.4
cid1Δ cid14Δ	105	2.9	1.67 ± 1.15	82.9	16.5 ± 12.9
cid1Δ cid16Δ	49	0	N/A	91.8	20.3 ± 11.2
TOTAL	361			89.1	18.1 ± 12.3
*cid1Δ (EV**)*	17	11.8	1 ± 0	100	22.9 ± 9.3
cid1Δ (cid1 o/e)	54	44.4	1.21 ± 0.41	87.0	27.2 ± 13.1
cid1Δ (cid12 o/e)	44	4.5	1 ± 0	97.7	20.2 ± 7.9
cid1Δ (cid13 o/e)	13	0	N/A	76.9	16.5 ± 10.5
cid1Δ (cid14 o/e)	21	0	N/A	90.4	15.9 ± 7.8
cid1Δ (cid16 o/e)	23	0	N/A	95.6	18.3 ± 7.6

* % oligo(U): the proportion of total retrieved sequences containing non-templated UMP residue(s) at the 5'-3' ligation junction.

† % poly(A): the proportion of total retrieved sequences containing non-templated AMP residue(s) at the 3' end of the templated sequence

** EV = empty vector (pREP1-NTAP)

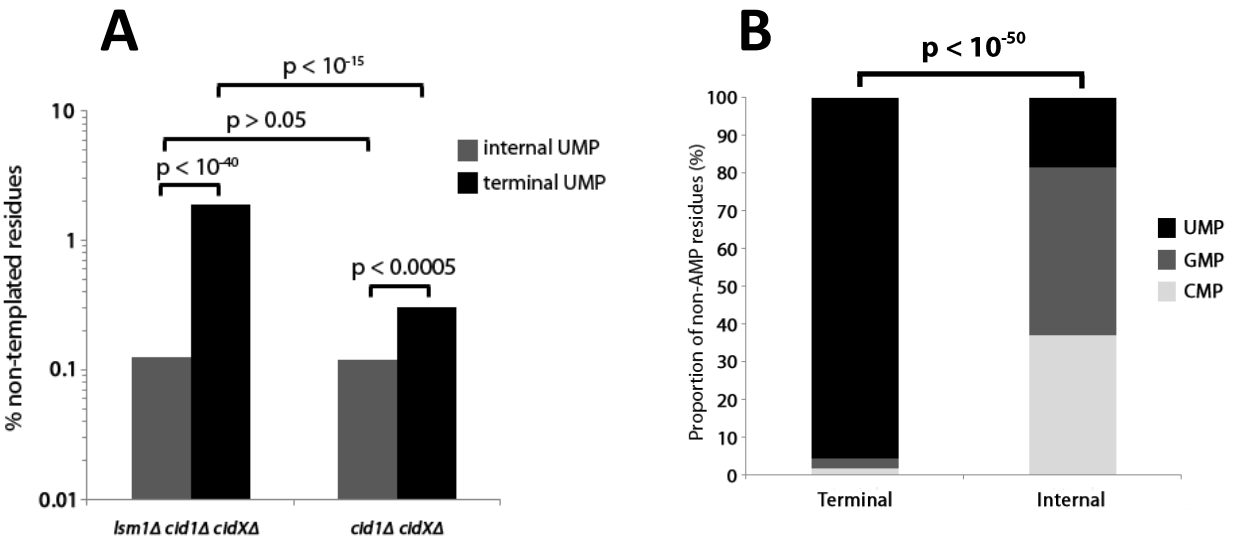


Figure 3.1: A second mRNA 3' uridylyltransferase is active in *S. pombe* cells. cRACE 5'-3' end sequencing data from a series of *lsm1Δ cid1Δ* and *lsm1*$^{+}$ *cid1Δ* strains containing additional deletions in the other *cid* family members were pooled and analysed cumulatively to provide a large sample set for the identification of second uridylyltransferase activity. (A) The frequency of non-templated UMP residues either within poly(A) tails or at the 3' terminus of transcripts were determined with respect to the status of the *lsm1* locus and population sizes compared using Pearson's χ^2 test. *cidXΔ* indicates a deletion in a second *cid* family member. (B) The relative nucleoside composition of the pools of non-templated residues other than AMP were compared depending on their positioning within poly(A) tails (internal) or at the 3' terminus of transcripts (terminal).

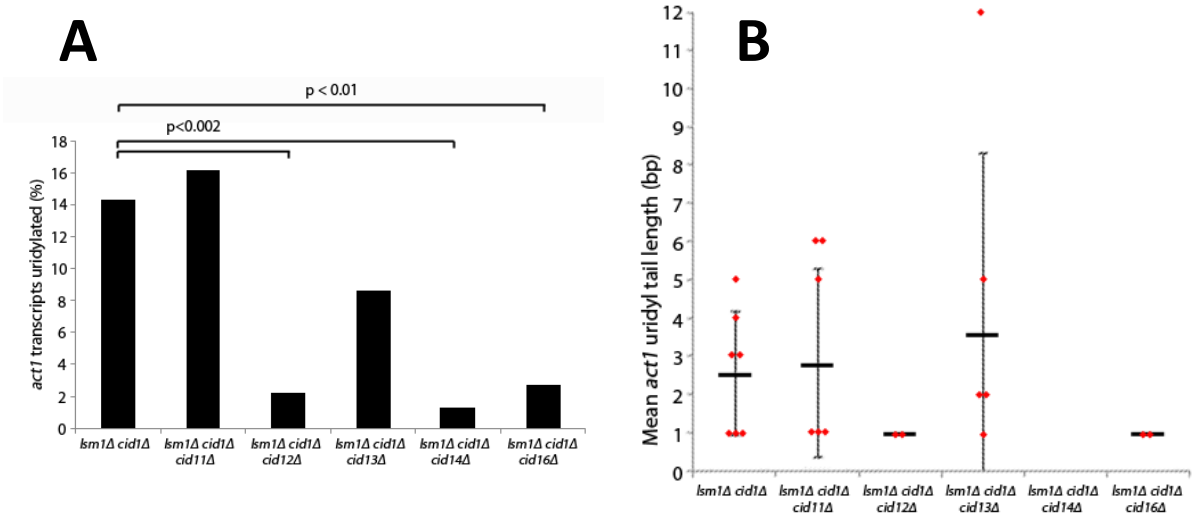
Finally, the canonical *act1* transcriptional start site (TSS) at position -57 with respect to the start codon is immediately preceded by two dTMP residues at positions -58 and -59; one complication of the cRACE methodology is that, due to the circularisation reaction ligating the 5' and 3' ends of individual mRNAs, the presence of one or two UMP residues at the 5'-3' ligation junction could derive from either the 3' or the 5' end and consequently represent either 3' terminal mono-/di-uridylylation or transcription initiation at positions -58/-59 respectively. Here, as in previous studies (Rissland *et al.* 2007), the incidence of apparent -58/-59 initiation is dependent on the activity of several 3' end-modifying/interacting enzymes, namely Cid1, Lsm1 and Dcp1, suggesting that this variation represents authentic 3' end uridylylation rather than single base-pair variation in TSS. Furthermore, 36.3% of uridylylation events catalysed by the second uridylyltransferase are 3 nt or greater in length and thus must unequivocally derive from the addition of at least one UMP residue to the 3' terminal of capped mRNA. From these multiple strands of evidence, it is concluded that a second uridylyltransferase is active at low frequency on *act1* transcripts in *S. pombe*.

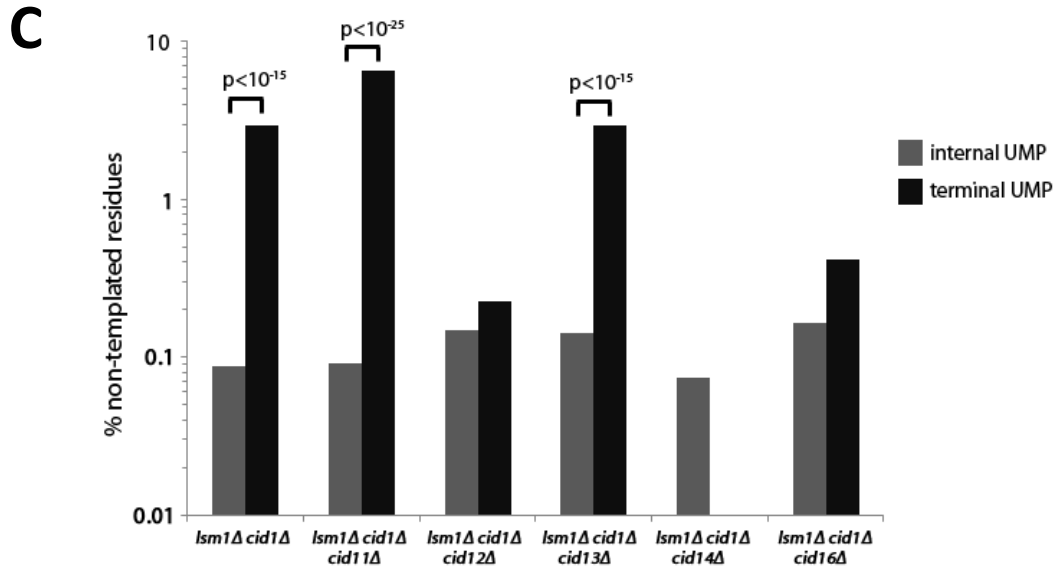
3.3 *act1* cRACE analyses of *lsm1Δ cid1Δ cidXΔ* genotypes

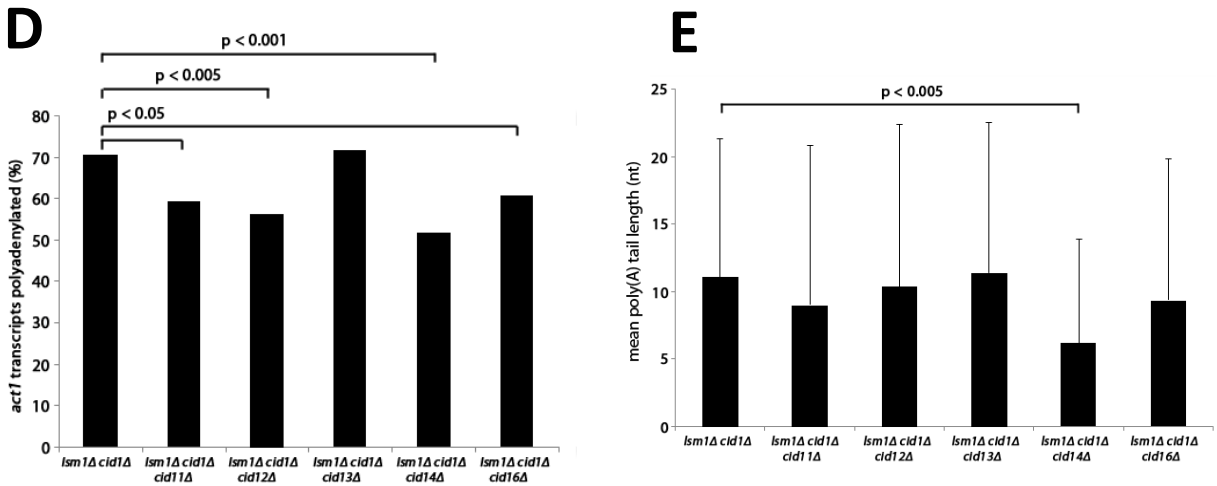
Given the low frequency of Cid1-independent uridylylation in a previous study (4.3% of *act1* transcripts in *cid1Δ* cells; Rissland *et al.* 2007) and the observation that deletion of the decapping activator *lsm1* significantly increased the abundance of uridylylated *act1* mRNAs (25 to 55%; Rissland and Norbury 2009), a battery of *lsm1Δ cid1Δ cidXΔ* mutants were constructed, where *cidX* represents each of the five other Cid family members in *S. pombe*, in order to enrich specifically for the products of the second uridylyltransferase activity and sequentially screen other Cid family member deletions for loss of this residual activity. *lsm1::kanMX* and *cid1::LEU2* deletion strains were crossed and kan^R leu^+ progeny isolated by replica plating of tetrad-dissected asci to restrictive media. These *lsm1Δ cid1Δ* strains were then crossed with individual $ura4^+$ deletion mutants of each of the five remaining Cid family members in *S. pombe* (*cid11-cid14, cid16*) and kan^R leu^+ ura^+ triple-positive mutants selected as above (a full list of strains used in this study is provided in Appendix A). All genotypes were confirmed by back-crossing against *wt* to re-isolate individual alleles with

Mendelian distribution and by colony PCR-based amplification of products spanning sites of recombination for each allele. All triple-mutant strains showed a slow-growth defect consistent with the effect of the *lsm1Δ* allele (Rissland and Norbury 2009). Additionally, the *cid12Δ* and *cid14Δ* deletions, both of which exhibit independent slow-growth phenotypes, caused additive growth defects with *lsm1Δ*, making the *lsm1Δ cid1Δ cid12Δ* and *lsm1Δ cid1Δ cid14Δ* triple mutants particularly slow-growing.

Between 32 and 87 *act1* capped mRNA sequences were retrieved for each of the six genotypes examined (*lsm1Δ cid1Δ* alone and *lsm1Δ cid1Δ cid11-14,16Δ*; see Table 3.1). Analysis of poly(A) tail status from each of the six genotypes recapitulated previous observations, confirming the veracity of the experimental procedure – pooled sequences from strains carrying the *lsm1Δ* allele showed significantly reduced polyadenylation (61.7% vs. 89.1%) and length (9.6 ± 10.9nt vs. 18.1 ± 12.3nt; $p < 10^{-20}$ by Student's t-test) when compared to pooled *lsm1*$^{+}$ strains, consistent with previous observations of increased deadenylation in *lsm1Δ* mutants to compensate for impaired decapping-mediated mRNA degradation (Rissland and Norbury 2009). Surprisingly, polyadenylation was decreased (51.7% vs. 70.3%) and mean poly(A) tail length significantly reduced (6.1 ± 7.8nt vs. 11 ± 10.3nt; $p < 0.005$, Students' t-test) for the *lsm1Δ cid1Δ cid14Δ* genotype compared to the *lsm1Δ cid1Δ* background (Figure 3.2D,E). Cid14 has been previously shown to be a nuclear poly(A) polymerase homologous to Trf4p/Trf5p in *S. cerevisiae,* and to add short oligo(A) tails to diverse nuclear RNAs prior to degradation by decapping and/or 3'→5' exonucleolysis (Win *et al.* 2006; Buhler *et al.* 2007; Buhler *et al.* 2008; Wang *et al.* 2008; Lemieux *et al.* 2011). While Cid14-mediated oligoadenylation and 3'→5' exonucleolysis is unlikely to be a significant degradation pathway for polyadenylated mRNA, it is possible that the inactivation of this degradation pathway has an indirect effect on the deadenylation pathway; alternatively, Cid14 may directly or indirectly regulate the expression of a component of the deadenylation-decapping pathway.


A
p < 0.01
p<0.002
act1 transcripts uridylated (%)
lsm1Δ cid1Δ
lsm1Δ cid1Δ cid11Δ
lsm1Δ cid1Δ cid12Δ
lsm1Δ cid1Δ cid13Δ
lsm1Δ cid1Δ cid14Δ
lsm1Δ cid1Δ cid16Δ
B
Mean act1 uridyl tail length (bp)


C
p<10-15
p<10-25
p<10-15
% non-templated residues
internal UMP
terminal UMP
lsm1Δ cid1Δ
lsm1Δ cid1Δ cid11Δ
lsm1Δ cid1Δ cid12Δ
lsm1Δ cid1Δ cid13Δ
lsm1Δ cid1Δ cid14Δ
lsm1Δ cid1Δ cid16Δ


D
p < 0.001
p < 0.005
p < 0.05
act1 transcripts polyadenylated (%)
E
p < 0.005
mean poly(A) tail length (nt)

***(previous page) Figure* 3.2: Activity of the second uridylyltransferase in the *lsm1Δ cid1Δ* genetic background.** Total RNA was purified from a series of *lsm1Δ cid1Δ* strains containing a tertiary deletion in each additional *cid* family member. 12µg RNA was treated successively with SAP and TAP and capped mRNAs intramolecularly ligated with T4 RNA ligase I to form circular RNAs ligated 5' to 3'. *act1* mRNA ligation sites were amplified by reverse transcription and nested PCR using divergent primers and sequenced to retrieve 5' and 3' end positions and the nature of any 3' non-templated additions. Sequences were compared between genotypes for (A) percentage of transcripts possessing 3' oligo(U) tails and (B) length of these oligo(U) tails (red diamonds: individual uridylylation events; bars represent mean tail length ± 1 s.d.). (C) Frequency of non-templated UMP residues contained within poly(A) tails (internal) or at the 3' teminal of transcripts (terminal) were compared for each genotype. Finally, frequency (D) and mean length ± s.d. (E) of non-templated poly(A) tails were compared between genotypes (error bars respresent 1 s.d.). *p*-values were calculated using Pearson's χ^2 test (A, D), Poisson regression analysis (C) and Student's *t* test (E).

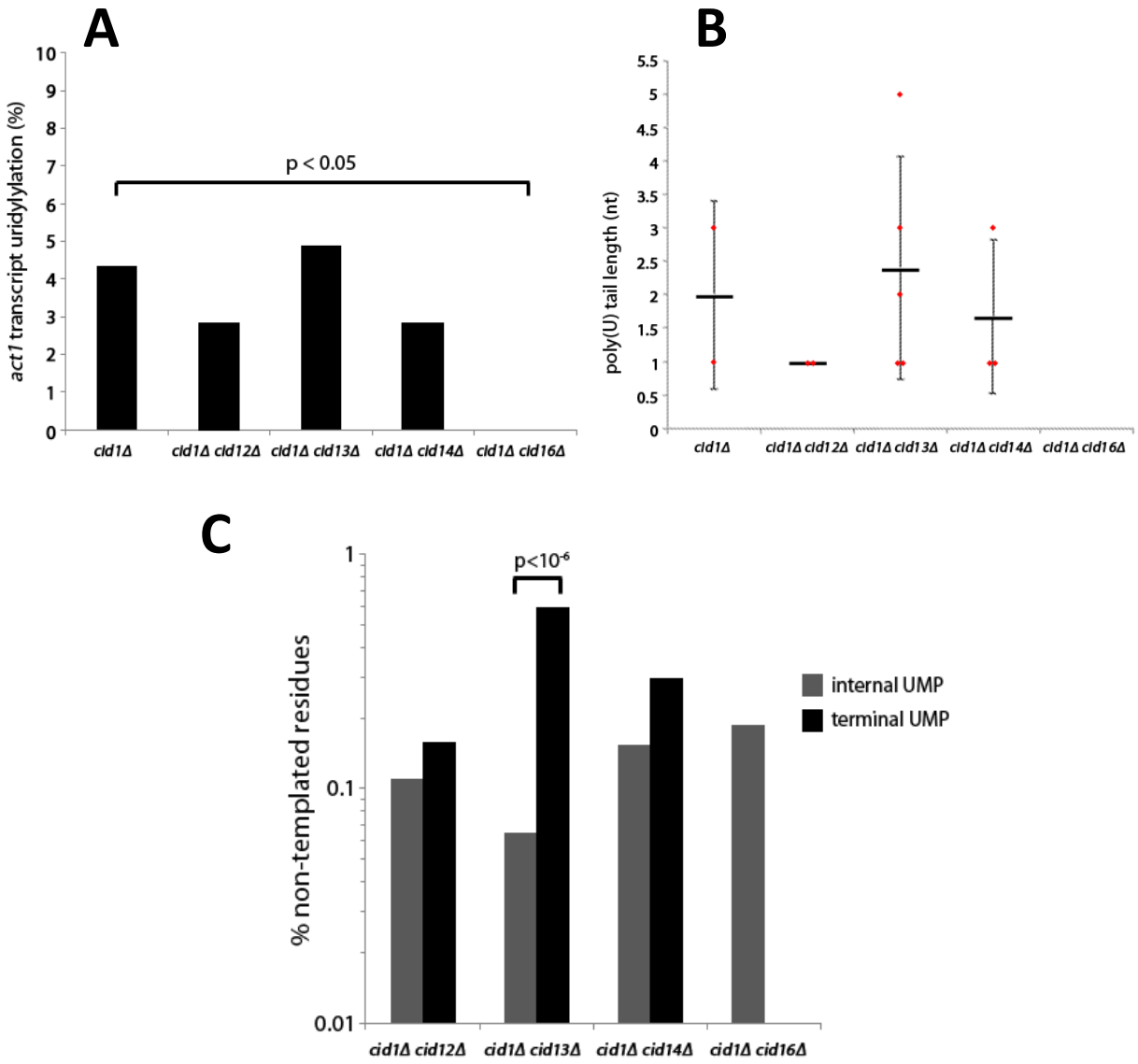
The second uridylyltransferase was found to be active in these assays, with 12.9% of transcripts retrieved from the *lsm1Δ cid1Δ* background possessing 3'-terminal oligo(U) tails, ranging in length from 1-5 nt, an incidence significantly above that observed for internal UMP residues ($p < 10^{-15}$ by Poisson regression analysis; Figure 3.2A-C). Significant enrichment of 3'-terminal UMP residues was also observed in the *lsm1Δ cid1Δ cid11Δ* and *lsm1Δ cid1Δ cid13Δ* genetic backgrounds ($p < 10^{-25}$, 10^{-15} respectively; Poisson regression analysis); oligo(U) tails of up to 12 nt in length occurred at a frequency of 18.8% and 7.8% respectively (Figure 3.2A-C). Surprisingly, *lsm1Δ cid1Δ cid12Δ*, *lsm1Δ cid1Δ cid14Δ* and *lsm1Δ cid1Δ cid16Δ* genotypes all showed a statistically significant reduction in second uridylyltransferase activity compared to the *lsm1Δ cid1Δ* genetic background, exhibiting only 2.3%, 0% and 3.9% addition of exclusively monouridylyl tails ($p < 0.002$, 0.002 and 0.01 respectively by Pearson's χ^2 test; Figure 3.2A-B). Importantly, the presence of exclusively monouridylyl tails prevents unequivocal attribution of these UMP residues to the activity of a uridylyltransferase, as it is possible that they derive exclusively from 5' TSS variation as discussed in **3.2** above. Though *lsm1Δ cid1Δ cid14Δ* alone showed a complete loss of uridylylation, the incidence of 3'-terminal UMP residues for all three genotypes was not significantly different from that of the control internal UMP residues (Figure 3.2C), demonstrating that all three genotypes resulted in the loss of second

uridylyltransferase activity. From these data, though Cid11 and Cid13 can be dismissed as second uridylyltransferase candidates, the identity of the enzyme cannot be unequivocally determined.

3.4 *act1* cRACE analyses of *cid1Δ cidXΔ* genotypes

It was noted during the analysis of the *lsm1Δ cid1Δ cidXΔ* triple-mutant cRACE data (above) that two of the mutants demonstrating loss of the second uridylyltransferase activity, *lsm1Δ cid1Δ cid12Δ* and *lsm1Δ cid1Δ cid14Δ*, exhibited additive growth defects between *lsm1Δ* and *cid12/14Δ* alleles that caused them to grow significantly more slowly than the other genotypes. As previously mentioned, Cid14 is a nuclear RNA decay factor responsible for oligoadenylation of diverse nuclear RNAs (Win *et al.* 2006; Buhler *et al.* 2007; Buhler *et al.* 2008; Wang *et al.* 2008; Lemieux *et al.* 2011) whilst Cid12 resides in the RDRC required for RNAi signal amplification and efficient segregation of chromosomes in mitosis and meiosis (Motamedi *et al.* 2004; Win *et al.* 2006). Though deletion of neither gene appears to show synthetic phenotypic enhancement or attenuation with deletion of the exclusively cytoplasmic decapping activator *lsm1*, the additive effects of the *lsm1Δ cid1Δ cid12Δ* and *lsm1Δ cid1Δ cid14Δ* allele combinations were sufficient to cause growth defects in these two strains more severe than those in other *lsm1Δ cid1Δ cidXΔ* genotypes. Given this observation, it is possible that the activity of the second uridylyltransferase is sensitive to the stress and/or growth conditions of the cell and that, whilst the *lsm1Δ* growth defect alone did not inhibit its activity, the more severe growth defects of the *lsm1Δ cid12Δ*/*lsm1Δ cid14Δ* genotypes did.

To test for second uridylyltransferase activity under less deleterious conditions, a second battery of Cid family double mutants were constructed in an *lsm1*$^{+}$ genetic background in the same manner as described in **3.2** above. Given the larger sequencing sample sizes required for experiments in which uridylylated mRNA was not enriched by the *lsm1* deletion, the analysis was restricted to deletions of the three remaining candidates (*cid1Δ cid12/14/16Δ*) with *cid1Δ cid13Δ* as a control. Between 49 and 105 capped *act1* mRNA sequences were retrieved for each of the four


A
act1 transcript uridylylation (%)
p < 0.05
cid1Δ
cid1Δ cid12Δ
cid1Δ cid13Δ
cid1Δ cid14Δ
cid1Δ cid16Δ
B
poly(U) tail length (nt)
C
% non-templated residues
p<10-6
internal UMP
terminal UMP
D
act1 mRNA polyadenylation (%)
p < 0.02
E
act1 poly(A) tail length (nt)

(previous page) **Figure 3.3: Activity of the second uridylyltransferase in the *cid1Δ* genetic background.** *act1* mRNA 5' and 3' end positions and any 3' non-templated additions were sequenced as described in Figure 3.2. Sequences were compared between genotypes for (A) percentage of transcripts possessing 3' oligo(U) tails and (B) length of these oligo(U) tails (red diamonds: individual uridylylation events; Bars represent mean tail length ± 1 s.d.). (C) Frequency of non-templated UMP residues contained within poly(A) tails (internal) or at the 3' teminal of transcripts (terminal) were compared for each genotype. Finally, frequency (D) and mean length ± 1 s.d. (E) of non-templated poly(A) tails were compared between genotypes. *p*-values were calculated using Pearson's χ^2 test (A, D), Poisson regression analysis (C) and Student's *t* test (E).

---

genotypes analysed (see Table 3.1). As before, the *cid1Δ cid14Δ* deletion showed a reduced rate of polyadenylation compared to other genotypes (82.9% vs. 91.4% for *cid1Δ cid13Δ*; $p < 0.02$ by Pearson's χ^2 test; Figure 3.3D), though no significant difference in poly(A) tail length was observed between samples in the *lsm1*$^+$ genetic background (Figure 3.3E).

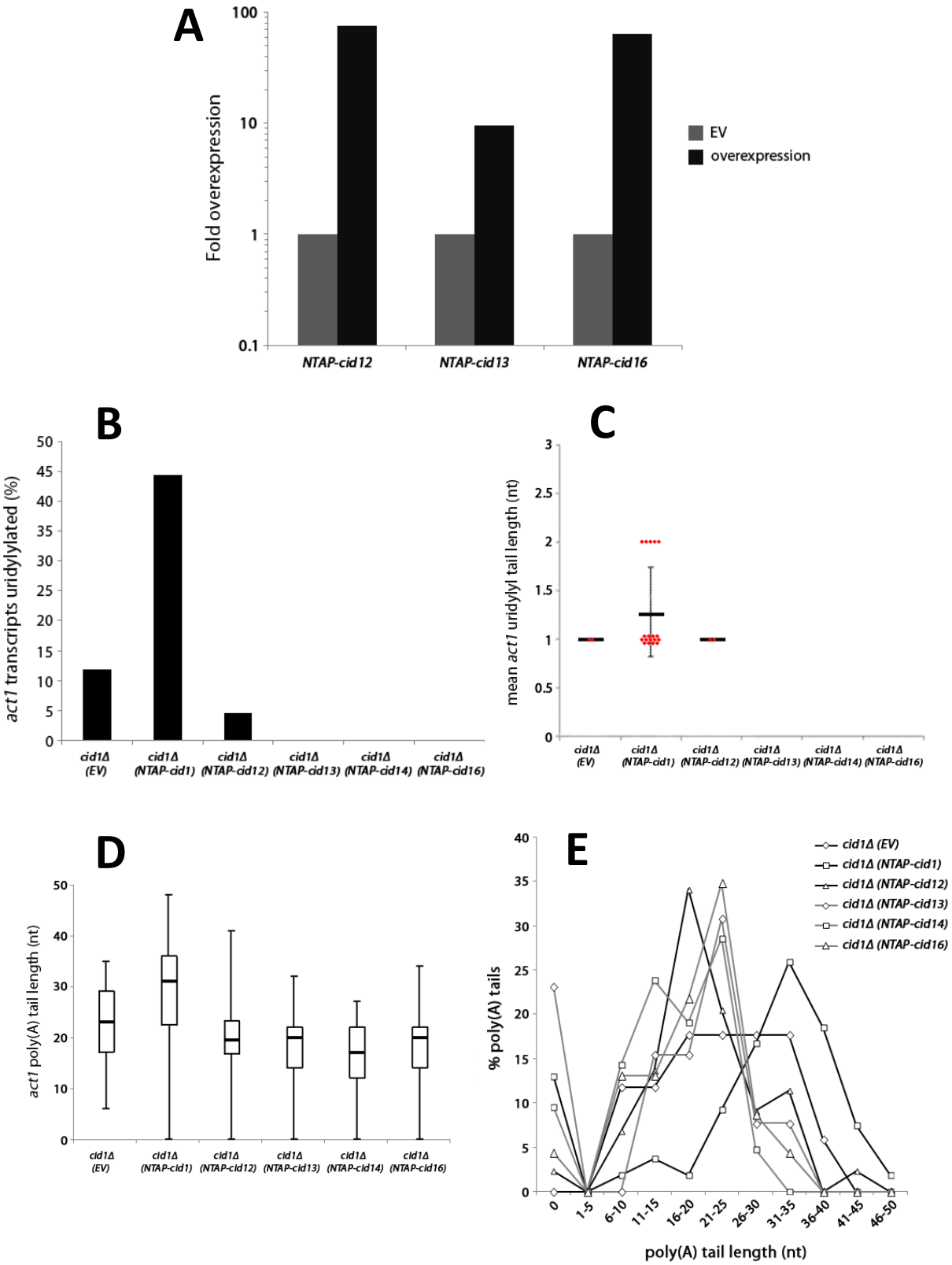
When compared to the uridylylation events observed in a *cid1Δ* single mutant in a previous study (Rissland *et al.* 2007), a statistically significant enrichment of terminal vs. internal UMP residues was seen in the *cid1Δ cid13Δ* strain ($p < 10^{-6}$; Poisson regression analysis) as previously observed in the *lsm1Δ* background (Figure 3.2A). However, no such enrichment was seen for *cid1Δ cid12Δ, cid1Δ cid14Δ* or *cid1Δ cid16Δ* strains, suggesting that the second uridylyltransferase may be inactive in these strains despite the reduced growth defects (Figure 3.3A-C). While a statistically significant lack of uridylylation is observed in the *cid1Δ cid16Δ* strain ($p < 0.05$ by Pearson's χ^2 test; Figure 3.3A), it is noted that the lower sample size of sequences analysed for this strain (n=49) may have precluded detection of the extremely low-incidence monouridylylation events observed in the *lsm1Δ cid1Δ cid16Δ* background (Figure 3.2A). Consistent with this hypothesis, such events were observed in the *cid1Δ cid12Δ* background at a similar frequency to the *lsm1Δ cid1Δ cid12Δ* background (2.9% and 2.3% respectively), suggesting these events may be independent of decapping factors and the activity of the second uridylyltransferase.

Importantly, one of the oligouridylylation events retrieved from the *cid1Δ cid14Δ* strain is the addition of a U_3 tail rather than the monouridylylation otherwise ubiquitous in second

uridylyltransferase-incompetent samples (Figure 3.3B); these longer oligo(U) tails are frequent products of the second uridylyltransferase but are absent, or extremely rare, in Cid1-catalysed uridylylation (Figure 4.3A). As discussed in **3.2** above, a U_3 tail cannot be attributed to TSS heterogeneity as tails of U_{1-2} can be, demonstrating that this is an unambiguous uridylylation event in a *cid1Δ* strain. This observation, combined with the well-described roles for Cid14 as a nuclear ncPAP (Win *et al.* 2006; Buhler *et al.* 2007; Buhler *et al.* 2008; Wang *et al.* 2008; Lemieux *et al.* 2011) suggests that Cid14 is unlikely to be the second uridylyltransferase, and thus that either Cid12 or Cid16 must represent the second uridylyltransferase.

3.5 *act1* cRACE analyses of NTAP-CidX overexpression cultures

The lack of second uridylyltransferase activity in multiple Cid-family double mutants observed above necessitated an alternative approach to assay individual Cid family members. To this end, cDNAs for all Cid family members except Cid11 (**3.6** below) were amplified by PCR and cloned into the *LEU2*-containing *pREP1-NTAP* vector, allowing expression of NTAP-tagged forms of these proteins under the control of the thiamine-repressible *nmt1* promoter (Maundrell 1993). All plasmids were sequenced to confirm the wild-type sequence had been incorporated. These plasmids, alongside the *pREP1-NTAP-cid1*$^{+}$ vector (Rissland *et al.* 2007) were transformed via the LiAc method (Bahler *et al.* 1998) into a *cid1::kanMX leu1-32* strain; transformants were selected on minimal medium lacking leucine and containing 30 μM thiamine. Transformation was confirmed by colony PCR amplification across *pREP1-cidX* ligation sites. Transformants were grown under repressive conditions to saturation then washed and transferred to thiamine-free medium for 18 hr to induce expression of the *cid* family members. Expression of NTAP-tagged mRNAs was confirmed by gene-specific qRT-PCR, and relative fold enrichment over endogenous levels was determined by comparison to a *pREP1-NTAP* empty vector control transformant (representative experiment in Figure 3.4A). Though the extent of overexpression differed between Cid family members, all showed robust overexpression of at least 9-fold in all experiments.


A
Fold overexpression
100
10
1
0.1
EV
overexpression
NTAP-cid12
NTAP-cid13
NTAP-cid16
B
act1 transcripts uridylylated (%)
cid1Δ (EV)
cid1Δ (NTAP-cid1)
cid1Δ (NTAP-cid12)
cid1Δ (NTAP-cid13)
cid1Δ (NTAP-cid14)
cid1Δ (NTAP-cid16)
C
mean act1 uridylyl tail length (nt)
D
act1 poly(A) tail length (nt)
E
% poly(A) tails
poly(A) tail length (nt)
0
1-5
6-10
11-15
16-20
21-25
26-30
31-35
36-40
41-45
46-50

(previous page) **Figure 3.4: Activity of the second uridylyltransferase upon overexpression of NTAP-tagged *cid* family members.** *cid* family cDNAs were cloned into the thiamine-inducible pREP1-NTAP vector and transformed into a *cid1Δ* strain. Overexpression was induced by growth for 18 hr in minimal medium lacking thiamine and (A) quantified by qPCR in comparison to endogenous transcript levels in an empty vector control for all experiments. A representative experiment is displayed here. *act1* mRNA 5'/3' ends and non-templated additions were sequenced as described in Figure 3.2 and analysed for (B) frequency and (C) length of oligo(U) tails (bars represent mean tail length ± 1 s.d.). Poly(A) tail length was compared between genotypes for (D) median, quartile and maxima/minima and (E) relative size distribution across 5 nt bins.

As before, total RNA was purified from transformants grown in inductive conditions and capped *act1* mRNA 5'/3' ends were sequenced by cRACE, with between 13 and 54 sequences retrieved per genotype (Table 3.1). Overexpression of *NTAP-cid1* resulted in a robust increase in *act1* mRNA uridylylation compared to the empty vector control (44.4% vs. 11.7%) and compared to the 25% previously reported for the activity of endogenous *cid1* (Figure 3.4B; Rissland *et al.* 2007; Rissland and Norbury 2009). Despite the increased frequency of uridylylation in *NTAP-cid1* overexpressing strains, no increase in the length of retrieved oligo(U) tails was observed, with the majority of uridylylation events still constituting addition of a single UMP residue (Figure 3.4C). This suggests that mono- or di-uridylylation is likely the Cid1-catalysed modification capable of directing decapping-mediated degradation of *act1* mRNA rather than a low-frequency population of longer oligo(U) tails. Consistent with this observation, mean poly(A) tail length is greater in *NTAP-cid1* overexpression cultures than empty vector controls (27.22 ± 13.1 nt vs. 22.9 ± 9.3 nt) and the profile of poly(A) tail lengths shows a marked enrichment of poly(A) tails >30nt in length in response to *NTAP-cid1* overexpression (Figure 3.4D,E). Comparison within the *NTAP-cid1* overexpression strain further showed significantly longer poly(A) tails on uridylylated transcripts than on non-uridylylated ones (31.1 ± 8.1 nt vs. 24.0 ± 15.3 nt; $p < 0.05$ by Students' t-test), supporting the hypothesis that mono-uridylylation is able to promote decapping independently of deadenylation (Rissland and Norbury 2009).

In contrast, overexpression of no other Cid family member resulted in increased uridylylation relative to the empty vector control (Figure 3.4B,C); indeed, uridylyl tails were only retrieved from

cultures overexpressing NTAP-Cid12, and these were at a rate indistinguishable from that of endogenous secondary uridylyltransferase activity in an *lsm1*$^{+}$ genetic background (4.5% vs. 4.9% in *cid1Δ cid13Δ*). The inability to enhance the second uridylyltransferase activity by overexpression was surprising; however, it is possible that different regulatory partners and/or post-transcriptional modifications may represent rate-limiting components of the second uridylyltransferase pathway and that, without concurrent overexpression of these components, second uridylyltransferase activity cannot be enriched by overexpression.

3.6 NTAP-Cid16 is a highly processive PUP *in vitro*

Given that the second uridylyltransferase appeared refractory to definitive identification by the mRNA sequencing-based methods described above, an alternative approach was used involving direct assay of *cid* family member polymerase activities. All six *cid* family members were amplified from cDNA and/or genomic DNA and cloned into the pREP1-NTAP-Cid1 thiamine-repressible expression vector and NTAP-tagged forms were overexpressed as described previously (**3.5**). *In vivo*-assembled protein complexes containing NTAP-tagged forms of each protein were purified from whole-cell lysates using IgG-Sepharose beads; successful purification was confirmed by the direct detection of the NTAP-tag with goat anti-mouse HRP-tagged secondary antibody. Successful purification of singlet bands of the expected molecular weight was achieved for all NTAP-tagged family members except for NTAP-Cid11 (Figure 3.5B). Interestingly, it was noted during the construction of pREP1 overexpression vectors that Cid11 cDNA could not be amplified from multiple cDNA samples prepared from exponentially growing cells or from a vegetative growth cDNA library (data not shown), suggesting that Cid11 is not expressed in exponentially growing cells. This observation is supported by whole-genome profiling experiments which observed expression of *cid11* mRNA exclusively during meioses I and II (Mata *et al.* 2002) and suggests that Cid11 may play specific roles in meiosis and sporulation rather than exponential growth.

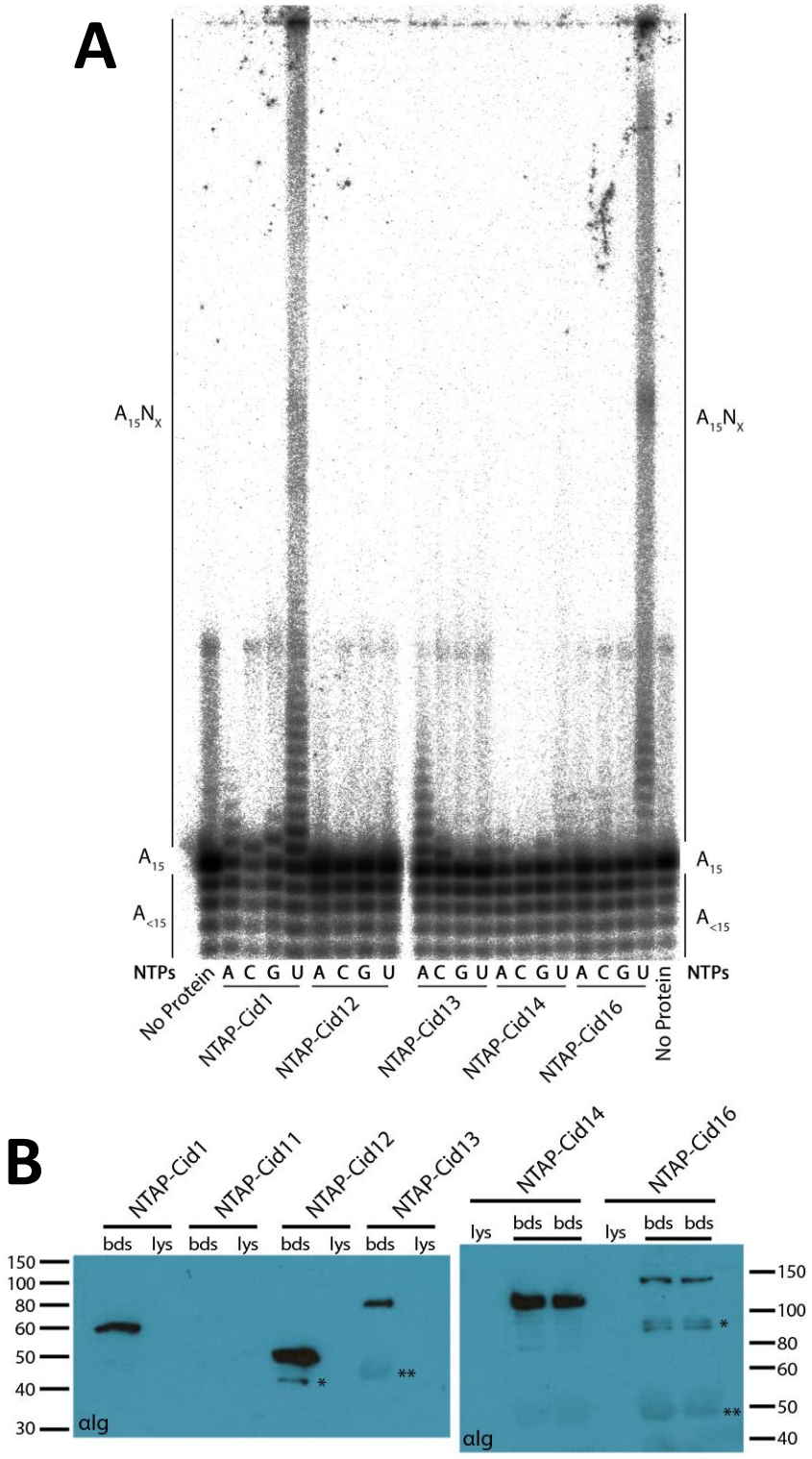


Figure 3.5: Polynucleotidyl polymerase assays of NTAP-tagged *cid* family members purified from *S. pombe*. (A) NTAP-tagged forms of each *cid* family member were purified from *S. pombe* and incubated with 5'-^{32}P-A_{15} in the presence of 0.5μM ATP, CTP, GTP or UTP (A, C, G and U respectively). As a negative control, 5'-^{32}P-A_{15} was incubated with ATP but without polymerase (No Protein). (B) Purified proteins were quantified by western blotting against the NTAP-tag in whole cell lysate (lys) and IgG-Sepharose beads (bds). * = non-specific bands; ** = antibody crossreactivity with IgG heavy chain derived from the IgG-Sepharose beads.

IgG-Sepharose beads loaded with each NTAP-tagged Cid1 family member were combined with a 5'-^{32}P-labelled A_{15} oligoribonucleotide and each of ATP, CTP, GTP or UTP individually to assay for ribonucleotide-specific polymerase activity. After 30' at 30°C, products were resolved on a 7 M urea/polyacrylamide gel at single-nucleotide resolution to visualise 3'-end polymerisation activity (Figure 3.5A). As has been previously reported, NTAP-Cid1 expressed in *S. pombe* possessed highly processive PUP activity but very limited PAP, PCP and PGP activity (Rissland *et al.* 2007). It is noted that the 16 nt product terminating with a non-templated GMP residue exhibits reduced mobility compared to other mono-nucleotidylated products (Figure 3.5A).

Whilst NTAP-Cid12 possessed no nucleotidyltransferase activity with any nucleotides tested in this assay, NTAP-Cid13 exhibited PAP activity capable of adding up to 8 nt poly(A) tails to the A_{15} oligoribonucleotide. NTAP-Cid13 was also able to add a single CMP or UMP residue to the RNA substrate, but possessed no activity with GTP as a substrate. NTAP-Cid14 was largely inactive under the overexpression and/or reaction conditions used here, but was still capable of adding one or two AMP residues and a single GMP residue to the A_{15} substrate. Intriguingly, addition of complexes containing purified NTAP-Cid14 to this assay reproducibly caused the disappearance of the artefactual band at approximately 30 nt, suggesting that NTAP-Cid14 itself, or some protein copurifying therewith, may be able to catalyse metabolism or dissociation of this species.

Strikingly, NTAP-Cid16 possessed robust and processive PUP activity similar to that of NTAP-Cid1 and was capable of adding poly(U) tails of several hundred nucleotides in length to the A_{15} substrate. Unlike NTAP-Cid1, however, NTAP-Cid16 did not exhibit any polymerase activity with ATP, GTP or CTP as a substrate, suggesting that NTAP-Cid16 possesses tightly controlled specificity for UTP as a substrate. Both NTAP-Cid1 and NTAP-Cid16 appear to exhibit periodicity in their polymerisation reactions with an enrichment of products approximately every 15 nt; this observation has been reported previously (Rissland *et al.* 2007; Rissland and Norbury 2008) and is likely to be an artefact of intra- or inter-molecular interactions between the poly(U) tail and the A_{15} oligoribonucleotide.

Consistent with this hypothesis, periodicity is not seen with poly(U) tails added by NTAP-Cid16 on non-A_{15} substrates (Figure 4.2).

It is thus concluded that, under the assay conditions utilised here, NTAP-Cid1 and NTAP-Cid16 alone are able to catalyse the addition of extensive homopolymeric poly(U) tails to an A_{15} primer. Though NTAP-Cid12 is inactive in this assay and thus could be a poly(U) polymerase, the observations here, coupled with the *act1* mRNA sequencing data presented above, strongly support the conclusion that Cid16 is the second uridylyltransferase in *S. pombe.*

3.7 H336 mutagenesis supports identification of Cid16 as the second uridylyltransferase

Recent structural investigations of recombinant *S. pombe* Cid1 by our group (Yates *et al.* 2012) and others (Lunde *et al.* 2012) have elucidated the nucleotide binding-site architecture and identified a key role for a single histidine residue, H336, in UTP selectivity. The Nε of this histidine side-chain forms a hydrogen bond with the O4 carbonyl of the UTP molecule (replaced by a primary amide group in CTP); mutation of H336 to alanine reduces PUP activity and increases PAP activity (Lunde *et al.* 2012; Yates *et al.* 2012). Alignment of the full-length sequences of Cid1 and Cid11-14 and the C-terminal 371 aa of Cid16 (containing the sequences equivalent to full-length Cid1; use of full-length Cid16 resulted in misalignment of the large 831 aa N-terminal extension with the Cid1 catalytic domain) was performed with the EMBL-EBI maintained ClustalW2 multiple sequence alignment programme (Larkin *et al.* 2007) and status of the H336 residue and other essential catalytic residues in other Cid family members determined (Figure 3.6, marked with the third green arrowhead). The ΦG[GS]-X_{9-13}-DΦ[DE]Φ (where Φ is a hydrophobic amino acid) motif characteristic

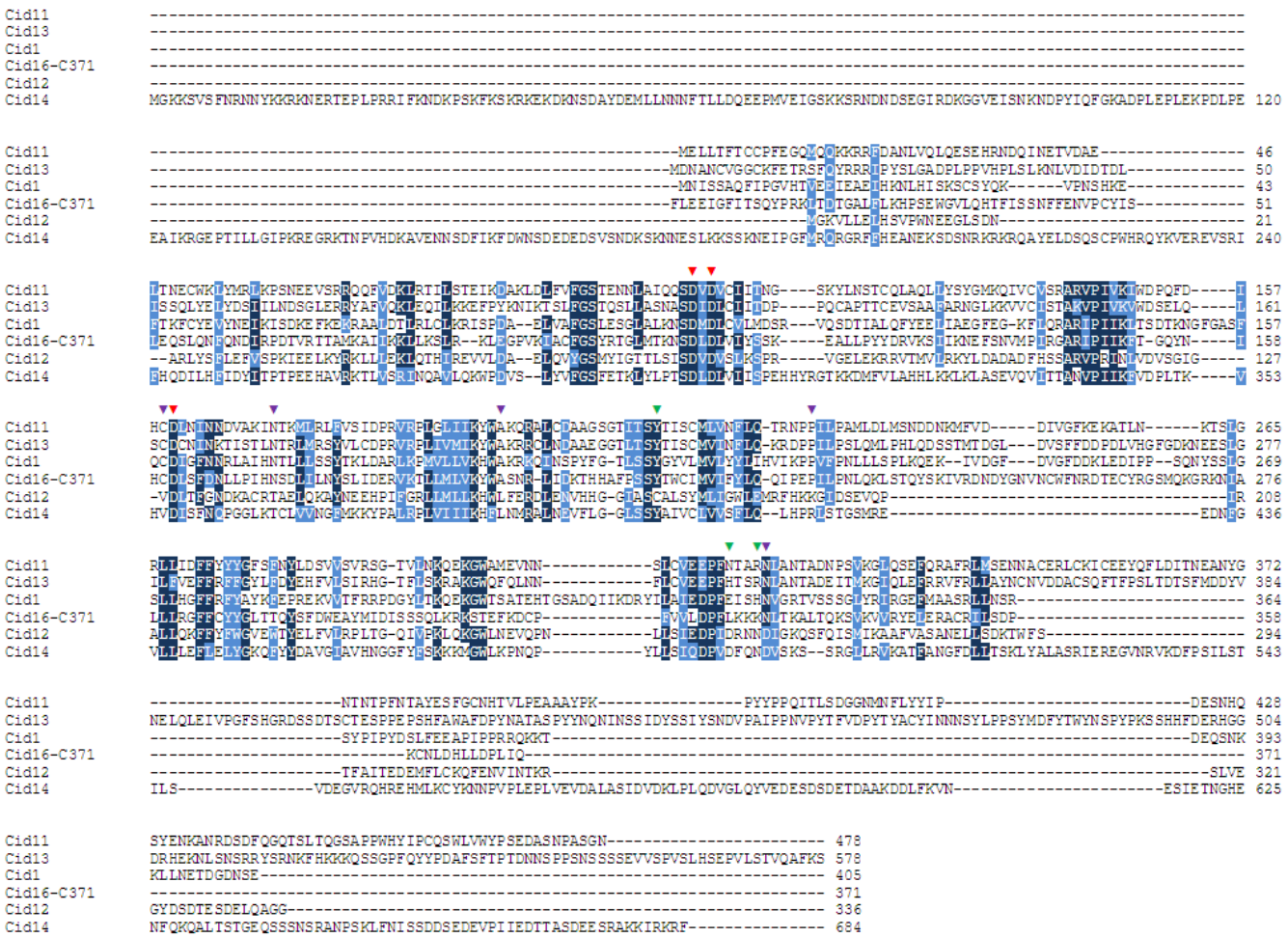


Figure 3.6: Alignment of *cid* family members in *S. pombe*. Sequences were aligned using CLUSTALW2 (Larkin *et al.* 2007) and residues exhibiting identity (dark blue) and similarity (light blue) identified. The Mg^{2+}-coordinating aspartates (red arrowheads), amino acids currently known to mediate NTP specificity (green arrowheads) and residues analysed in **3.8** below (purple arrowheads) are indicated. Due to complications arising from its long N-terminal extension, only the C-terminal 371 aa of Cid16, containing the catalytic domains, are displayed.

Figure 3.7: Phylogenetic tree of *S. pombe cid1* family members. The phylogeny was compiled with the CLUSTAL Phylogeny toolW2 (Larkin *et al.* 2007) using the alignment presented in Figure 3.6 and visualised using TreeView (Page 1996). Scale bar indicates evolutionary distance (residue changes per position).

of Pol β superfamily nucleotidyltransferases (Aravind and Koonin 1999) is conserved in all family members, as are the three divalent cation-coordinating aspartate residues (Figure 3.6, red arrowheads), suggesting that, at least by this criterion, all family members are capable of catalytic activity. Surprisingly, however, despite its central role in UTP selectivity, the H336 residue is not conserved in any other *S. pombe* Cid-family members, instead being substituted by a lysine (K) in Cid16, an arginine (R) in Cid11/Cid13 and an asparagine (N) in Cid12/Cid14.

To investigate the effects of these substitutions, their effects on the catalytic activity of recombinant Cid1 were investigated. H336K, H336R and H336N mutations were introduced by site-directed mutagenesis into a truncated form of Cid1 lacking the N-terminal 31 aa (tCid1) expressed from the pGEX6P-1 *E. coli* expression vector. All mutated vectors were sequenced to confirm the lack of any second-site mutations. Overexpression of GST-tagged forms of each mutant protein and the unmutated tCid1 control was induced in BL21 (DE3) (pLysS) *E. coli* strains with IPTG, purified by GST-Sepharose bead pull-down and human rhinovirus 3C protease cleavage from the GST tag. All three mutant proteins were strongly expressed in *E. coli* and purified with similar yields (0.5-1 ng/μL) to the tCid1 control. All protein samples yielded a singlet protein band of the expected molecular weight in SDS-PAGE analyses and were >95% pure (data not shown). The ability of equimolar quantities of tCid1 and each H336 mutant to transfer individual nucleotides to a 5'-^{32}P-labelled A_{15} primer was assessed as described previously (**3.6**). The polymerase assay gel is displayed in Figure 3.8.

As has previously been reported (Rissland *et al.* 2007), tCid1 possesses robust PUP activity similar to that of the NTAP-Cid1 purified from *S. pombe* (Figure 3.5A) and has similar poorly processive PCP and PGP activity; however, the PAP activity of recombinant tCid1 is significantly greater than that of NTAP-Cid1 from *S. pombe* and is capable of adding up to ~15 nt poly(A) tails to the A_{15} substrate (see also Figure 3.9A). The cause of this enhanced PAP activity is unknown but could relate either to partial UTP selectivity roles for the N-terminal 31 aa of full-length Cid1 which are absent in the

recombinant tCid1, or to the lack of Cid1 cofactors and/or post-translational modifications in the recombinant sample that copurify with NTAP-Cid1 expressed in *S. pombe*. Importantly, each of the H336 mutants tested exhibited different effects on the catalytic activity of tCid1. H336K appeared to slightly attenuate PAP and PGP activity of tCid1 (as evidenced by the increased quantity of unpolymerised 5'-^{32}P-A_{15} template) without affecting PCP activity. Whilst it also resulted in reduced PUP activity in comparison to wild-type tCid1, tCid1-H336K still exhibited robust PUP activity and was capable of adding poly(U) tails of up to ~100 nt in length to the A_{15} template. In contrast, the H336R mutation appeared to greatly impair the PUP activity of tCid1, resulting in a predominant product with only a single UMP residue added, without significantly altering the PAP activity. Most strikingly, the H336N mutation greatly reduced the PUP activity of tCid1 (with a majority of mono- or di-uridylylated products) but significantly enhanced its PAP activity, allowing processive addition of poly(A) tails of up to 25 nt in length. During preparation of this thesis, identical activity for an H336N mutant was reported by another group (Lunde *et al.* 2012), substantiating these observations.

The respective activities of each of the tCid1-H336 mutants align well with observations of the activity of the *S. pombe* Cid-family members whence they derive. NTAP-Cid16 purified from *S. pombe* was found to be a robust and highly specific PUP (Figure 3.5A), whilst the corresponding H336K was unique among the H336 mutations in possessing robust and processive PUP activity. Interestingly, though a large-scale cladistic analysis of ncPAPs from diverse species placed Cid16 in a remote branch from Cid1 (Figure 1.1), a restricted cladistic analysis of *S. pombe* Cid family members (Figure 3.7) suggests that Cid16 catalytic domains are most similar to those of Cid1. Consistent with this, an *in silico* evolution-based phylogenetic algorithm designed to identify evolutionary relationships within protein families places Cid1 and Cid16 in close apposition when the same population of ncPAPs to Figure 1.1 is analysed, suggesting that the distant relationship there observed may be an artefact of the long N-terminal Cid16 extension (S. Kelly, pers. comm.). Whilst

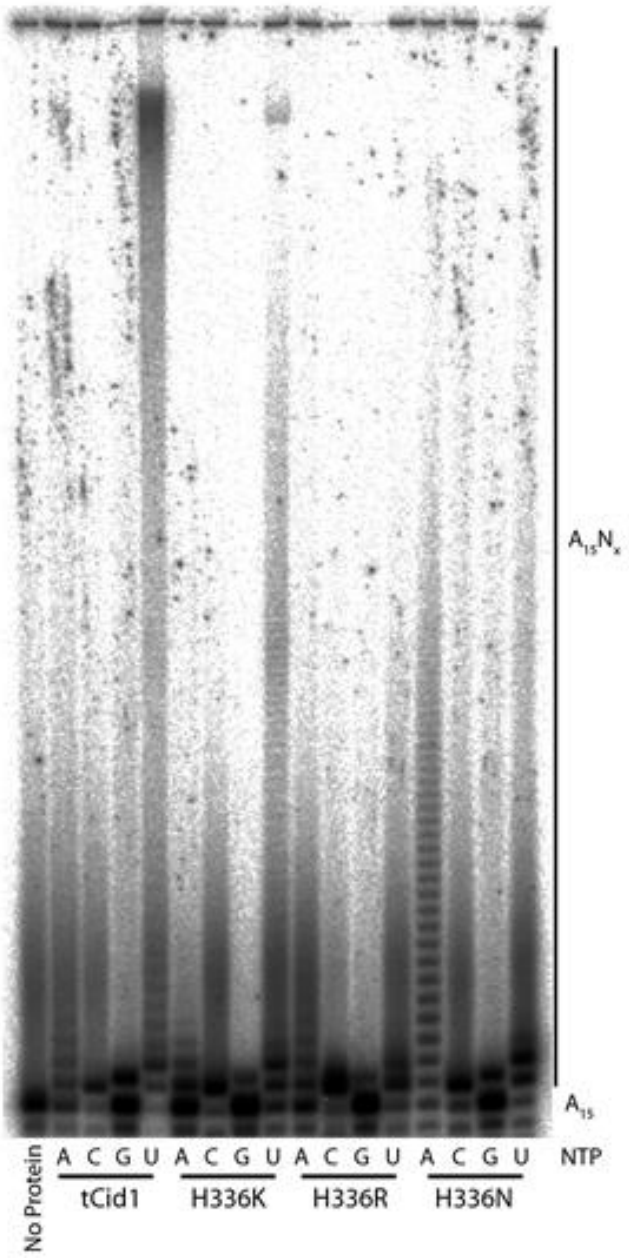


Figure 3.8: *cid* family-directed H336 mutations have differing effects on PAP and PUP activity. tCid1, tCid1-H336K (resembling Cid16), tCid1-H336R (resembling Cid11/Cid13) and tCid1-H336N (resembling Cid12/Cid14) were incubated with 5'-^{32}P-A_{15} and either ATP, CTP, GTP or UTP (A, C, G and U respectively). As a negative control, 5'-^{32}P-A_{15} was incubated with ATP but without polymerase (No Protein).

little to no catalytic activity was found for NTAP-Cid12 and NTAP-Cid14 under the reaction conditions utilised here (Figure 3.5A), Cid14 has been well-characterised *in vivo* as a nuclear poly(A) polymerase responsible for promoting the degradation of aberrant nuclear RNAs (Win *et al.* 2006; Buhler *et al.* 2007; Buhler *et al.* 2008; Wang *et al.* 2008; Lemieux *et al.* 2011). Consistent with such an activity, the Cid12/Cid14-directed H336N mutant exhibits both enhanced PAP activity and markedly attenuated PUP activity, possibly due to the inability of asparagine to form a stabilising H-bond with the uracil base of UTP. Finally, the Cid11/Cid13-directed mutant, H336R, also exhibits greatly decreased PUP activity but does not result in a compensatory enhancement in PAP activity as seen with H336N. This is surprising given the significant PAP activity of the NTAP-Cid13 preparation (Figure 3.5A); however, it is possible that other mutations in the nucleotide binding site of tCid1 are required to fully recapitulate the Cid11/Cid13 binding mechanism (explored in **3.9** below).

3.8 Cid12 and Cid14 may possess a structural architecture different from that of Cid1

The phylogenetic analysis displayed in Figure 1.1 places Cid11 and Cid13 in the same broad clade as Cid1 and Cid16 as well as a number of other known PUPs from *Aspergillus*, *Arabidopsis*, *C. elegans*, humans and mice. This suggests that these enzymes may represent evolutionary revertants which have regained PAP activity in an otherwise PUP-dominated clade, and thus that Cid1, Cid11, Cid13 and Cid16 are likely to share a common evolutionary ancestor. By contrast, Cid12 and Cid14 cluster together in a different clade typified by the nuclear ncPAPs Trf4p/Trf5p from *S. cerevisiae*. Upon examination of the alignment of *S. pombe* Cid family members described previously (Figure 3.6), five residues were identified that were conserved within the two clades but differed between them (Figure 3.6, purple arrowheads) and which thus may represent important determinants of Cid1 clade-specific activity and/or UTP selectivity.

To test this hypothesis, each of these Cid1 codons was changed to that for the corresponding amino acid found in the Cid12/Cid14 clade (C159V, N171T, A196L, P229R and N337D mutations) by site-directed mutagenesis of the pGEX6P-1-tCid1 vector. As before, vectors were sequenced to exclude

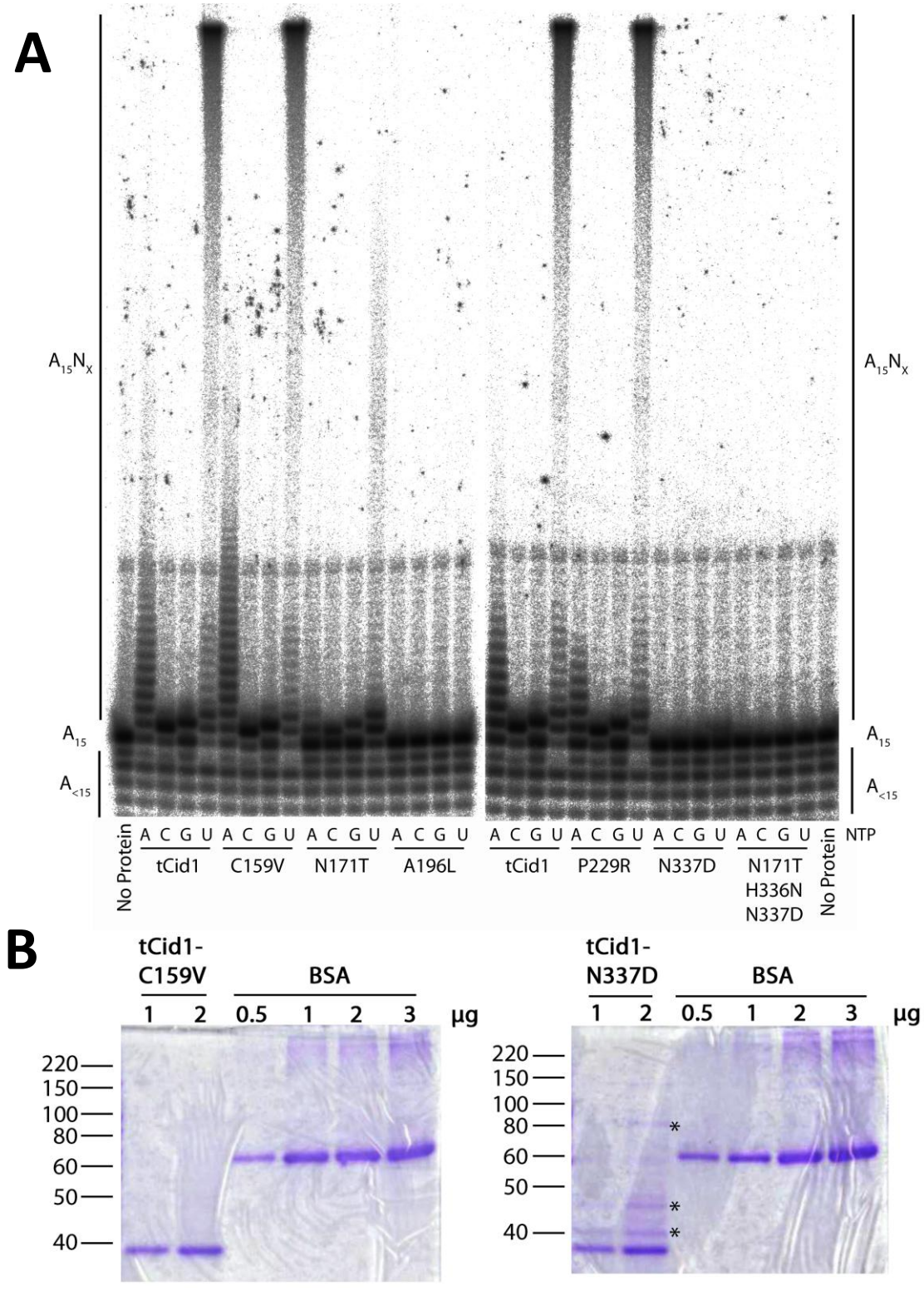


Figure 3.9: *cid12/cid14*-directed mutants demonstrate varying effects on tCid1 activity. (A) tCid1, tCid1-C159V, tCid1-N171T, tCid1-A196L, tCid1-P229R, tCid1-N337D and tCid1-N171T/H336N/N337D were incubated with 5'-^{32}P-A_{15} and either ATP, CTP, GTP or UTP (A, C, G and U, respectively). As a negtive control, 5'-^{32}P-A_{15} was incubated with ATP but with without polymersae (No Protein). (B) Representative purifications of active (tCid1-C159V) and inactive (tCid1-N337D) purifications were separated on Coomassie-stained SDS-PAGE gels alongside BSA for comparative quantitation. * = prominent non-specific bands.

second-site mutations. All mutant proteins were expressed in *E. coli*, purified by glutathione-Sepharose affinity chromatography and assayed for NTP-specific activity on an A_{15} oligoribonucleotide as described above (**3.7**). Figure 3.9A demonstrates that two mutants, C159V and P229R, showed no effect on the activity of tCid1, suggesting they are not required for catalytic activity under the reaction conditions used here. The N171T mutation resulted in severely attenuated, but not abolished, polymerase activity using all four nucleotides, with the long poly(U) tails of tCid1 mostly reduced to mono- or di-uridylyl additions, and only a small quantity of mono-adenylation, -cytidylation and -guanidylation observed. It is thus concluded that, whilst the N171T mutation was still catalytically active, some aspect of catalytic regulation or substrate/template recognition was severely impaired and thus that N171 is a residue with important functions in the context of the Cid1/11/13/16 architecture but which diverges in the Cid12/14 clade. The remaining two mutations, A196L and N337D, completely abolished catalytic activity with all nucleotides, suggesting that integrity of A196 and N337 is essential for tCid1. The A196L and N337D mutants also greatly reduced the expression of GST-tCid1 in *E. coli* cells (data not shown) and were purified at approximately 20-fold lower concentrations than wild-type and non-disrupting mutations (0.02-0.06ng/µL vs. 0.5-2ng/µL, respectively). As a consequence, the preparations of these proteins were much less pure than those of other mutants, with approximately 25% of proteinaceous material composed of contaminating species (representative tCid1-N337D and tCid1-C159V preparations presented in Figure 3.9B; asterisks indicate major contaminants). The poor expression and purification of tCid1-A196L and -N337D, coupled with complete loss of catalytic activity, suggests that these proteins may be partially or completely unfolded. In support of a gross structural defect in these proteins, the lack of activity in the N337D mutation could not be rescued by complementary second- and third-site mutations that recapitulated the Cid12/Cid14-clade NTP binding pocket, since an analogously expressed and purified tCid1-N337D/H336N/N171T triple mutant was similarly difficult to purify and also entirely inactive in the polymerase assay (Figure 3.9A).

To investigate the possible structural bases for the divergent effects of the mutations tested above, the five mutated residues were plotted on a ribbon diagram of tCid1 complexed with UTP recently solved by our group (Yates *et al.* 2012). The position of each residue is shown in several orientations in Figure 3.10. P229 lies in a surface loop between helices αG and αH and is remote from both the catalytic and nucleotide-binding regions, supporting the observation that this residue does not play a role in catalysis. C159, by contrast, lies in the β4 strand within the anti-parallel β-sheet of the N-terminal domain (NTD) and is immediately adjacent to D160, a member of the catalytic triad. Despite this much more central location, the side-chain of C159 does not protrude into the catalytic cleft but instead resides on the opposite face of the β-sheet, facing towards helices αB and αD. As such, though C159 resides in a more central location, its side-chain is remote from the NTP-binding and catalytic regions and therefore unlikely to play a role in catalysis or interaction.

The functionally essential A196 residue lies in the αF helix that also contains positively-charged residues essential for binding the negatively-charged phosphates of UTP, and its side-chain is deeply embedded in the hydrophobic core of the C-terminal domain (CTD). The substitution of the much larger leucine side-chain is expected to disrupt the tight packing of this hydrophobic core and could cause large-scale disruptions in the folding of the protein. The second essential residue, N337, is immediately adjacent to the UTP-binding H336 residue discussed above (**3.7**), but unlike H336 faces away from the nucleotide-binding cleft, lying instead on the surface of the protein between several loops. This location was initially surprising given the importance of N337 to the activity and, potentially, structural integrity of tCid1; however, careful examination of the N337 microenvironment showed the close apposition of the backbone carbonyl of R323 to the amine of the N337 side-chain and of the backbone amide of I328 to the side-chain carbonyl, suggesting that the N337 side-chain may form conserved H-bonds with the protein backbone at both of these residues (Figure 3.11). Consistent with this, the measured lengths of the putative H-bonds (2.90Å and 2.85Å respectively) both fall within the optimum reported length for O-H-N H-bonds (2.85-2.95Å; Baker and Hubbard 1984). Importantly, both R323 and I328 lie in a loop between the

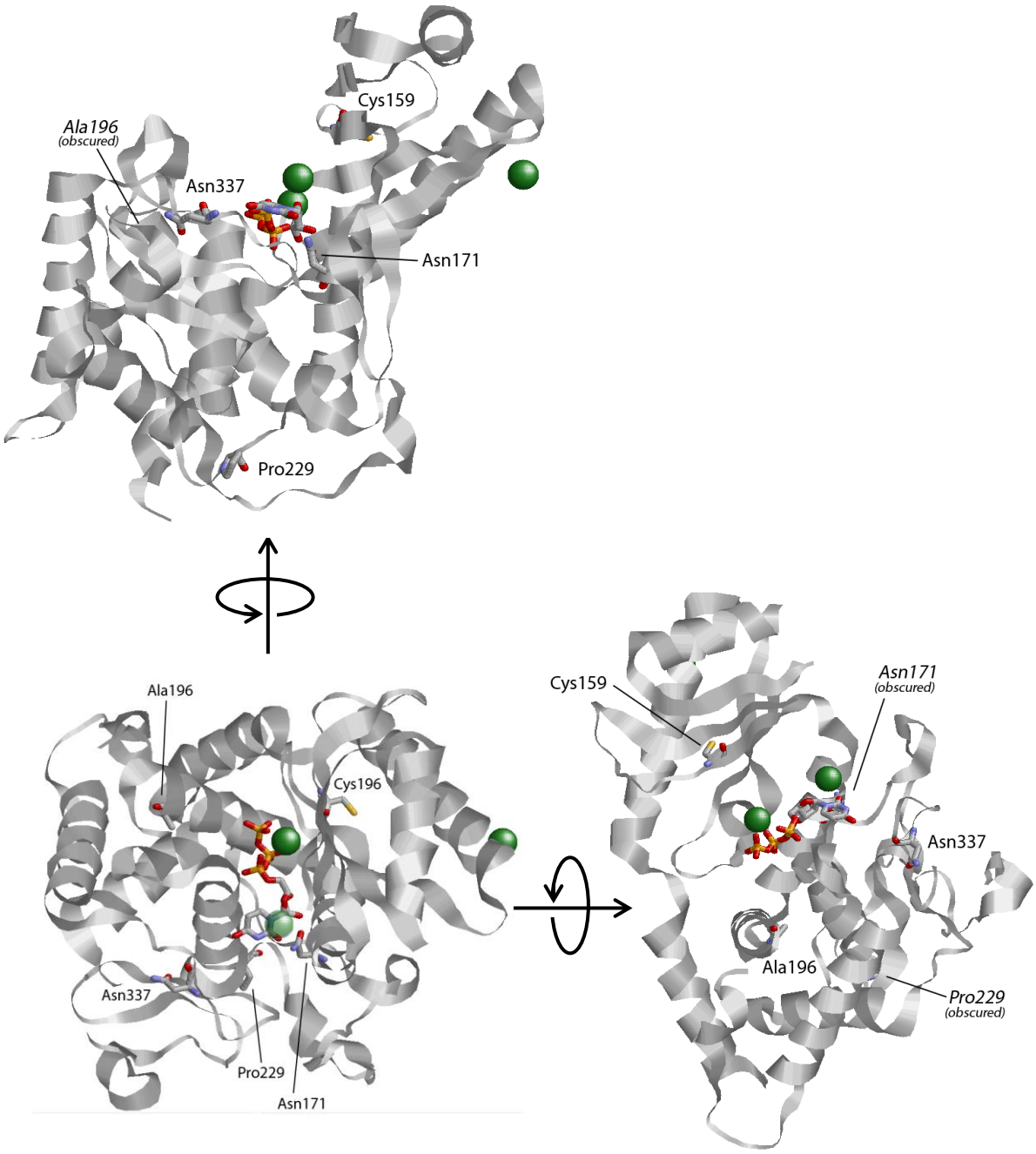


Figure 3.10: Location of *cid12/cid14*-directed mutants in the tCid1 structure. Three different orientations of the tCid1 structure (Yates *et al.* 2012) are shown with the relevant amino acids labelled and displayed in stick format. A bound UTP molecule is shown in stick format and coordinated Mg^{2+} ions as space-filled green balls.

proposed "β-trapdoor" region and the essential nucleotide recognition motif (NRM) loop that contains the H336 residue (Yates *et al.* 2012); even the relatively conservative substitution of N337D would disrupt these H-bonds and could therefore disrupt the structural architecture of a large region of the tCid1 protein, including the NRM loop and the β-trapdoor, and consequently severely disrupt all catalytic activity.

Finally, the N171 residue lies within the αE helix of the NTD and, uniquely among the residues investigated in this section, lies in close juxtaposition with the bound UTP molecule. Furthermore, N171 forms several H-bonds with UTP, its side-chain carbonyl group binding the 2'-OH of the ribose moiety whilst the side-chain amine group binds the O2 carbonyl of the uracil ring itself (shown in Figure 3.12; Lunde *et al.* 2012; Yates *et al.* 2012). It is also possible that the side-chain amine forms an H-bond with the phenyl-hydroxyl group of the uracil-stacking Y212 side-chain, though at 3.49Å this bond would be at the limit of the reported length for protein O-H-N sidechains (Baker and Hubbard 1984). Importantly, N171 is also reported to contact the 2'-OH of tCid1-bound 3'-deoxy-ATP and, via a water molecule, the N7 tertiary amide of guanosine in a coordinated GTP molecule (Lunde *et al.* 2012), suggesting that it is likely to interact with all coordinated NTPs. Loss of the H-bonds between N171 and the base and/or ribose of a coordinated NTP are likely to destabilise but not abolish the tCid1-NTP complex and thus may explain the extensive, but not complete, loss of catalytic activity seen with the N171T mutation.

On the basis of these structural observations, it is concluded that Cid12 and Cid14 have conserved a number of residues that, when compared to the Cid1-containing clade, must result in significantly different structural architecture and/or catalytic mechanisms. Such an observation would support a model in which a Cid12/Cid14 ancestor diverged evolutionarily from the Cid1 clade much earlier than the other Cid family members, and suggests therefore that more recent evolutionary duplication or divergence events are likely to have generated the two uridylyltransferases, Cid1 and Cid16, as well as the revertant PAP activity of Cid11 and Cid13.

3.9 Cid11 and Cid13 likely use multiple amino acid residues to select ATP over UTP

One surprising observation made in the course of H336 mutagenesis in section **3.7** above was the poor PAP activity of the Cid11/Cid13-directed H336R mutation. Given that Cid13 was shown to possess PAP activity (**3.6**), it was reasoned that additional, second-site mutations must exist in Cid11/Cid13 compared to Cid1 that complement the H336R mutation and permit more processive ATP-selective polymerase activity. One strong candidate as a second-site mutation was the R340 residue of Cid1, which is conservatively substituted by lysine (K) in Cid16, Cid12 and Cid14 but non-conservatively substituted by an asparagine (N) in both Cid11 and Cid13 (Figure 3.6). Furthermore, solution structures of tCid1-UTP complexes (Munoz-Tello *et al.* 2012; Yates *et al.* 2012) show that the R340 side-chain protrudes into the nucleotide-binding pocket of tCid1 in a way that could sterically hinder interaction of the bulkier adenine base (Figure 3.13). Consistent with such a model, a structure of Cid1 with 3'-deoxy-ATP from another group (Lunde *et al.* 2012) shows the adenine base arranged in an *anti* orientation with respect to the ribose, resulting in extrusion of the N6 amine further into the nucleotide binding pocket than UTP where steric conflicts with the R340 side-chain could occur. The corresponding residue in the *Trypanosoma brucei* ncPAP TUT4 stabilises protein-base interactions via coordination of a water molecule (Stagno *et al.* 2007).

To test the above hypothesis, the R340N mutation was generated in pGEX6P-1-tCid1 both singly and in combination with the H336R mutation and sequenced to confirm accuracy of mutagenesis. The single and double mutants were purified and assayed for NTP-specific polymerase activity as performed above; the gel-resolved products are presented in Figure 3.14. The H336R mutation attenuated PUP activity as previously reported but surprisingly, in this assay also caused a reduction in PAP activity compared to tCid1. A similar reduction of both PUP and PAP activity was observed with the R340N mutation, while the H336R/R340N double mutant was indistinguishable in terms of PAP/PUP activity from the two single mutants.

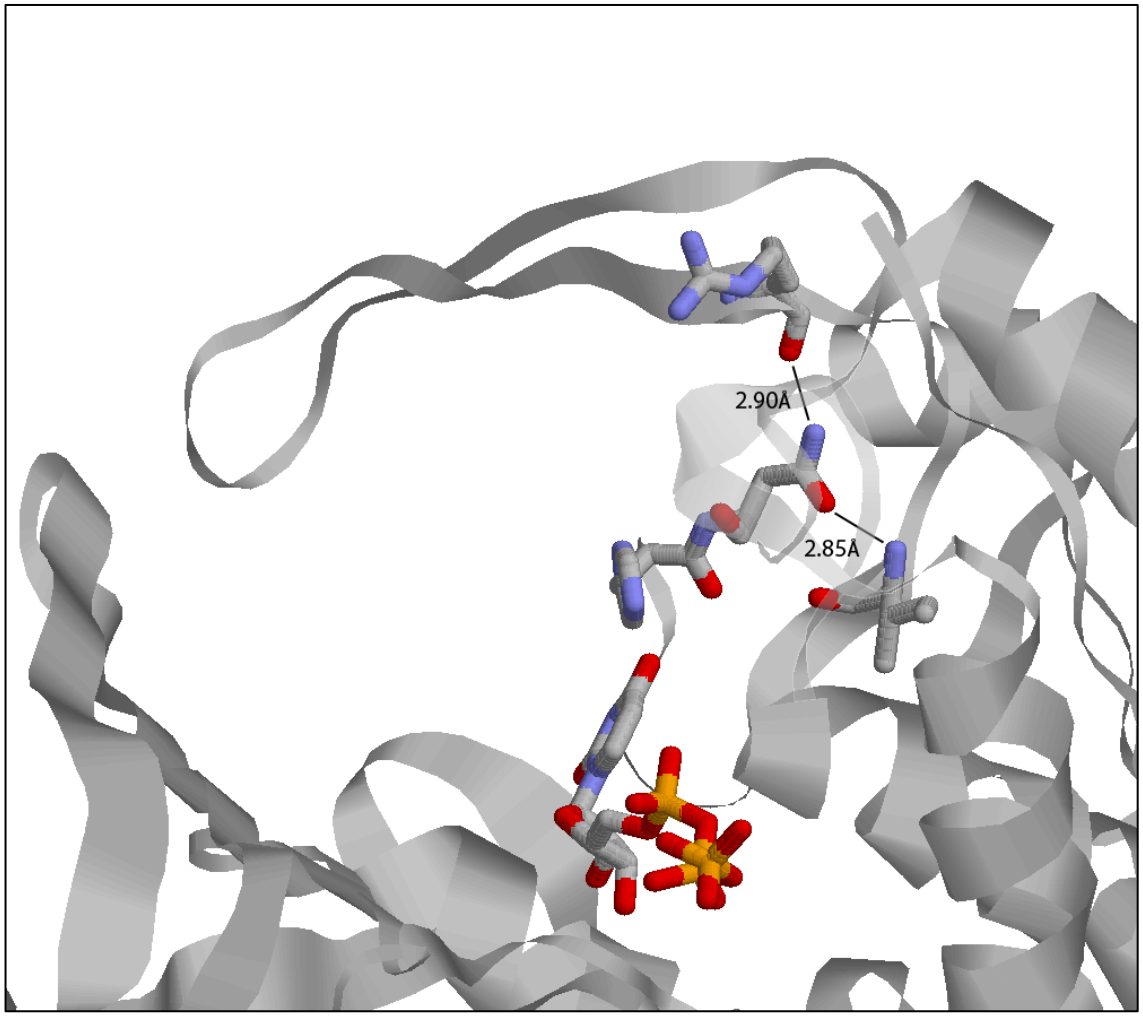


Figure 3.11: Proposed H-bonding pattern of the N337 side chain. Close-up of the UTP binding loop of the tCid1 structure (Yates *et al.* 2012) showing UTP, H336, N337, R323 and I328 are shown in stick format with the putative H-bonds and H-bond distances marked.

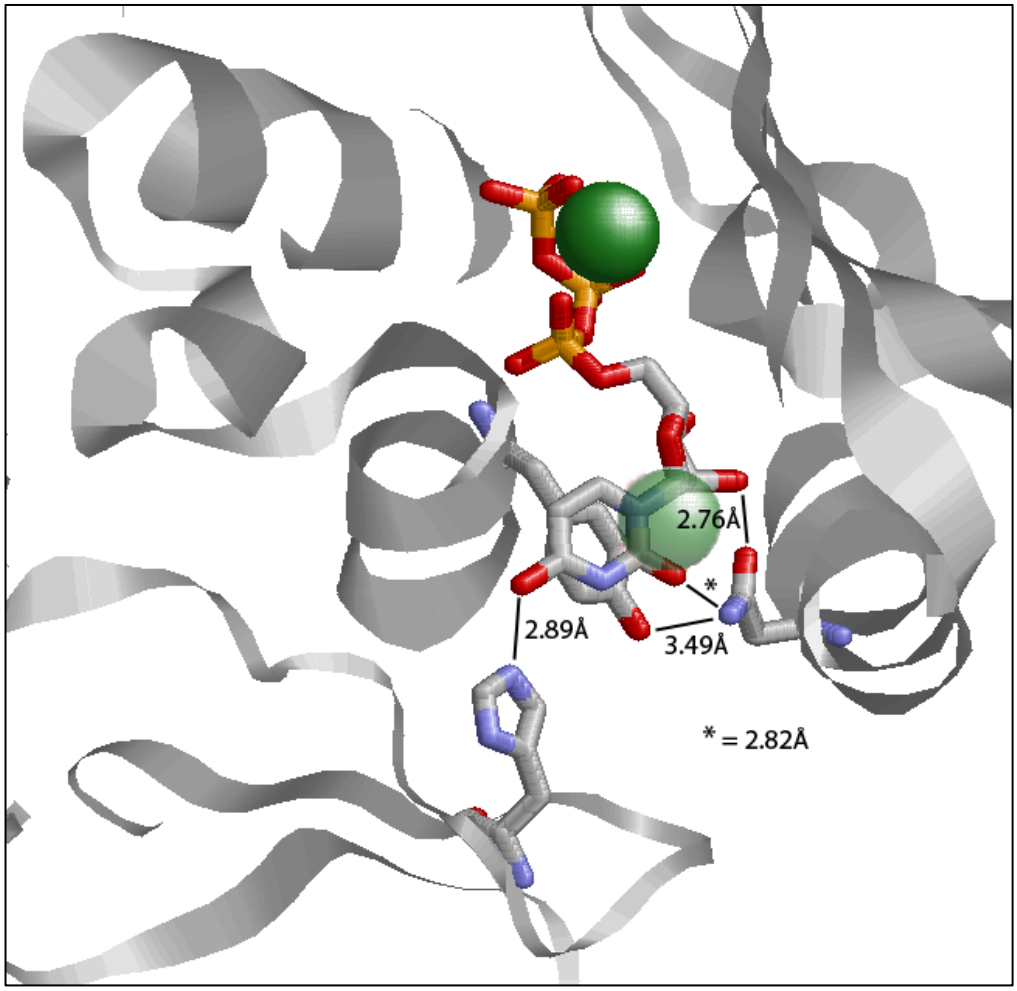


Figure 3.12: Proposed H-bonding network involving N171. The NTP-binding pocket in the tCid1 structure (Yates *et al.* 2012) is displayed with UTP-coordinating residues (N171, Y212, H336) and the coordinated UTP displayed in stick format. Coordinated Mg^{2+} ions are displayed as space-filled green balls. Proposed H-bonds involving N171 and their lengths in are displayed.

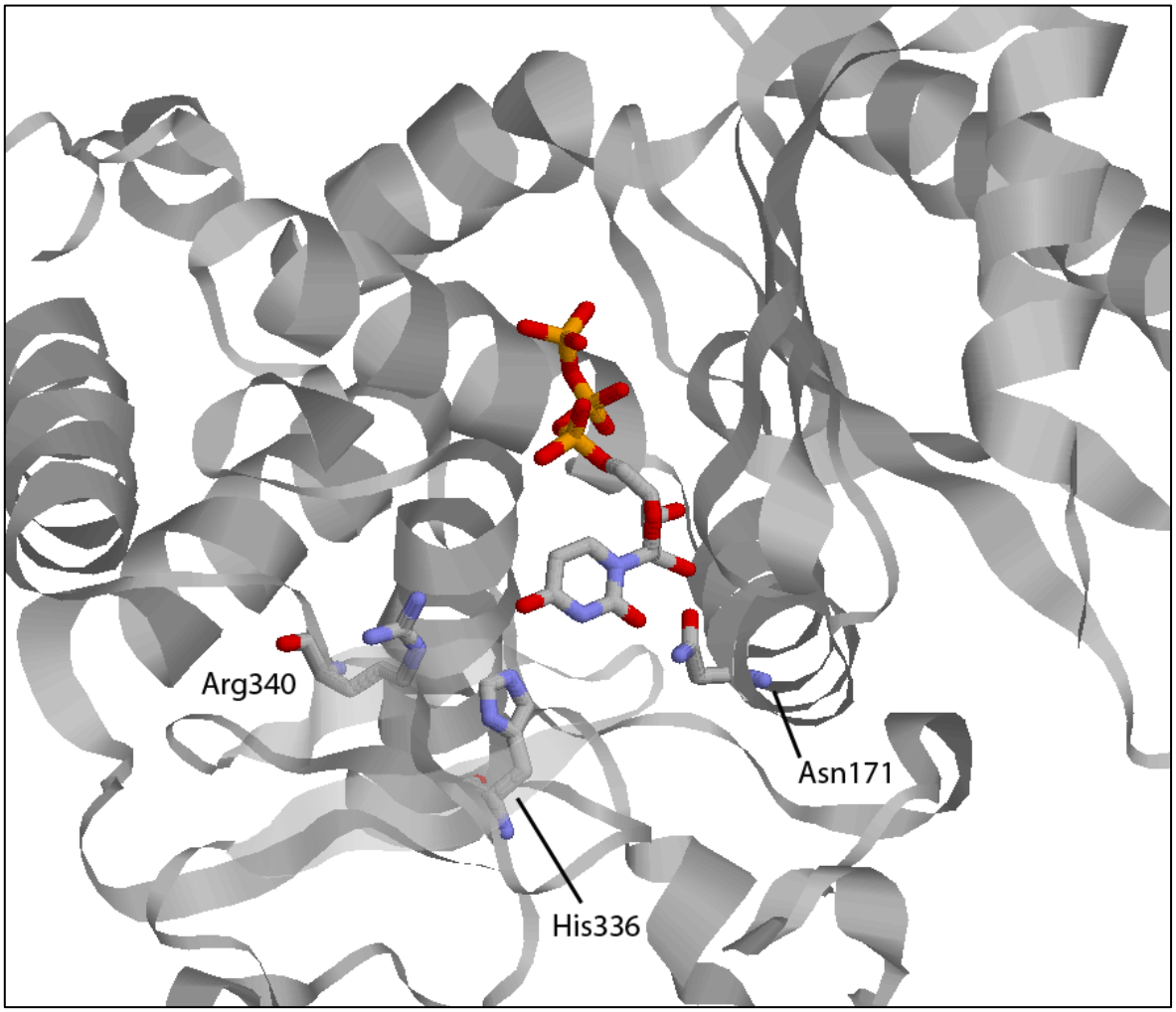


Figure 3.13: Positioning of the R340 side chain within the NTP binding pocket. The NTP-binding pocket of tCid1 (Yates *et al.* 2012) is shown with N171, H336, R340 and the coordinated UTP displayed in stick format.

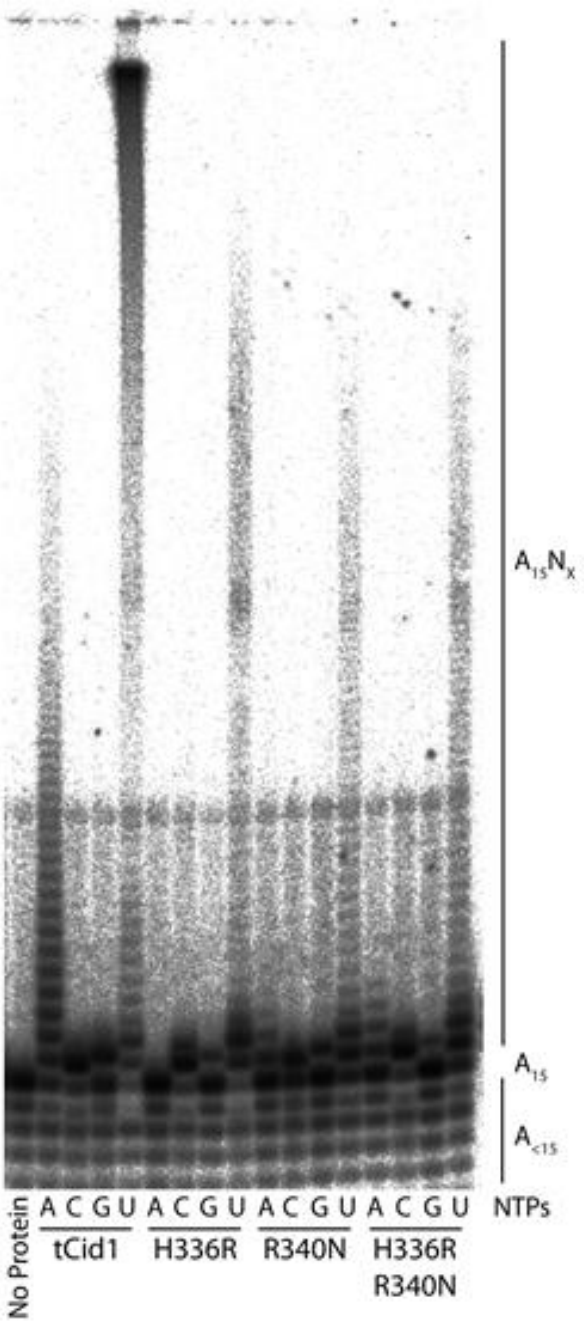


Figure 3.14: *cid11/cid13*-directed tCid1 mutations do not recapitulate PAP activity. tCid1, tCid1-H336R, tCid1-R340N and tCid1-H336R/R340N were incubated with 5'-^{32}P-A_{15} and either ATP, CTP, GTP or UTP (A, C, G and U, respectively). 5'-^{32}P-A^{15} was incubated with ATP but no polymerase as a negative control (No Protein).

The retardation of PAP activity upon the R340N mutation suggests that, instead of sterically hindering the adenine base in the active site, the R340 side-chain may play a role in tCid1-ATP and/or tCid1-UTP interactions; alternatively, R340 may have a structural role in the tCid1 nucleotide binding pocket, for example by interacting with residues of the αG helix containing the critical Y212 which interacts with the uracil ring through π-orbital stacking (Lunde *et al.* 2012; Munoz-Tello *et al.* 2012; Yates *et al.* 2012). The reduced PAP activity of H336R in this assay suggests that an arginine side-chain at position 340 may be sufficient to sterically exclude both UTP and ATP from the nucleotide-binding site. Finally, the lack of synthetic PAP activity enhancement in the H336R/R340N mutant suggests that these two residues alone are not sufficient to interconvert PAP and PUP activities in the tCid1 backbone, and that other residues or structural features must exist within Cid11/Cid13 nucleotide-binding site that participate in conferring ATP selectivity.

3.10 Discussion

3.10.1 Summary

Extensive sequencing of *act1* mRNA 5' and 3' ends in the *cid1Δ* genetic background has unequivocally proven the existence of a second *S. pombe* mRNA uridylyltransferase; however, due to repression of this activity upon deletion of several different *cid* family members, it is impossible by genetic means alone to distinguish between *cid12* and *cid16* as encoding this second uridylyltransferase. Complementary *in vitro* PUP activity assays with NTAP-tagged *cid* family members overexpressed in *S. pombe* showed that NTAP-Cid16 possesses robust and highly specific PUP activity, while NTAP-Cid12 has no activity in this assay. Mutagenesis of the H336 residue critical for UTP selectivity (Lunde *et al.* 2012; Yates *et al.* 2012) in a recombinant tCid1 background supported the identification of Cid16 as the second uridylyltransferase, while mutagenic testing of residues differing between the Cid1/11/13/16 and Cid12/14 clades suggested that Cid12 and Cid14 likely have a significantly different structural

architecture and catalytic mechanism to Cid1. While Cid11 and Cid13 appear to have diverged from Cid1 relatively recently in evolutionary history, mutagenic analysis with tCid1 suggests that the evolutionary remodelling of the NTP-binding pocket likely involved several different residues and has not been fully characterised here.

3.10.2 The activity of ncPAP family members in S. pombe

The *in vitro* PUP assays conducted here with NTAP-purified versions of all the *S. pombe* ncPAPs represents the first systematic attempt to measure the activity of all family members under parallel conditions. This assay accurately recapitulated the previous ncPAP assays for Cid1 and Cid13 (Read *et al.* 2002; Rissland *et al.* 2007) which exhibited predominantly PUP and PAP activity respectively; it also provided the first evidence that NTAP-Cid16 is a robust and highly-selective PUP when purified from *S. pombe* under *nmt1* promoter-driven overexpression.

By contrast, an extremely limited amount of activity was recorded with NTAP-Cid14, which was capable of adding only one or two AMP or one GMP residue to an A_{15} substrate at low efficiency (Figure 3.5A). While a previous report found Cid14 to possess robust PAP activity (Buhler *et al.* 2007), this analysis was conducted using recombinant Cid14 purified from *E. coli* and thus is qualitatively different to the assay conducted here. Furthermore, no activity was recorded at all with NTAP-Cid12, despite it being the most abundantly purified of the NTAP-ncPAP proteins (Figure 3.5B). There are several possible reasons for this lack of activity. Firstly, both Cid12 and Cid14 function *in vivo* as components of heterotrimeric complexes – RDRC (Motamedi *et al.* 2004) and the *Sp*TRAMP complex (Buhler *et al.* 2007) respectively – and are essential to the function of these complexes. Robust overexpression of either NTAP-Cid14 or NTAP-Cid12 without concurrent overexpression of the other components of their respective complexes is likely to result in a loss of stoichiometry between component subunits and, consequently, a large proportion of NTAP-Cid14 and NTAP-Cid12 remaining in the cell as unassembled monomers, which may

well be unable to interact with RNA targets or may be catalytically inactive. Alternatively, though not necessarily exclusively, the presence of the NTAP tag on these proteins could interfere with their assembly into their functional heterotrimers and/or could interfere with their import into the nucleus, sequestering them in the cytoplasm. Finally, while Cid14 is known to be capable of acting on ssRNA substrates *in vivo* (Win *et al.* 2006; Buhler *et al.* 2007; Wang *et al.* 2008), the endogenous targets of Cid12 are currently unknown and, given the role of the RDRC in cotranscriptionally synthesising dsRNA (Motamedi *et al.* 2004; Colmenares *et al.* 2007), they could be structurally markedly different to the A_{15} oligoribonucleotide used in these *in vitro* assays, which may consequentially not represent a suitable substrate of Cid12. Consistent with this idea, the highly-conserved PI[I/V]K motif found in all other *S. pombe* ncPAPs (Figure 3.6) and hypothesised from its position and orientation to be involved in binding the 3' end of the substrate RNA (personal observations) is not conserved in Cid12, which instead contains a PRIN motif at this location.

A different problem was encountered during the testing of NTAP-Cid11, which was repeatedly found to be refractory to expression under the vegetative growth conditions used here. *In vivo*, Cid11 expression is tightly restricted to the meiotic life cycle, peaking at meiosis I and II and rapidly declining thereafter (Mata and Bähler 2006). No expression was observed in genome-wide vegetative cell cycle expression analyses (Rustici *et al.* 2004) and attempts to amplify the Cid11 cDNA from vegetative poly(A)$^+$ mRNA or from a vegetative cDNA library were unsuccessful (data not shown). These observations suggest that Cid11 is under extremely stringent transcriptional and/or translational control during vegetative growth and that this control may be counteracting the *nmt1* promoter-driven expression used in these assays. Though the reason for this tight transcriptional control is unknown, it is likely that Cid11 plays some role in the promotion of meiotic progression, for example by suppressing mitotic transcripts or by promoting the expression/translation of meiotic transcripts; either of these activities may well be deleterious to mitotic growth, necessitating the robust repression of Cid11 at these times. It is here noted that the

cid11 gene lies entirely within, but antisense to, the 3' UTR of the *avt5* gene (Wood *et al.* 2002; Wood *et al.* 2012) which is constitutively expressed during vegetative growth (Rustici *et al.* 2004) and may consequently repress *cid11* transcription, either by steric inhibition of converging RNA Pol II complexes or by the formation of dsRNA and consequent initiation of an RNAi response.

3.10.3 The structure and activity of Cid12

Due to the lack of assayed NTAP-Cid12 activity *in vitro* and the suppression of second uridylyltransferase activity in *cid1Δ cid12Δ* genetic backgrounds *in vivo* it is important to note that, while all available evidence supports the identification of Cid16 as the second uridylyltransferase, it remains formally possible that Cid12 possesses PUP activity and is responsible for the observed secondary uridylyltransferase activity. Indeed, ncPAPs possessing specific uridylyltransferase activity have been identified as components of RDRC-like complexes in several organisms. *C. elegans* CDE-1 resides in a complex with the RdRP EGO-1, the helicase DRH-3 and a Tudor-domain protein EKL-1 *in vivo* and, like Cid12, is required for efficient chromosome segregation; however, CDE-1 appears to direct the degradation rather than synthesis of siRNAs (van Wolfswinkel *et al.* 2009). In *Tetrahymena,* two PUPs, Rdn1 and Rdn2, reside in different RDRC-like complexes with the RdRP Rdr1 and, in the case of Rdn1-containing complexes, one of two accessory proteins, Rdf1 and Rdf2 (Lee *et al.* 2009). Importantly, these uridylyltransferases are proposed to enable the synthesis of full-length dsRNA products through a unique template 3' end looping mechanism (Talsky and Collins 2010). Perhaps of greatest relevance to the Cid12 pathway is the *C. elegans* ncPAP RDE-3 which, together with the RDRC-like DRH-3/EKL-1/(EGO-1 or RRF-1) complex and the putative exoribonuclease MUT-7, is required for the synthesis of WAGO class 22G siRNAs (Tabara *et al.* 1999; Chen *et al.* 2005; Gu *et al.* 2009). Though *in vitro* tethering assays could detect no catalytic activity for RDE-3 (Kwak and Wickens 2007), mutagenesis assays found that, as

with Cid12 (Motamedi *et al.* 2004), the catalytic residues of RDE-3 were essential for its activity, suggesting that this enzyme is active *in vivo* (Chen *et al.* 2005).

A number of observations support the identification of Cid16 over Cid12 as the second uridylyltransferase. Most striking is the robust and highly selective PUP activity of Cid16 (Figure 3.5), confirming that Cid16 is likely to act exclusively as a uridylyltransferase *in vivo*. In addition, the functional restriction of Cid12 to the exclusively nuclear RDRC and probable activity on nascent transcripts at heterochromatic genomic loci (Motamedi *et al.* 2004; Colmenares *et al.* 2007) suggest it is unlikely that Cid12 is able to act on mature mRNAs in the cytoplasm (though it is noted that one report identified checkpoint-regulatory functions of Cid12 not found in other members of the RDRC, suggesting a possible RDRC-independent activity; Win *et al.* 2006). Furthermore, mutagenesis studies of two residues known to be vital for UTP selectivity in tCid1, H336 (Figure 3.8) and N171 (Figure 3.9) have shown that, while Cid16-directed substitutions (H336K; N171 is conserved) retain processive and specific PUP activity, Cid12-directed substitutions (H336N, N171T) at these sites do not, instead exhibiting robust and specific PAP activity and attenuation (but not abolition) of all catalytic activities, respectively (Lunde *et al.* 2012; Yates *et al.* 2012). Interestingly, the *C. elegans* RDE-3 ncPAP exhibits identity with Cid1 at both these positions, suggesting that it may indeed be a hitherto unidentified PUP (personal observations).

A remarkable and somewhat unexpected observation was that the highly conserved Y212 residue, which stabilises protein-NTP interactions non-specifically via π-orbital stacking with the nucleic acid base, was not conserved in Cid12, which instead encodes a non-aromatic cysteine residue at this position (Figure 3.6). This residue, though not perfectly conserved, was uniformly aromatic (Y/F/W) in a wide variety of other ncPAPs exhibiting PAP and PUP activity, including Trf4p/Trf5p, trypanosomal mitochondrial TUTases and all mammalian ncPAPs. In addition, substitution of a number of residues that are

conserved among *S. pombe* ncPAPs of the Cid1 clade (Cid1/11/13/16) for their analogues in the Cid12/Cid14 clade (Figure 1.1) results in complete loss of tCid1 activity with all NTPs and greatly reduced expression/purification of these proteins, suggesting they may be structurally unstable (Figure 3.9, 3.10, 3.11). Cumulatively, these observations suggest that Cid1 and Cid12 exhibit significant structural and/or mechanistic differences in substrate binding and catalysis, further supporting the identification of Cid16 as the Cid1-like mRNA PUP in *S. pombe*. Mechanistic differences would not be unprecedented – trypanosomal mitochondrial and cytoplasmic TUTases (the only other ncPAPs for which the structures have been determined) coordinate UTP not via an H336-like direct interaction but through a network of H-bond stabilised H_2O molecules assembled by two conserved aspartates in the NTP binding pocket (Stagno *et al.* 2007; Stagno *et al.* 2010). Additionally, some (but not all) trypanosomal TUTases contain a large, ~150 aa insertion within their catalytic domain, which loops out to form a separate, solvent-exposed domain of unknown but essential function (Aphasizhev *et al.* 2002; Aphasizhev *et al.* 2004; Stagno *et al.* 2010).

It is thus concluded that, while it is possible that Cid12 acts as a PUP in *S. pombe*, it is highly unlikely to be responsible for the secondary uridylylation of *act1* mRNA. The nature of the catalytic activity of Cid12 and the role that this activity plays in the synthesis of siRNAs remains a major unresolved question in the study of RNAi and heterochromatin formation; experiments directed towards purification of native RDRC complexes and testing of substrates of varying dsRNA/ssRNA structure likely represent the best route forwards in the study of the enzyme.

Chapter 4
Characterisation of the Cid16 poly(U) polymerase

4.1 Summary

The identification of mRNA uridylyltransferase activity in Cid16 is important both for a complete understanding of the function of uridylylation in *S. pombe* and for the current understanding of Cid1, since both uridylyltransferases act on at least the *act1* transcript, and may consequently possess partial functional redundancy. However, Cid16 is almost completely uncharacterised in *S. pombe*, with the only available information coming from genome-wide studies of expression and subcellular localisation. The experiments presented in this chapter form the first steps towards a full biochemical and functional characterisation of Cid16.

Detailed analysis of NTAP-Cid16 activity revealed a highly selective poly(U) polymerase capable of selectively adding long homopolymeric poly(U) tails. Compared to NTAP-Cid1, NTAP-Cid16 possesses enhanced selectivity for both UTP and RNA substrate, exhibiting nucleotidyltransferase activity only on A_{15} and C_{15} substrates where it predominantly, though not exclusively, adds poly(U) tails. Consistent with Cid16-dependent uridylylation events observed *in vivo*, purified NTAP-Cid16 also exhibits greater processivity than NTAP-Cid1, preferring to add U_{5-8} tails compared to the U_{1-2} of NTAP-Cid1.

Analysis of the Cid16 sequence *in silico* reveals a single set of well-conserved catalytic residues located in the C-terminal part of the protein, with a long N-terminal extension which, though suggested by protein disorder predictions to be largely structured, exhibits no significant similarity to

any other known protein. This N-terminal extension is not required either for catalytic activity or for UTP selectivity of NTAP-Cid16, though it does enhance the activity of the protein. Deletion of a zinc-finger protein suggested by genome-wide genetic interaction studies to be functionally paired with Cid16 had no effect on polymerase activity in the assays utilised here. Finally, microarray-based transcriptome-wide expression analysis of a *cid16* deletion strain revealed an extremely small number of expression changes, suggesting that Cid16 may function independently of RNA stability, or may exhibit redundancy with the more active Cid1. However, two of the upregulated transcripts were found to be meiosis-specific genes lying within the 3' UTRs of genes involved in the regulation of RNAi and/or heterochromatin formation, suggesting an intriguing possible regulatory mechanism involving antisense transcript regulation of RNAi components.

4.2 NTAP-Cid16 is a highly selective poly(U) polymerase

NTAP-Cid16 was shown in the previous chapter to be a highly processive poly(U) polymerase when presented with an A_{15} substrate in the presence of 0.5 µM UTP. However, *in vivo*, UTP must compete with the other three nucleotides for interaction with Cid16. Furthermore, ATP is found in various eukaryotes in two- to ten-fold molar excess over other NTPs (Woodland and Pestell 1972; Osorio *et al.* 2003; Cao *et al.* 2005) suggesting that, *in vivo*, ATP may have a selective advantage over other NTPs.

To test whether Cid16 was able to synthesise homopolymeric poly(U) tails effectively in the presence of physiological concentrations of NTPs, two complementary assays were used. In both cases, NTAP-Cid16 was purified from *S. pombe* as previously described and incubated with 5'-^{32}P-A_{15} oligoribonucleotide substrate in the presence of 0.5 µM UTP alone or all four nucleotides with a ten-fold molar excess of ATP (0.5 µM ATP, 0.05 µM CTP, GTP and UTP). In the first assay, reaction products were incubated with RNAse A or RNAse T1, which cleave after pyrimidine (CMP and UMP) residues or after GMP residues respectively, or with an RNAse A/RNAse T1 mixture which generates oligo(A) structures terminating in a single CMP, GMP or UMP residue. The individual digestions

allow determination of the position of the first CMP/UMP or GMP residue in the non-templated tail whilst the double digestion shows the existence of any poly(A) tails, which are uniquely resistant to the RNAse A/RNAse T1 double digestion. In the second assay, polymerase reaction products were incubated with RNAse III, an enzyme that selectively degrades dsRNA in a sequence-independent fashion, and with a molar excess of unlabelled A_{15}, C_{15}, G_{15} or U_{15}. In the presence of non-templated homopolymeric sequences of each nucleotide, the complementary unlabelled oligoribonucleotide will base-pair with the tail and elicit degradation by RNAse III. Between them, these assays indicate both the efficacy of CMP/UMP incorporation at a single-nucleotide level and the ability of Cid16 to synthesise homopolymeric tracts efficiently. The results of these two assays are presented in Figure 4.1A and Figure 4.1B, respectively.

In the presence of UTP alone, both NTAP-Cid1 and NTAP-Cid16 efficiently synthesised long poly(U) tails which were resistant to RNAse T1 but completely degraded back to the A_{15} substrate by RNAse A alone or a combination of RNAse A/RNAse T1, as expected given the necessarily homopolymeric status of these tails (Figure 4.1A). In the presence of physiological concentrations of all four nucleotides, the degradation of polymerised tails was indistinguishable from that in the UTP-only assay, with all polymerase tails synthesised by both proteins cleaved at the first nucleotide by RNAse A. These observations confirm that both NTAP-Cid1 and NTAP-Cid16 are highly specific for the incorporation of either CMP or UMP as the first residue of a non-templated tail (Figure 4.1A). Though CMP and UMP cannot be distinguished in this assay, the strong enrichment of PUP activity over polycytidylyl polymerase (PCP) activity in NTAP-Cid1 assays and the absence of PCP activity from NTAP-Cid16 (Figure 3.5) suggest that the first residue is much more likely to be UMP than CMP.

In the presence of all four nucleotides at physiological concentrations, NTAP-Cid16 was able to synthesise long poly(N) tails that were almost entirely degraded by RNAse III in the presence of excess cold A_{15} (Figure 4.1B), confirming that the polymeric tails synthesised by NTAP-Cid16 *in vitro* are predominantly or entirely composed of UMP residues. Though a reduction in polymeric reaction

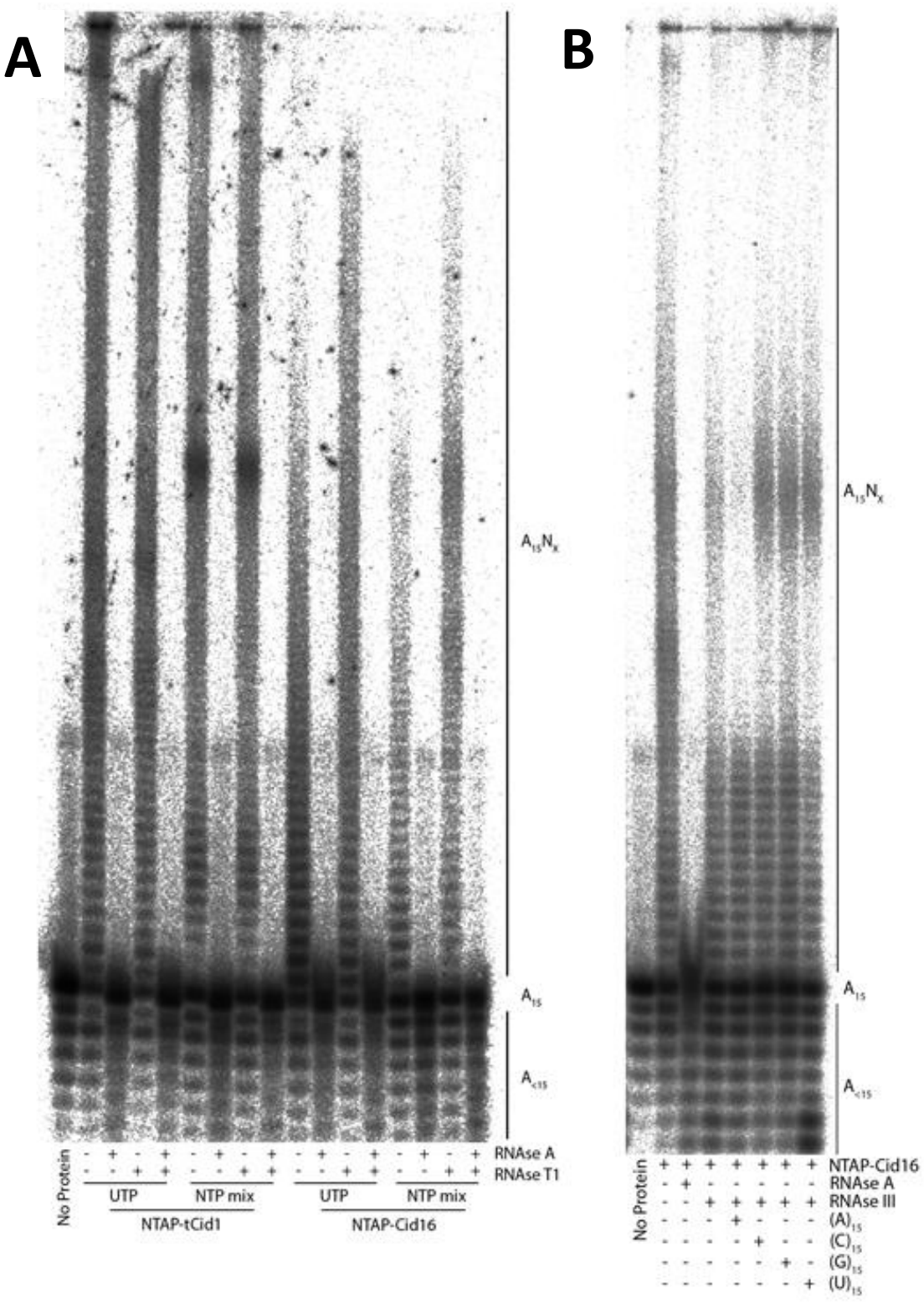


Figure 4.1: UTP selectivity testing shows Cid16 to be a highly-selective PUP *in vitro*. NTAP-Cid16 or NTAP-Cid1 were incubated with 5'-^{32}P-A_{15} in the presence of 0.5 µM UTP alone or an NTP mixture (0.5 µM ATP, 0.05 µM CTP, GTP, UTP). Reaction products were (A) treated with RNAse A, RNAse T1 or RNAse A/RNAse T1 as indicated or (B) with RNAse III alone or in the presence of excess unlabelled A_{15}, C_{15}, G_{15} or U_{15} as indicated. 5'-^{32}P-A_{15} was incubated with ATP but no polymerase as a negative control in both experiments (No Protein).

products was observed in the presence of C_{15}, G_{15} and U_{15} compared to reactions lacking RNAse III, this degradation is likely to reflect residual RNAse III-mediated degradation of poly(U) tails interacting with residual 5'-^{32}P-A_{15} substrate rather than the presence of minor populations of poly(C), poly(G) and poly(A)-extended substrates, since a similar level of degradation was seen in the presence of RNAse III alone without added oligoribonucleotides (Figure 4.1B). The combination of these assays shows that NTAP-Cid16 is highly selective for the incorporation of UMP residues both as the first residue of a non-templated tail and as the major constituent of long polymeric tails *in vitro*, confirming that, even in the presence of physiological concentrations of all four NTPs, NTAP-Cid16 is a highly specific PUP on oligo(A) substrates.

4.3 NTAP-Cid16 is able to distinguish between and select certain RNA substrates

All analyses conducted thus far have utilised 5'-^{32}P-A_{15} as a substrate in order to most closely recapitulate known *in vivo* substrates of Cid1 and Cid16, namely polyadenylylated mRNA (Rissland 2008). However, given that Cid16 has been shown *in vivo* in **Chapter 3** to uridylylate non-adenylated mRNA species, it appears that a poly(A) tail is not obligatory for its PUP activity. Furthermore, previous investigations have shown that the use of a U_{15} rather than A_{15} substrate can change the processivity and/or nucleotide specificity of NTAP-Cid1 (Rissland 2008), suggesting that RNA substrate-dependent polymerase activity may be a feature of cytoplasmic PUPs. As such, the activity and nucleotide specificity of NTAP-Cid16 on two other homopolymeric oligonucleotides (C_{15} and U_{15}) was tested using the polymerase activity assay described previously. In order to test the hypothesis that cytoplasmic Cid family members may exhibit RNA substrate-dependent polymerase activity, NTAP-Cid1 and NTAP-Cid13 were also tested in parallel. Whilst polymerase assays were attempted with G_{15} as a substrate, technical complications relating to poor 5' labelling by polynucleotide kinase and/or degradation of G_{15} meant that reliable activity measurements could not be obtained with this substrate (data not shown).

NTAP-Cid1 was previously reported to exhibit significantly enhanced PAP activity and slightly enhanced PCP and polyguanidylyl polymerase (PGP) activity on a U_{15} substrate compared to an A_{15} substrate (Rissland 2008). Though the assays conducted here broadly recapitulated those observations, a reduction in PUP activity on the U_{15} substrate compared to A_{15} was observed in the current study (Figure 4.2A). Furthermore, this study describes a greater level of PGP activity than previously reported, resulting in accumulation of products containing a $G_{6\text{-}9}$ tail (Figure 4.2A). Interestingly, this PGP activity was observed in the previous study using recombinant tCid1, suggesting that differences in reaction conditions and/or the proteins copurified with NTAP-Cid1 may be responsible for these differences. Despite these differences, both this and the previous study (Rissland *et al.* 2007) support an hypothesis in which, whilst NTAP-Cid1 is still a PUP on U_{15} substrate, its permissivity for other polymerase reactions is enhanced. By contrast, no polymerase activity with any NTP was seen with either NTAP-Cid16 or NTAP-Cid13 on the U_{15} substrate, despite the fact that the same preparations were active on A_{15} (data not shown), suggesting that NTAP-Cid16 and NTAP-Cid13 act only on certain substrates and are able to exclude U_{15}.

Assays using 5'-^{32}P-C_{15} showed that both NTAP-Cid1 and NTAP-Cid16 were active on this substrate; however, they exhibited remarkably different activity profiles. Unlike its activity on A_{15}, NTAP-Cid1 was a poor PUP on C_{15}, adding up to seven UMP residues to a small proportion of template molecules. Furthermore, PAP activity was completely abolished. By contrast, both PCP and, most noticeably, PGP activity were significantly enhanced on the C_{15} template, with NTAP-Cid1 adding up to approximately 15 CMP residues and creating a smear of template products by PGP activity that included high molecular weight species at the top of the gel (Figure 4.2B). The smearing is likely an

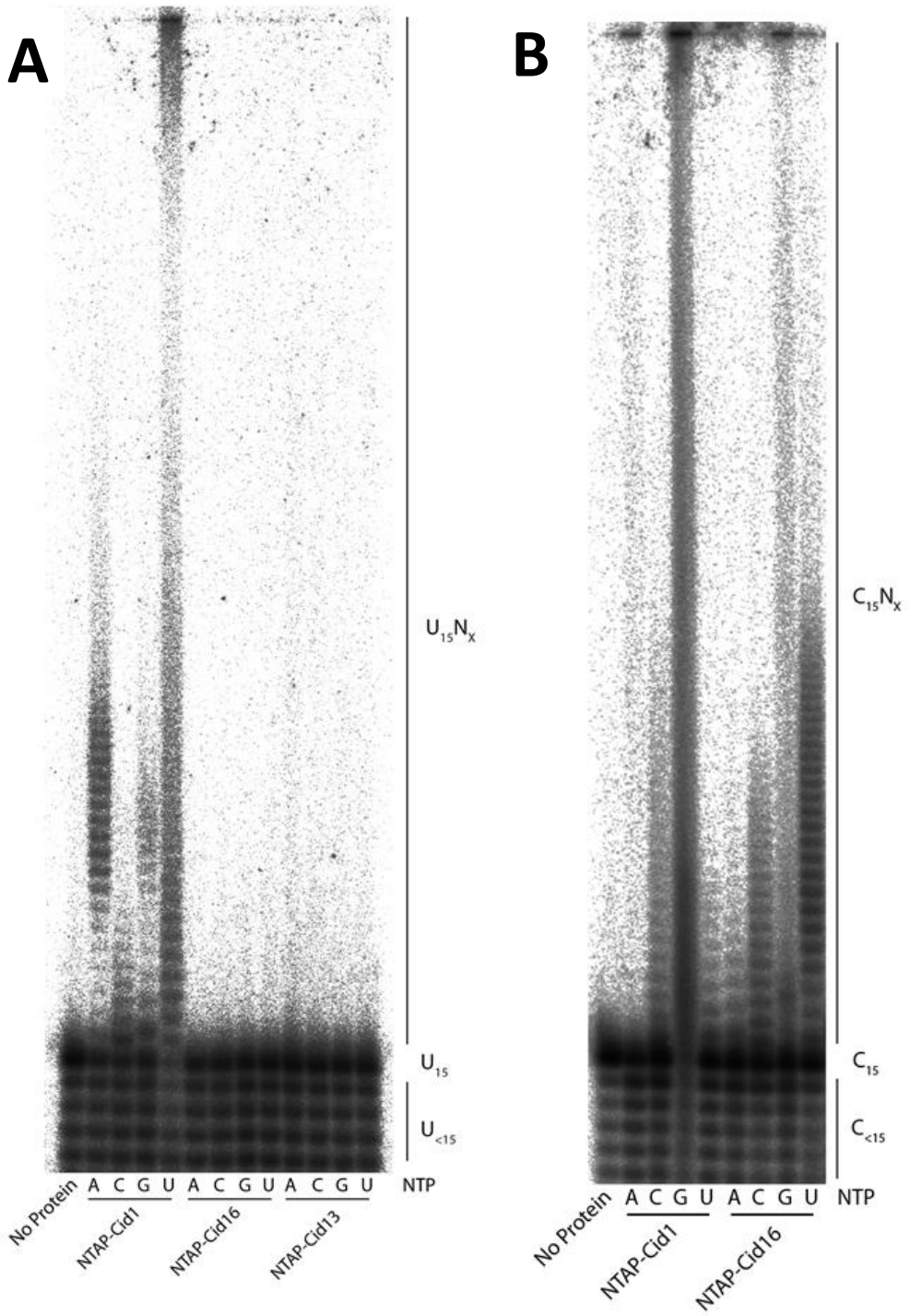


Figure 4.2: *cid* family members exhibit different activities on different RNA substrates. NTAP-Cid1, NTAP-Cid13 or NTAP-Cid16 were incubated with 5'-^{32}P-U_{15} or 5'-^{32}P-C_{15} in the presence of ATP, CTP, GTP or UTP (A, C, G, and U respectively). The relevant labelled template was incubated with ATP but no polymerase as a negative control (No Protein).

artefact of the retarded migration of 3' GMP-modified species discussed in **3.6** above which results in a series of non-overlapping ladders initiated from C_{15} as well as 3'-truncated $C_{<15}$ species. Notably, whilst NTAP-Cid16 similarly exhibited enhanced PCP and PGP activity on the C_{15} substrate, these effects were more modest, with the majority of modified substrate molecules containing up to twelve additional CMP residues and only two to three GMP residues. Unlike with NTAP-Cid1, PAP activity was enhanced with NTAP-Cid16, resulting in the addition of up to five AMP residues. NTAP-Cid16 also retained significantly more PUP activity than NTAP-Cid1, adding 25-30 UMP residues to the C_{15} substrate. Critically, NTAP-Cid16 still exhibited greatest activity with UTP as a substrate on C_{15}, suggesting that its predominant activity on C_{15} in the presence of all four NTPs is polyuridylylation. Unfortunately, this hypothesis could not be fully tested, since degradation of C_{15} by RNAse A precluded the performance of UTP selectivity assays on this substrate.

From these observations, it is suggested that NTAP-Cid1 nucleotidyltransferase activity is much more dependent on substrate sequence than was previously expected. Furthermore, whilst NTAP-Cid16 exhibits some substrate-dependent activity fluctuations, it is significantly more selective than NTAP-Cid1 both for NTP and for oligonucleotide substrate and thus may represent a higher-stringency PUP.

4.4 Cid16 is a more processive PUP than Cid1

Prior to this study, only four uridylylation events had been retrieved from *cid1Δ* strains (Rissland *et al.* 2007; Rissland 2008), preventing meaningful statistical analysis of Cid16 uridylylation activity *in vivo*. Pooling and analysis of observed *act1* mRNA uridylylation events in all *cid1Δ* strains presented in **Chapter 3** demonstrated that uridylyl tails added by Cid16 are, on average, significantly longer than those added by Cid1 (2.54 ± 2.40 nt vs. 1.28 ± 0.46 nt; $p < 0.005$ by Student's *t*-test). In addition, whilst no uridylyl tails added by Cid1 (NTAP-Cid1 overexpression strain) exceeded 2 nt in length, 36.3% of all *cid1Δ* uridylyl tails were 3 nt or longer (Figure 4.3A). Given the low frequency of uridylylation by Cid16 on *act1* mRNA (4.9% in *cid1Δ cid13Δ*; Table 3.1), any uridylyl tails observed in

cid1Δ cells are likely to represent individual uridylylation events rather than repeated rounds of mono- or di-uridylylation. These observations therefore suggest that Cid16 may be a more processive PUP than Cid1.

Given this observation, poly(U) polymerase assays using the A_{15} substrate were performed in which NTAP-Cid1 and NTAP-Cid16 were purified in parallel and resolved on the same gel (total of three independent experiments) and the size distribution of poly(U) polymerase products was compared (Figure 4.3B). The *in vitro* polymerase assay products aligned well with the uridylyl tails retrieved for each enzyme *in vivo* – NTAP-Cid1 assays exhibited abundant mono- and di-uridylylated products that were markedly more abundant than products with longer tails. By contrast, mono- and di-uridylylated products were almost entirely absent from NTAP-Cid16 assays, which instead showed the highest abundance among products with 5-8 nt extensions (Figure 4.3B). The lack of $A_{15}U$/$A_{15}U_2$ products in the NTAP-Cid16 assays was somewhat surprising given that the most common uridylylation event in *cid1Δ* strains *in vivo* was monouridylylation. This difference is likely to be a consequence of the greatly increased effective concentration of Cid16 in the assay, which promotes processivity and/or multiple rounds of uridylylation; however, it could also be due to biological factors including a different affinity of Cid16 for A_{15} and mRNA, or the loss of an exonuclease or processivity-regulating partner of Cid16 during the purification.

To test whether this difference was statistically significant, profiles of ^{32}P signal intensity along the length of the NTAP-Cid1 and NTAP-Cid16 lanes were generated. These were normalised for total ^{32}P signal in the lane, then bins for each individual band from $A_{15}U$ to $A_{15}U_{15}$ were identified and the total radioactivity per band was determined for each protein. Finally, background signals for each band size were determined from a 'No Protein' control lane and the enrichment of each product band over background was determined for NTAP-Cid1 and NTAP-Cid16. This analysis allows the relative quantity of uridylylation products in NTAP-Cid1 and NTAP-Cid16 assays to be directly compared at single-base resolution (Figure 4.3C).

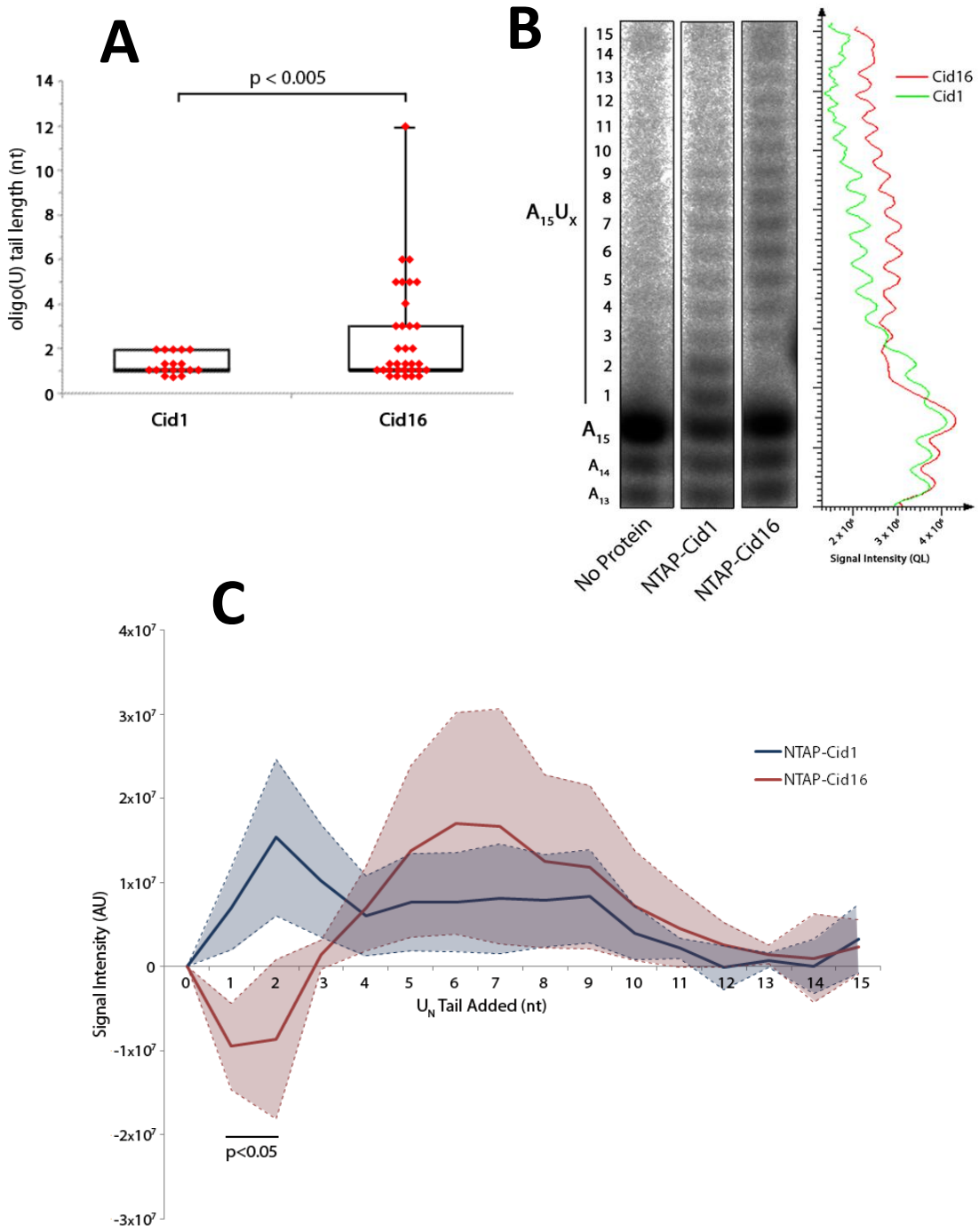


Figure 4.3: Cid16 is a more processive PUP than Cid1. (A) All polyuridylylation events attributable to *cid1* or *cid16* from **Chapter 3** were retrieved and compared. Red diamonds represent individual uridylylation events; boxes show median, upper quartile and maximum values; *p*-values determined using Student's *t*-test. (B) NTAP-Cid1 or NTAP-Cid16 were incubated with 5'-^{32}P-A_{15} and UTP and the densitometry profile for sizes 15-30 nt determined using AIDA Image Analysis software. (C) Mean densitometry profiles for three experiments (representative experiments presented in (B) were averaged and compared following normalisation for total loaded radioactivity and to the No Protein control lane densitometry; solid lines represent average densitometry for the three experiments, shaded area ± 1 s.d.. *p*-values were determined using Student's *t*-test.

Importantly, NTAP-Cid1 showed a statistically significant enrichment for mono- and di-uridylylation products compared to NTAP-Cid16 ($p < 0.05$; Student's t-test). It is noted that the negative values for NTAP-Cid16 in these bands, indicative of a lower ^{32}P signal in the NTAP-Cid16 lane than the No Protein control, are likely an artefact of bleeding of the intense A_{15} band into the $A_{15}U$ and $A_{15}U_2$ bins in the No Protein control lane; this bleed is reduced in the NTAP-Cid1 and NTAP-Cid16 lanes due to the shifting of a proportion of total signal further up the gel (Figure 4.3B). Whilst NTAP-Cid16 consistently exhibited a greater accumulation of U_5-U_8 products in this analysis, the enrichment fell below the level of statistical significance. From this analysis, and from the mRNA sequencing observations discussed above, it is concluded that Cid16 adds longer oligo(U) tails than Cid1 both *in vitro* and *in vivo*.

4.5 *in silico* analysis of the Cid16 protein

Despite the extensive biochemical characterisation carried out above, relatively little is known about the nature of the Cid16 protein, its structure and how these relate to possible functional roles. As an initial approach towards characterising the Cid16 protein, a suite of *in silico* prediction and analytical tools were used to investigate various aspects of Cid16 structure, sequence and homology within and across species.

The Cid16 protein is a large protein of 1202 aa with a M_W of 138.5 kDa and pI of 9.00 predicted by the ExPASy ProtParam tool (Gasteiger *et al.* 2005), suggesting it may be positively charged *in vivo*. Cid16 is almost twice the size of the next largest *S. pombe* Cid family member, Cid14, at 684 aa, though it is of similar size to PUPs retrieved from other species. Analysis with the domain prediction programme CDD (Marchler-Bauer *et al.* 2011) locates the catalytic NTP transferase domain (NTD) and poly(A) polymerase/2',5'-oligoadenylate synthase associated (PAP/25A) domain towards the C-terminus of the protein, resulting in the presence of a long, ~900 aa N-terminal extension. Homology searching with the Basic Local Alignment Search Tool (BLAST) and Position-Specific Iterated BLAST (PSI-BLAST; Altschul *et al.* 1990; Altschul *et al.* 1997) demonstrated that, whilst the C-

terminal catalytic domains are well conserved (as seen in Figure 3.6) and can retrieve Cid family members from a variety of species with high statistical significance, the long N-terminal extension retrieves no sequence in any species with similarity that falls below the significance threshold ($E < 0.001$), suggesting that the N-terminal extension sequence is unique to *S. pombe*.

The protein disorder prediction tool DISOPRED (Ward *et al.* 2004) predicted that the N-terminal extension is largely structured, with only a limited, ~200 aa region of partial disorder immediately N-terminal to the catalytic domains (Figure 4.4A); this prediction was replicated in several other disorder prediction programmes using different algorithms (FoldUnfold, IUPred; Dosztányi *et al.* 2005; Galzitskaya *et al.* 2006). The protein folding prediction programme PSIPRED (Jones 1999) predicted a large structured, mostly α-helical N-terminal region for Cid16 (full prediction provided in Appendix E) in addition to recapitulating with reasonable accuracy the structure of the catalytic domains predicted by sequence alignment with the X-ray diffraction structure of Cid1 (Lunde *et al.* 2012; Munoz-Tello *et al.* 2012; Yates *et al.* 2012). The apparently highly structured nature of this N-terminal extension was somewhat unexpected given that several other known or suspected PUPs with N-terminal extensions, including the Cid16 homologue SjPAP in the closely related fission yeast *Schizosaccharomyces japonicus*, are predicted to have a partially or completely unstructured N-terminal region (Figure 4.4B,C); however, it is noted that the human PUP ZCCHC6 (Figure 4.4D) shows a highly structured region in its N-termini corresponding to catalytically-inactive duplication of the catalytic domains (Rissland 2008).

Despite this apparently well-structured N-terminus, both the PRODOM (Bru *et al.* 2005) and InterProScan (Quevillon *et al.* 2005) domain prediction programmes found no identifiable domain motifs or structures in this region. While InterProScan identified a possible signal peptide at the extreme N-terminus of the Cid16 protein, survey with the specialised signal peptide prediction tool SignalP (Petersen *et al.* 2011) found this prediction to be below the threshold of significance. In addition, specialised prediction programmes for specific structural motifs including transmembrane

helices (HMMTOP; Tusnady and Simon 2001) and coiled-coil motifs (COILS; Lupas *et al.* 1991) found no evidence of these structures in the Cid16 N-terminal extension. Surprisingly, the CDD domain prediction programme (Marchler-Bauer *et al.* 2011) predicted a second TRF4-like multidomain in the N-terminal extension, extending from 315-899 aa in the Cid16 protein (Figure 4.5), suggesting that Cid16 may exhibit the dual catalytic domain arrangement observed in human ZCCHC11 and ZCCHC6 (Rissland 2008) and *C. elegans* CDE-1 (observations herein). However, examination of the sequence in this predicted region shows that much of the similarity lies in the extreme N- and C-terminal sequences of the TRF4 multidomain with very poor sequence conservation in the central, catalytic domain-containing regions. No conserved ncPAP motif can be identified in this region suggesting that, if it does represent a duplicate, inactive catalytic domain, it is far more divergent than those observed in other duplicate-domain proteins. Alternatively, it is possible that Cid16 contains a single conserved TRF4 multidomain element that has undergone structural rearrangement resulting in the relocation of the core catalytic domains to the C-terminus while leaving the N- and C-termini *in situ* (Figure 4.5).

Finally, observations on Cid16 transcript expression and localisation were retrieved from a set of reported genome-wide analyses. Cid16 expression was found to be constant throughout the cell cycle and not to fluctuate during meiosis/sporulation, pheromone response or in response to several different environmental stressors including oxidative stress, cadmium exposure, heat stress and sorbitol treatment (Mata *et al.* 2002; Chen *et al.* 2003; Rustici *et al.* 2004; Mata and Bähler 2006), suggesting that Cid16 is a constitutively expressed protein. Interestingly, a genome-wide YFP-fusion protein localisation study (Matsuyama *et al.* 2006) found that an overexpressed Cid16-YFP fusion protein localised to a small number of cytoplasmic granules in *S. pombe*. Though the identity of these granules is currently unclear, this specific subcellular localisation suggests that Cid16 activity *in vivo* may be restricted to certain subcellular compartments and, consequently, to the subset of cellular RNAs contained therein.

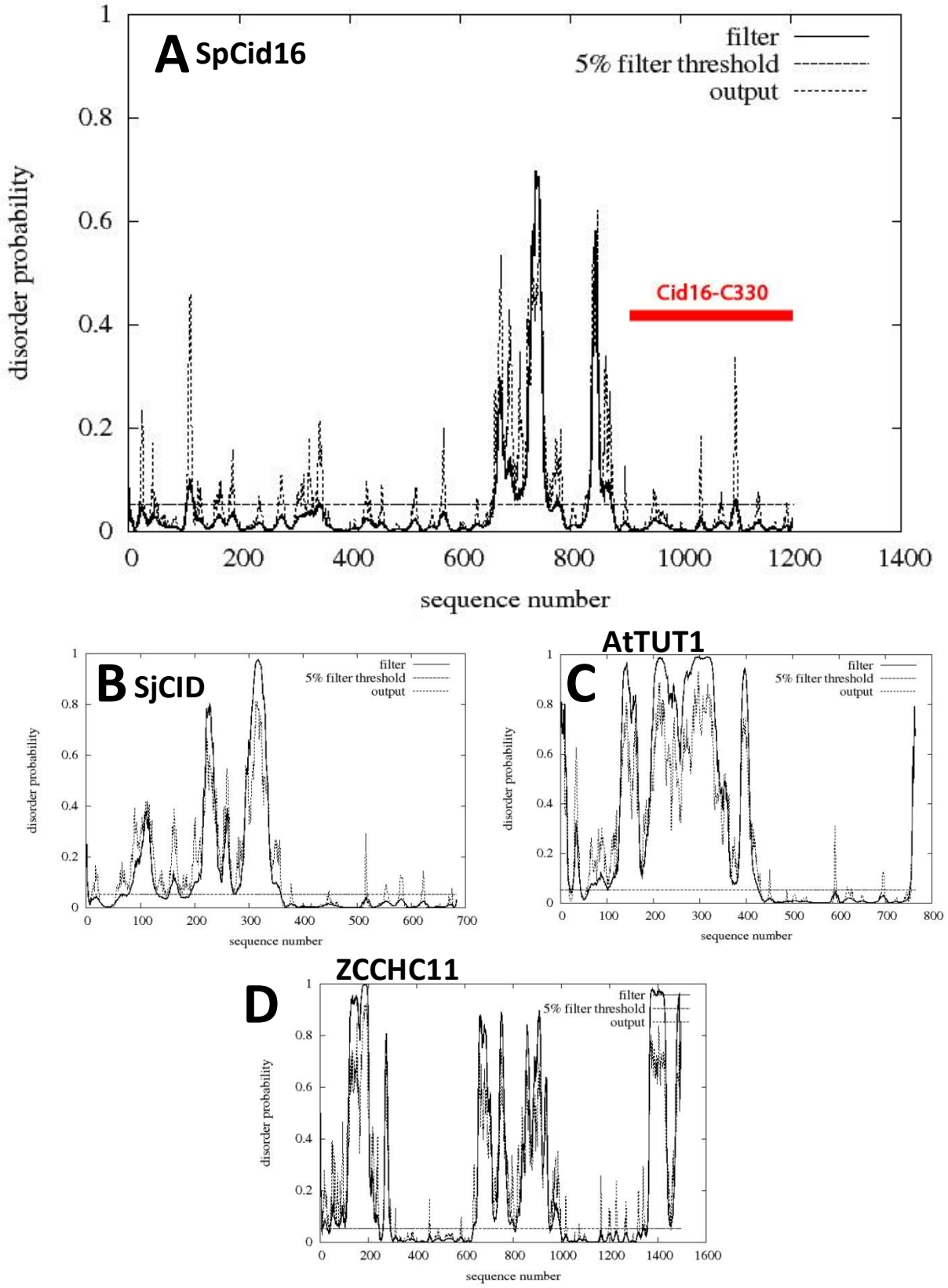


Figure 4.4: Structure/disorder predictions for several ncPAP family members possessing long N-terminal extensions. The full-length sequences of (A) *S. pombe* Cid16, (B) *S. japonicus* PAP/Cid16, (C) *A. thaliana* TUT1 and (D) human ZCCHC6 were analysed for disorder likelihood using the DISOPRED algorithm (Ward *et al.* 2004). The location of the Cid16-C330 construct used below is indicated in (A).

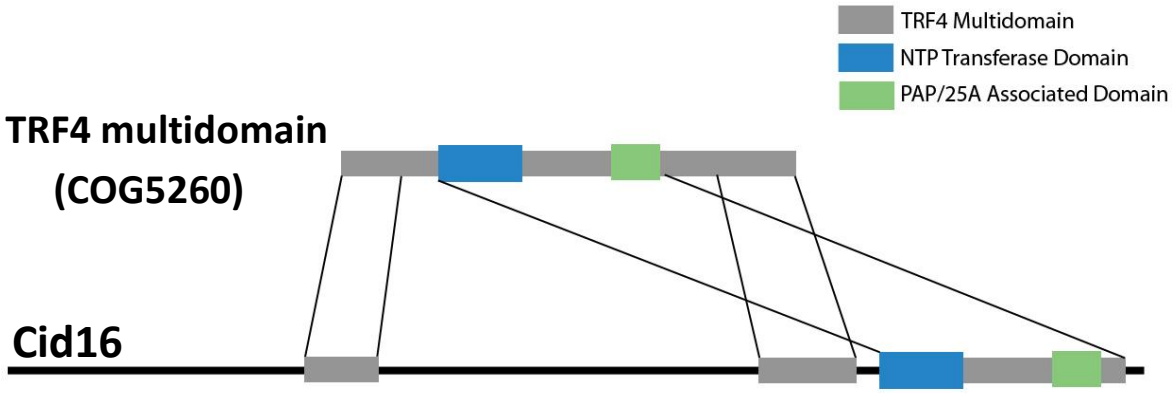


Figure 4.5: Possible TRF4 multidomain rearrangement in *S. pombe* Cid16. Arrangement of multidomain fragments is based on alignment of the CDD TRF4 multidomain sequence (CDD ID: COG5260) with the putative N-terminal TRF4 multidomain of Cid16 predicted by the CDD domain prediction programme (Marchler-Bauer *et al.* 2011) and the separate C-terminal catalytic domain.

4.6 The Cid16 N-terminal extension is not required for uridylylation activity

On the basis of the above observations, it was hypothesised that the potentially highly-structured N-terminal extension of Cid16 may play important role(s) in the activity and/or function of this protein, for example by controlling nucleotide or RNA template specificity, processivity or specific activity, or other key functional requirements such post-translational regulation or sub-cellular localisation. In order to test the effect of the Cid16 N-terminal extension on the activity of the enzyme, two different truncations containing only the C-terminal catalytic domains were constructed, Cid16-C371 (containing the last 371 aa of Cid16, including the region orthologous to full-length Cid1) and Cid16-C330 (containing the last 330 aa, the region orthologous to recombinant tCid1). These constructs were cloned into the *S. pombe* pREP1-NTAP expression vector and used to express N-terminal truncated Cid16 fragments in parallel with full-length Cid16. These three proteins were purified and quantified by western blotting for the NTAP tag as described previously (Figure 4.6B). All three proteins were overexpressed and purified successfully; NTAP-Cid16-C371 and -C330 were detected at significantly higher concentrations than full-length NTAP-Cid16 (marked by the arrow in Figure 4.6B), though the quantity of NTAP-Cid16 is likely underrepresented due to inefficient transfer of the ~160 kDa NTAP-Cid16 protein from the 10% SDS-PAGE gel to the membrane.

Polymerase activity assays on 5'-^{32}P-A_{15} with the three forms of Cid16 are presented in Figure 4.6A. Interestingly, the Cid16-C330 truncation, but not the longer Cid16-C371 truncation, retained nucleotidyltransferase activity on the A_{15} substrate. Though NTAP-Cid16-C330 appeared to be less processive and/or active than full-length NTAP-Cid16 (as evidenced by the lack of high M_W products towards the top of the gel), it was still capable of adding long poly(U) tails to the RNA substrate *in vitro*. Furthermore, the NTAP-Cid16-C330 truncation retained the high substrate specificity of full-length Cid16, as evidenced by the lack of PAP, PCP or PGP activity. It is thus concluded that, though the long N-terminal extension of Cid16 may enhance the activity of Cid16, it is not essential for this activity or for UTP selectivity, suggesting these characteristics are innate in the catalytic domains.

The difference between these two truncations suggests that the 41 aa region present only in NTAP-Cid16-C371 is likely to have a destabilising and/or inhibitory effect on Cid16-C330 activity; given that this region is predicted by DISOPRED to be partially disordered (Figure 4.4A), it is possible that a folding defect in this protein is responsible for the lack of activity. It is noted that, in recombinant Cid1 expressed in *E. coli*, proteins possessing the 31 aa corresponding to this region were unstable and experienced N- and C-terminal degradation, supporting this hypothesis (Rissland *et al.* 2007).

4.7 A putative Cid16 functional partner is not required for activity in vegetative cells

During the course of investigations into the function and activity of Cid16, a large-scale epistasis map (E-MAP) screen of ~60% of the *S. pombe* genome was published (Ryan *et al.* 2012). As well as providing detailed genetic interaction profiles for individual genes, this work identified genes that are likely to reside in the same functional complex/pathway by their common genome-wide genetic interaction profiles. Of particular relevance to this study, *cid16* was reported to show a genetic interaction profile significantly overlapping with that of a previously uncharacterised gene, *SPAC13F5.07c*, leading the authors to identify a putative heterodimeric functional complex of unknown function containing these two proteins (Ryan *et al.* 2012).

SPAC13D5.07c encodes a small, 18.5kDa protein containing a PARP-type zinc finger, a motif initially identified in poly(ADP)-ribose polymerase (PARP), where it interacts with single-stranded DNA breaks (Gradwohl *et al.* 1989; Mazen *et al.* 1989; Gradwohl *et al.* 1990; Molinete *et al.* 1993), and subsequently found in human DNA ligase III (Wei *et al.* 1995). Genome-wide YFP-tag studies show SPAC13F7.05c-YFP localised to both the nucleus and the cytosol (Matsuyama *et al.* 2006) whilst transcriptome-wide expression studies demonstrate transient induction of *SPBC13F5.07c* expression early in meiosis (Mata *et al.* 2002), as well as robust induction in response to oxidative stress (Chen *et al.* 2003). Though PARP-type zinc fingers have so far only been shown to interact with DNA, the partial localisation of this protein to the cytoplasm, as well as overlapping *SPAC13F5.07c* genetic

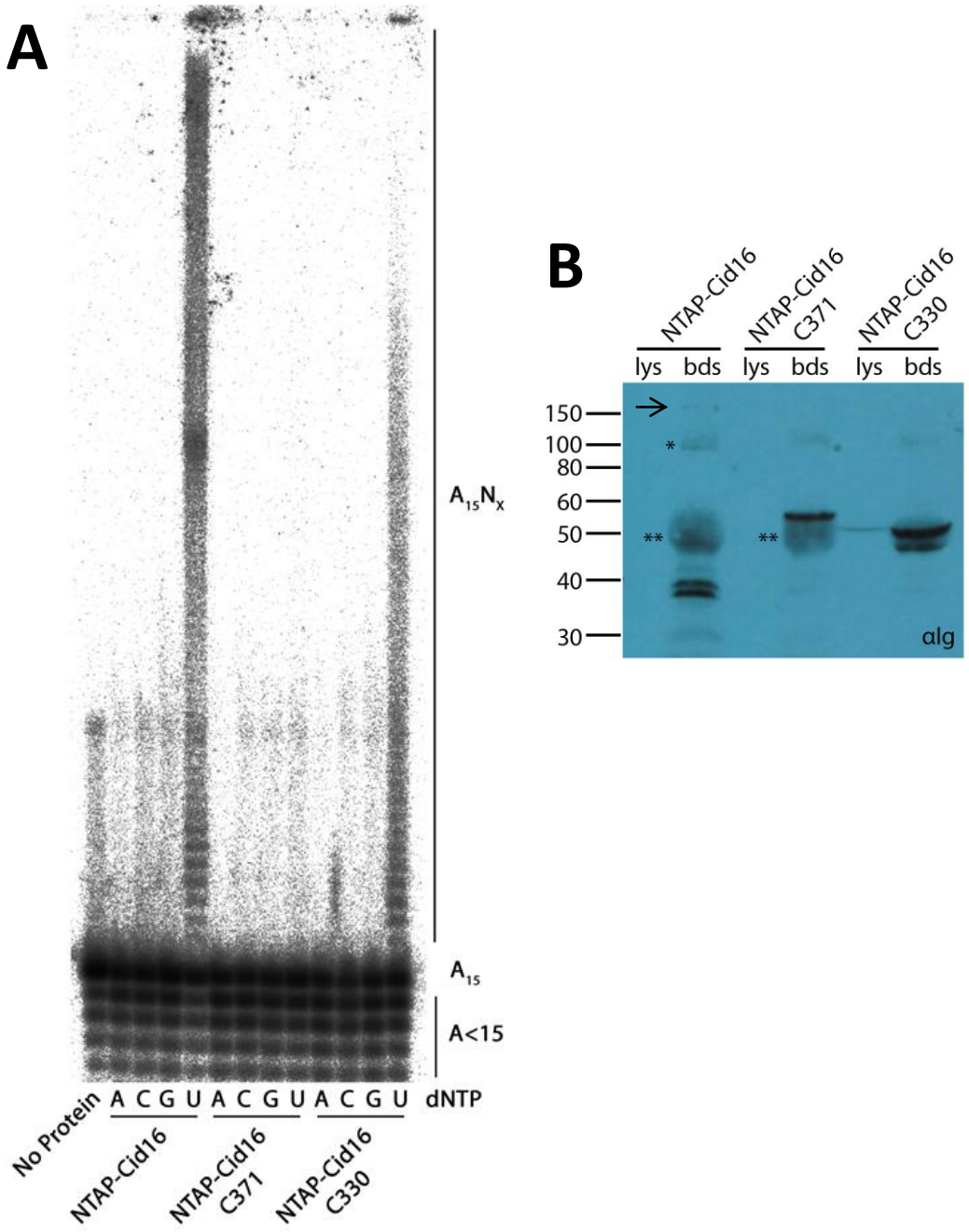


Figure 4.6: The Cid16 N-terminal extension is not required for catalytic activity or UTP selectivity. Full-length NTAP-Cid16, NTAP-Cid16-C371 or NTAP-Cid16-C330 were incubated with 5'-^{32}P-A_{15} and either ATP, CTP, GTP or UTP (A, C, G and U, respectively. 5'-^{32}P-A_{15} was incubated with ATP but no polymerase as a negative control (No Protein). (B) Successful purification of proteins was confirmed by western blotting. Arrow indicates full-length NTAP-Cid16; * = non-specific band; ** = probe antibody crossreactivity with IgG heavy chain derived from the IgG-Sepharose beads used in purification.

interaction profiles with *cid16Δ*, suggests that this protein may represent an interaction partner of Cid16, in particular as a possible RNA binding and/or specificity factor.

To test this hypothesis, the pREP1-NTAP-Cid16 vector was transformed into a *SPAC13F5.07c::kanMX leu1-32 ura4-D18 ade6-M210* strain (a kind gift of Dr. Timothy Humphrey) and NTAP-Cid16 purified from these cells was assayed for nucleotidyltransferase activity in comparison to NTAP-Cid16 purified from *cid1Δ SPAC13F5.07c*$^{+}$ cells. However, deletion of *SPAC13F5.07c* was found to have no effect on the *in vitro* activity of NTAP-Cid16, which was able to extend the A_{15} substrate with similar processivity and nucleotide specificity following purification both genetic backgrounds (Figure 4.8). Whilst it is important to note that these observations do not necessarily preclude an involvement of SPAC13F5.07c in the activity of Cid16, either under certain cellular conditions (such as during meiosis or following oxidative stress, conditions inductive for *SPAC13F5.07c* expression) or in some aspect that is not assayed under the assay conditions here (interaction with mRNA targets or with other complex subcomponents), it is concluded that SPAC13F5.07c is not required for efficient *in vitro* PUP activity of NTAP-Cid16 purified from exponentially growing cells.

4.8 Cid16 regulates the abundance of few mRNAs in exponentially growing *S. pombe*

Due to the sequence-specific nature of the cRACE protocol used to identify Cid16 uridylylation events *in vivo*, the only RNA currently known to experience Cid16-mediated uridylylation is the *act1* mRNA (**3.2** above). Given the similarity of uridylylation patterns between Cid1 and Cid16, and the previously-described role of Cid1-mediated uridylylation in RNA degradation (Rissland and Norbury 2009), it was considered likely that Cid16-mediated uridylylation may also signal for degradation. In order to assay for possible degradative targets of Cid16, *wt* and *cid16Δ* total RNA samples were compared on a transcriptome-wide *S. pombe* microarray to identify transcripts whose levels varied significantly upon *cid16* deletion. This work was conducted with the assistance of Ms. Ayesha Hasan and Dr. Juan Mata, Department of Biochemistry, University of Cambridge.

Prototrophic *h+ wt* and *h+ cid16::kanMX* strains were grown in fully supplemented minimal media to mid-log phase and total RNA was harvested. These RNAs were used to synthesise Cy3- or Cy5-labelled cDNA primed with a combination of random hexamers and $(dT)_{20}$ to enrich for mRNA sequences, and the cDNAs were hybridised to a custom Agilent microarray prepared for the Mata Lab representing the full *S. pombe* mRNA transcriptome as well as all characterised functional ncRNAs and an extensive range of uncharacterised ncRNAs including antisense transcripts, CUTs and intergenic transcripts. Two biological replicates from cells grown on different days were performed, data from both arrays parsed and collectively analysed with Agilent GeneSpring GX software.

Comparison of the two biological replicates showed a high degree of replicability between the samples (Figure 4.8). Surprisingly, *cid16Δ* RNA samples exhibited an extremely low level of variation compared to the *wt* RNA sample, with only three genes showing >2-fold upregulation in both biological replicates (Table 4.1). The only mRNA showing reproducible >2-fold downregulation was that of the deleted *cid16* gene, whose signal was reduced to background; however, two additional transcripts were identified which represented possible downregulation targets which in one array showed significant downregulation and in the other were not detected (Table 4.2). The only known Cid16 uridylylation target, *act1,* showed no expression changes in response to *cid16* deletion (1.02- and 1.05-fold upregulated in the two replicates, respectively), though the low level of uridylylation on this target is unlikely to represent a proportion whose degradation would be detectable in this system. Though none of the genes identified in these arrays were also upregulated in *cid1Δ* microarrays (**5.2**), it was noted that *isp5*, one of the most significantly upregulated transcripts in a *cid1* deletion, was slightly but reproducibly upregulated in these arrays (1.52 and 1.90-fold in replicates 1 and 2, respectively; mean upregulation 1.71-fold). Cumulatively these observations suggest that, at least in the limited growth conditions tested here (exponential growth in minimal media with prototrophic strains), deletion of *cid16* has minimal effect on the stability of the vast majority of *S. pombe* mRNAs.

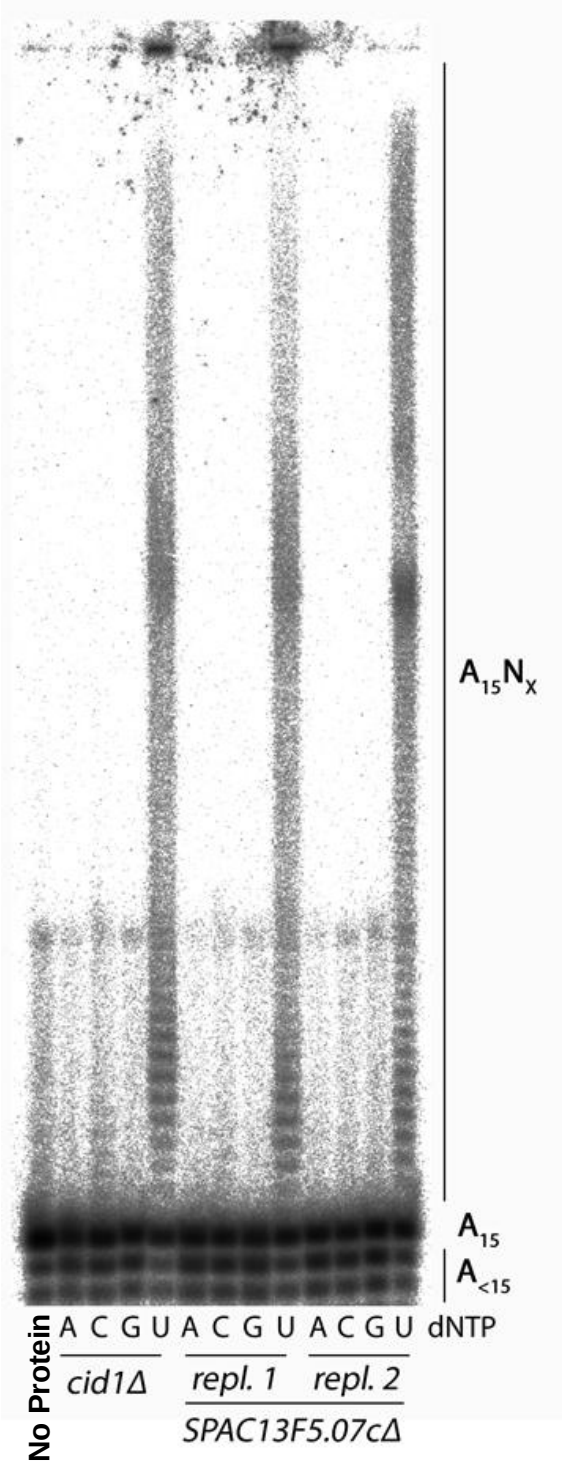


Figure 4.7: SPAC13F5.07c is not required for the activity of NTAP-Cid16 *in vitro*. NTAP-Cid16 was purified from *wt* or *SPAC13F5.07cΔ* cells and incubated with 5'-^{32}P-A_{15} in the presence of either ATP, CTP, GTP or UTP (A, C, G and U, respectively). 5'-^{32}P-A_{15} was incubated with ATP but no polymerase as a negative control (No Protein).

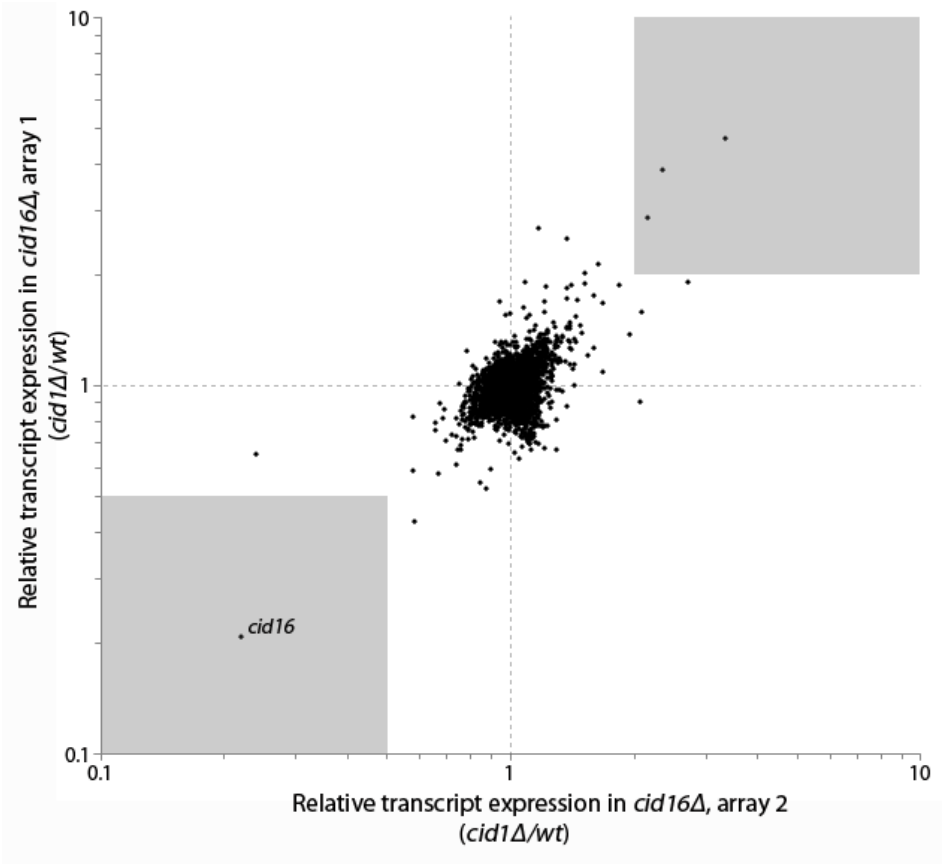


Figure 4.8: *wt:cid16Δ* expression arrays reveal few transcripts showing reproducible up/downregulation. Relative expression of transcripts in *cid16Δ* vs. *wt* was compared between replicates. Grey shading indicates regions exhibiting reproducible >2-fold up- or down-regulation.

Table 4.1: List of genes exhibiting reproducible >2-fold upregulation in *cid16Δ* strains.

Systematic Name	Common Name	Expression Change		Mean Expression Change	Comments
		Run 1	Run 2		
SPCC188.12	*spn6*	3.33	4.70	4.02	meiotic septin
SPBC428.07	*meu6*	2.33	3.87	3.11	meiotic chromosome segregation protein
SPBC1347.03	*meu14*	2.16	2.88	2.52	sporulation protein; genetic interaction with *spn6*

Table 4.2: List of genes exhibiting reproducible >2-fold downregulation in *cid16Δ* strains.

Systematic Name	Common Name	Expression Change		Mean Expression Change	Comments
		Run 1	Run 2		
SPAC17H9.01	*cid16*	0.22	0.21	0.21	*cid16Δ* strain
SPAC513.03	*mfm2*	0.05			M-factor precursor
SPAC1F8.03	*isp3/meu4*	0.05			sequence orphan; conjugation protein

The extremely small number of genes that exhibited significant expression changes in response to *cid16* deletion precludes statistical analysis. However, it is noted that all three genes upregulated in response to *cid16* deletion play important roles in meiotic chromosome regulation (Table 4.1), suggesting that *cid16* may play a role in the regulation of meiosis and/or the chromosome dynamics therein. More strikingly, the two most upregulated genes, *spn6* and *meu6*, are genomically located in closely-spaced convergent gene pairs respectively with *dcr1* and *clr4*, two central regulators of the RNAi response in *S. pombe* (reviewed in **1.5**); indeed, genomic annotations of transcript UTRs suggest that *spn6* and *meu6* lie entirely within the 3' UTRs of these two genes (Figure 4.9A,B respectively). This observation suggests that upregulation of *spn6* and *meu6*, either transcriptionally or post-transcriptionally, could interfere with the expression/stability of *dcr1* and *clr4* transcripts; surprisingly, however, no such fluctuations were seen, with *dcr1* (1.26- and 1.24-fold upregulation) and *clr4* (1.23- and 1.26-fold upregulation) both instead showing little response to *cid16* deletion other than a possible slight upregulation. A more complete understanding of the significance of these observations must await a more complete characterisation of *cid16* targets and activity.

4.9 Discussion

4.10.1 Summary

Though Cid16 was demonstrated in **Chapter 3** to be a second mRNA 3' uridylyltransferase *in vivo* and *in vitro*, almost nothing is known about the biochemical and physiological function of this protein. In this chapter, evidence is presented showing that NTAP-Cid16 is a highly selective PUP under physiological conditions, both for UTP and for the RNA substrate; this selectivity is markedly stronger than that of NTAP-Cid1, which shows considerable variation in catalytic activity depending on the RNA substrate supplied. NTAP-Cid16 also possesses greater processivity than NTAP-Cid1, adding longer oligo(U) tails both *in vitro* and *in vivo*.

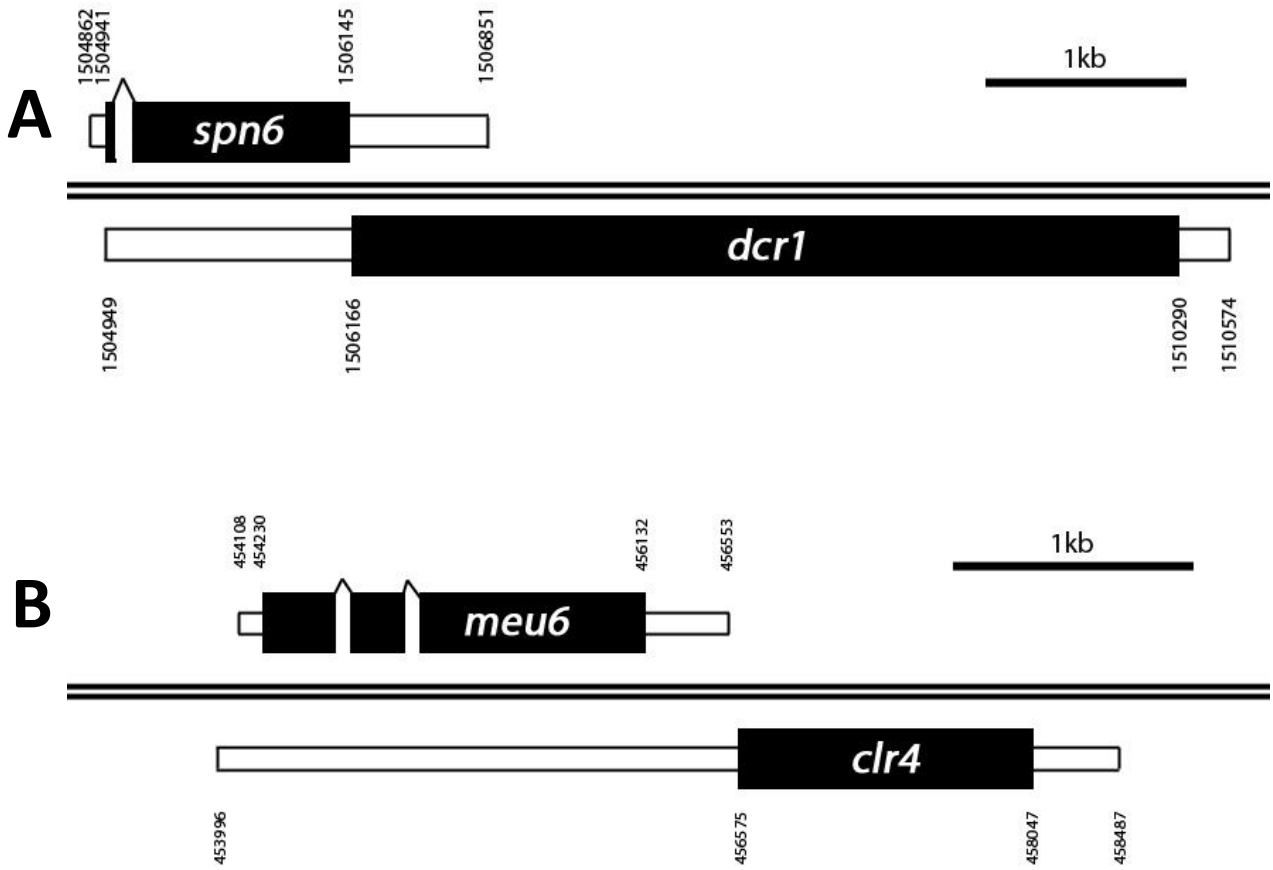


Figure 4.9: *spn6* and *meu6* lie in the 3' UTR of *dcr1* and *clr4*, respectively. Schematics of the (A) *spn6-dcr1* and (B) *meu6-clr4* convergent gene pairs were constructed using the most recent genome annotations on the *S. pombe* genome database PomBase (Wood *et al.* 2012). Absolute positions of 5'/3'-UTR and ORF start/finish positions are indicated in numbers above/below the schematics. *spn6-dcr1* lie on chromosome III and *meu6-clr4* on chromosome II.

The sequence of Cid16 reveals highly conserved 3' catalytic domains and a large, potentially structured N-terminal extension that exhibits no significant domain or sequence conservation with other known proteins. This domain enhances, but is not essential for, NTAP-Cid16 activity *in vitro* and has no effect on UTP selectivity, at least in the assays conducted here. Similarly, no function was found for a proposed Cid16 functional partner, SPAC13F5.07c, in the activity of Cid16 in these assays. Finally, microarray analysis of a *cid16Δ* strain reveals that Cid16 appears to have little effect on the abundance of *S. pombe* mRNAs, suggesting either stability-independent or Cid1-redundant activities; however, the finding that two of the targeted transcripts lie within the 3' UTRs of RNAi factors and exhibit meiotic accumulation provides intriguing insights into the possible functions of Cid16 and the existence of a novel mRNA-regulatory pathway during meiosis.

4.10.2 Cid1 and Cid16 activity on different RNA substrates

One of the more surprising observations to arise from the data presented above in fact concerned NTAP-Cid1 rather than NTAP-Cid16 – the finding that NTAP-Cid1 activity varies remarkably depending on the supplied RNA substrate. Previous studies, using U_{15} as a substrate, observed enhanced PAP, PCP and PGP activity with NTAP-Cid1 purified from *S. pombe*. The studies presented in this chapter further showed a loss of PUP activity on U_{15} and, when a C_{15} substrate was tested, near-complete loss of PUP activity coinciding with a robust enhancement of PCP and particularly PGP activity (Figure 4.2). While the molecular basis of this remarkable variation in activity is currently unknown, it is noted that the enhancement of PAP activity on U_{15} and of PGP activity on C_{15} would generate oligo(A) and oligo(G) tails respectively, and that both of these tails would be expected to interact either intra- or inter-molecularly with the labelled U_{15}/C_{15} substrate. As such, the apparent marked increase in activity using these nucleotides may be partially artefactual and reflect the stabilisation and/or enhancement of oligo(A)/oligo(G) tail synthesis via the formation of dsRNA structures. However, this explanation is not sufficient to explain the increased synthesis of non-

complementary tails on these substrates, and the conclusion must be drawn that the sequence of the RNA substrate is able to modify in some way the nucleotide specificity of NTAP-Cid1.

While these *in vitro* observations are striking, their significance *in vivo* is unclear. RNAs terminating with homopolymeric C_{15} and U_{15} tracts are likely to be extremely rare or absent *in vivo* (though it is noted that RNA Pol III transcripts such as U6 snRNA terminate in short templated oligo(U) sequences; Trippe *et al.* 1998; Trippe *et al.* 2003) and the observed activity of Cid1 on poly(A)$^+$ mRNA (Rissland *et al.* 2007) suggests that, on these native substrates, the PUP activity of Cid1 is likely to predominate. It is nonetheless possible that factors associating with Cid1 *in vivo* may be capable of modulating either its NTP selectivity or its access to RNA substrates in such a way as to induce activities other than polyuridylylation. Interestingly, the *Chlamydomonas reinhardtii* ncPAP Mut68 has been separately reported to uridylylate miRNAs and siRNAs (Ibrahim *et al.* 2010) and to polyadenylate 5' miRNA cleavage products (Ibrahim *et al.* 2006), suggesting that this enzyme may exhibit separate, catalytically distinct activities on different substrates *in vivo*. Similarly, the well-characterised human U6 TUTase (**1.3.3** Maintenance of the U6 snRNA 3' end above; Trippe *et al.* 2003; Trippe *et al.* 2006) has been reported to be able to substitute for canonical PAP and cotranscriptionally polyadenylate mRNAs in response to phosphatidylinositol signalling (Mellman *et al.* 2008; Mellman and Anderson 2009). While both of these reported bifunctionalities remain controversial, the wide variation in NTAP-Cid1 polymerase activity on different substrates leaves open the possibility that Cid1 may possess separate, non-PUP activities on specific RNAs *in vivo*.

Interestingly, NTAP-Cid16 exhibits significantly more stringent UTP and RNA selectivity than NTAP-Cid1 in these same assays; while PAP, PCP and PGP activities are enhanced on C_{15}, significant PUP activity remains and appears in these non-competitive assays to be the most robust activity on this oligoribonucleotide. Furthermore, NTAP-Cid16 exhibits no activity with any NTP on a U_{15} substrate, suggesting that NTAP-Cid16 and/or coimmunoprecipitated proteins are inherently able to distinguish between target RNA sequences (Figure 4.2). The latter observation is surprising given the

long poly(U) tails observed on A_{15} substrates, the synthesis of which requires the activity of NTAP-Cid16 on substrates with long homopolymeric uridylyl tracts at their 3' end; however, it is possible in this situation that the interaction of NTAP-Cid16 with the 5'-A_{15} stretch of the target ribonucleotide, or alternatively with a different A_{15} oligoribonucleotide, is sufficient to confer activity on these otherwise unfavourable substrates.

Given the above observations, it is potentially significant that Cid16, but not Cid1, is proven to oligouridylylate mRNAs without a poly(A) tail *in vivo*, as shown by the existence of exclusively oligo(U) tails on a proportion of *act1* mRNAs derived from *cid1Δ* strains (see Appendix D). While a small number of uridylylated mRNAs without poly(A) tails were observed in previous studies (Rissland *et al.* 2007; Rissland and Norbury 2009), it is important to note that, in these studies, both the *cid1* and the *cid16* loci were intact and therefore a proportion of the observed uridylylation events (expected to be about 20%, given the relative frequency of uridylylation events in *cid1*$^{+}$ and *cid1Δ* strains) are likely to derive from the activity of Cid16 rather than Cid1. Strikingly, of the six nonadenylated, uridylylated sequences retrieved from *cid1*$^{+}$ in the previous studies (Rissland *et al.* 2007; Rissland and Norbury 2009), two show oligo(U) tails of >2 nt in length (U_3 and U_{11}) and are therefore expected to derive from the action of Cid16 rather than Cid1 (see **4.4**). Additionally, overexpression of NTAP-Cid1 in this study resulted in the accumulation of uridylyl tails exclusively on polyadenylylated mRNAs (see Appendix D.12). It is important to note that, without extensive cRACE analysis of *cid1*$^{+}$ *cid16Δ* strains and further characterisation of Cid1 activity it is impossible to dismiss the possibility that Cid1 is able to uridylylate non-polyadenylylated mRNAs. However, a system in which Cid16, with its more stringent UTP selectivity, acts on both polyadenylylated and nonadenylated mRNAs while the less selective Cid1 is restricted to polyadenylylated mRNAs would be intriguing and could provide hints towards different physiological roles for and possible functional interactions between these uridylyltransferases.

4.10.3 The structure of Cid16

The finding that the Cid16 catalytic domains lie at the C terminus of the protein (Figures 4.5, 4.6A) leads to obvious questions regarding the function of the long N-terminal extension of ~900 aa. N-terminal truncation testing shows that, while this extension enhances the activity of the catalytic domain, it is not essential for this activity and has no effect on UTP selectivity in the *in vitro* PUP assays employed here. Instead, it is likely that the N-terminal extension is required for aspects of Cid16 activity that are irrelevant *in vitro,* for example the identification of and interaction with target RNAs or protein cofactors in the complex cellular milieu, or providing regulatory functions that modulate Cid16 activity in response to cellular stimuli. Alternatively, the observation that an overexpressed GFP-Cid16 fusion localises to discrete cytoplasmic dots in an ORFeome localisation screen (Matsuyama *et al.* 2006) suggests the possibility that the N-terminal extension could direct this subcellular localisation of Cid16. It is interesting to note that ZCCHC11 and ZCCHC6 from vertebrates also possess C-terminal catalytic domains, and that their N-terminal extensions contain domains essential for the interaction with cofactors and for efficient catalysis (Thornton *et al.* 2012) and that ZCCHC6 localises to cytoplasmic granules like Cid16 (Andrea Mikulasova, pers. comm.). *C. elegans* CDE-1, another PUP possessing a long N-terminal extension, also localises to cytoplasmic granules via interaction with RDRC components in the *C. elegans* germline (van Wolfswinkel *et al.* 2009). Though the above hypotheses regarding the function of the Cid16 N-terminal extension remain thus far speculative, they provide guidance for the further investigation of Cid16 function and suggest a number of experiments that will facilitate a greater understanding of this protein.

4.10.4 RNAi factors and convergent gene regulation by Cid16?

Comparison of *wt* and *cid16Δ* expression profiles by microarray demonstrated remarkable consistency between these samples and suggested that the vast majority of mRNAs in *S. pombe* are not regulated by Cid16 during vegetative growth in EMM (Figure 4.8 and Tables 4.1, 4.2). However, the observation that two of the three genes upregulated >2-fold upon *cid16* deletion, *spn6* and

meu6, lie in convergent gene pairs with the key RNAi regulatory factors (*dcr1* and *clr4* respectively; Figure 4.9) hints at intriguing potential functions for *cid16*. Gullerova and Proudfoot (2008) have reported that convergent genes (CGs) undergo transient, cell cycle-regulated heterochromatin formation in *S. pombe* and that this heterochromatin formation is responsible for cell cycle-specific regulation of CG expression. Heterochromatin formation is induced during G1 by the formation of dsRNA from overlapping transcription of CGs and consequent recruitment of the RNAi machinery; subsequently, cohesin recruited to the heterochromatically-marked CGs accumulates in the intergenic space between the genes and prevents further overlapping transcription, providing a novel link between the regulation of transcription termination and the heterochromatic machinery (Gullerova and Proudfoot 2008). While approximately 30% of all *S. pombe* genes lie in convergent gene pairs, this proportion rises to approximately 80% when genes encoding components of the RNAi machinery are considered (see Table 4.3; Gullerova *et al.* 2011). Remarkably, the orientation of RNAi factors into CGs is essential for their function *in vivo*, since 'tandemisation' of *dcr1*, *clr4* or *ago1* by downstream insertion of a tandemly-oriented *ura4*$^{+}$ cassette and consequent loss of G1-specific repression of these genes results in defects in heterochromatic silencing and chromosomal segregation (Gullerova *et al.* 2011).

Examination of the expression profiles of the partner genes in RNAi factor CGs reveals that the majority undergo robust up- or downregulation during the meiotic cell cycle (Table 4.3). For example, *spn6* and *meu6* are undetectable in genome-wide mitotic expression profiles (Rustici *et al.* 2004) and expressed at a very low level in RT-PCR analyses (Gullerova *et al.* 2011) but accumulate approximately 80-fold and 300-fold respectively over the course of the meiotic cell cycle (Mata and Bähler 2006). These observations suggest the possibility that, in addition to the mitotic cell cycle regulation observed by Gullerova *et al.* (2011), RNAi factors may undergo regulation over the course of the meiotic cell cycle through the robust upregulation of their convergent gene pairs. The role of RNAi in the establishment of meiotic heterochromatin and/or heterochromatic gene silencing is not well characterised; however, a deletion of *cid12* has been shown to cause both

Table 4.3: The majority of RNAi factor genes lie in convergent gene pairs containing meiotically regulated genes. RNAi factor genes listed in Gullerova *et al*. (2011) are listed alongside the orientation of the gene immediately 3' to them. Genes exhibiting convergent orientation are listed and their expression over the course of the meiotic life cycle provided (Mata and Bähler 2006). Grey shading indicates that the convergent gene lies entirely within the annotated 3'-UTR of the RNAi factor.

RNAi factor	Arrangement	Convergent gene	Meiotic expression
dcr1	convergent	*spn6*	Up
chp1	tandem		
tas3	convergent	*ucp8*	Complex
ago1	convergent	*mmi1*	Complex
arb1	tandem		
arb2	convergent	*crb3*	Up
rdp1	convergent	*ade3*	Down
cid12	convergent	*naa50*	Up
hrr1	convergent	*SPCC1739.04c*	Up
clr4	convergent	*meu6*	Up
rik1	convergent	*SPCC11E10.09c*	Up
raf1	tandem		
raf2	convergent	*SPCC970.06*	Up
pcu4	tandem		
stc1	convergent	*mug30*	Up
sir2	convergent	*zrt1*	Down
set2	convergent	*ino80*	Unchanged

mitotic and meiotic chromosome segregation defects in *S. pombe* (Win *et al.* 2006) while roles for RNAi machinery in meiotic chromosome segregation have been reported in higher organisms including *C. elegans* (Claycomb *et al.* 2009; Gu *et al.* 2009) and, potentially, zebrafish (Houwing *et al.* 2008). Furthermore, the piRNA-mediated silencing of transposable genetic elements has been reported in a number of species (Vagin *et al.* 2006; Houwing *et al.* 2007; Pélisson *et al.* 2007; Klattenhoff and Theurkauf 2008; Wang and Reinke 2008) and is important for the maintenance of genomic integrity during meiosis (Vagin *et al.* 2006; Houwing *et al.* 2007; Wang and Reinke 2008). The activity of PUPs has been shown to regulate meiotic chromosome segregation in *C. elegans*, where CDE-1 uridylylates and promotes the degradation of CSR-1-bound siRNAs (van Wolfswinkel *et al.* 2009).

It is important to note that any hypothesis implicating *cid16* in the regulation of meiotic RNAi CGs is, on the basis of the current evidence, purely speculative. While *spn6* and *meu6* have been shown to reside in RNAi-related CGs, the third gene upregulated >2-fold in *cid16Δ, meu14*, does not (though the upregulation of *meu14* may be a secondary effect of *spn6* upregulation, since these two genes are reported to show positive genetic interaction; Onishi *et al.* 2010). It is also unclear whether Cid16 regulates the expression of other meiotic transcripts in RNAi-related CGs or whether this effect is limited to the CGs containing *dcr1* and *clr4*, and whether the upregulation of *spn6* and *meu6* is sufficient to induce changes in the activity of the RNAi machinery in *cid16Δ* strains. Finally, the mechanism by which Cid16 may recognise and regulate these transcripts is unclear, though the meiotic upregulation of the putative functional partner *SPAC13F5.07c* (Mata and Bähler 2006; Ryan *et al.* 2012) could suggest a meiotic-specific cooperation of these proteins. Experiments to investigate these questions are currently underway and, it is hoped, will provide insights into this intriguing possible role for *cid16*.

Chapter 5
Cid1 and subtelomeric heterochromatin

5.1 Summary

Despite extensive study of Cid1 in *S. pombe*, the full characterisation of the *in vivo* function(s) of this protein has been impeded by the inability to identify Cid1 targets except by candidate-based sequence-specific assays. In this chapter, a novel approach using 4-thiouridine (4-SU) pulse-labelling of nascent transcripts and microarray technologies to generate a transcriptome-wide measure of RNA half-life ($t_{½}$) was used to compare $t_{½}$ for *wt* and *cid1Δ* strains. Surprisingly, no transcripts were identified that showed concurrent expression changes and $t_{½}$ changes, suggesting that deletion of *cid1Δ* does not result in significant changes in RNA $t_{½}$ under the conditions tested here.

Analysis of the transcripts that show transcriptional up-regulation in response to *cid1* deletion revealed an unexpected enrichment of genes in extended subtelomeric regions, suggesting that *cid1* may mediate a genome position-specific regulation mechanism reminiscent of effects seen upon deletion of heterochromatic and/or RNAi factors. To investigate this hypothesis, the *cid1Δ* allele was tested for genetic interaction at a transcript expression level with several mediators of heterochromatin and RNAi pathways. Whilst *cid1Δ* showed effects independent of the RNAi machinery and the histone variant *h2a.z*, an epistatic interaction with a deletion allele of the putative siRNA-degrading exoribonuclease gene *eri1* was observed for a subset of *cid1Δ*-upregulated transcripts, suggesting that *cid1* and *eri1* may lie in a common metabolic pathway. Finally, given early observations of a role for *cid1* in the regulation of mitotic arrest in response to perturbation of DNA replication (Wang *et al.* 2000), the response of the *cid1Δ*-upregulated transcripts *tlh1-4* to DNA

Pol δ/ε inhibition (and consequent S phase arrest) in the presence and absence of *cid1* was tested. Differences in the regulation of *tlh1-4* in response to S phase arrest induction were observed similar to the phenotypic variation of *cid1* deletion in these backgrounds, supporting a possible role for sub-telomeric transcript regulation in the *cid1* S phase arrest phenotype. These preliminary observations provide the first evidence for a possible role for *cid1* in the regulation of subtelomeric heterochromatin.

5.2 *cid1* deletion does not result in RNA half-life changes in a transcriptome-wide assay

Thus far, all mRNA targets of Cid1 uridylylation have been identified by direct sequencing of 5'/3' ends using the cRACE method described previously (**3.2** above). The transcript-specific nature of this method, and the large number of sequences required to confirm Cid1 uridylylation of an mRNA unequivocally, make large-scale target identification by this method unfeasible. Though attempts have been made to develop an analogous RNA circularisation-based deep-sequencing approach independent of transcript sequence in human cells, these attempts have so far been unsuccessful (Dr. Marie-Joëlle Schmidt, unpubl. obs.). Whilst microarray-based expression profiling has identified a set of transcripts whose levels fluctuate in response to *cid1* deletion, the reliance of this method on differences in steady-state transcript levels precludes determination of whether these species are upregulated co- or post-transcriptionally, and thus whether they are dependent on uridylylation-mediated degradation or on secondary effects (Dr. Abigail Stevenson and Dr. Olivia Rissland, unpubl. obs.).

Given the well-characterised role of mRNA uridylylation as a signal for post-transcriptional degradation (Shen and Goodman 2004; Mullen and Marzluff 2008; Rissland and Norbury 2009), deletion of *cid1* is expected to result not only in the upregulation of a target transcript but also an increase in RNA stability, quantifiable via the $t_{½}$. To date, one uridylylation target, *urg1*, has been shown to undergo a change in RNA $t_{½}$ upon deletion of *cid1* (Rissland and Norbury 2009). Given these observations, measurements of changes in RNA $t_{½}$ in response to *cid1* deletion likely represent

a useful surrogate metric for the identification of physiologically relevant Cid1 uridylylation targets. To this end, an attempt was made to identify new *cid1* targets on a whole-genome level using a microarray-based assay for RNA $t_{½}$ changes first described with human fibroblasts (Dölken *et al.* 2008) and subsequently optimised for use in *S. pombe* by Dr. Juan Mata at the Department of Biochemistry, Cambridge University, in whose laboratory this work was conducted (see Figure 2.2).

In this method, two biological replicates of prototrophic *wt* and *cid1Δ* strains were grown in supplemented minimal medium to mid-log phase then briefly pulsed with excess 4-SU which was incorporated into all RNA transcribed during the course of the pulse. RNA was subsequently extracted for each strain and split into separate "total RNA" and "transcribed RNA" pools. Thiolated transcripts in the "transcribed RNA" pool were selectively isolated by biotinylation and streptavidin bead purification. Successful purification was confirmed by Nanodrop RNA quantitation (data not shown). Labelling of "total RNA" and "transcribed RNA" from each strain with different fluorescent Cyanin dyes (Cy3 and Cy5) and competitive hybridisation to a microarray allowed determination of the ratio of nascently transcribed RNA to total RNA and, consequently, calculation of the specific RNA $t_{½}$. Finally, comparison of the derived $t_{½}$ between strains allowed identification of transcripts whose $t_{½}$ was dependent upon *cid1* integrity. A parallel *wt*:*cid1Δ* expression array correlated RNA $t_{½}$ changes with changes in steady-state RNA levels.

Comparison between biological replicates for retrieved RNA $t_{½}$ measurements for *wt* (Figure 5.1A) and *cid1Δ* (Figure 5.1B) strains and for the *wt*:*cid1Δ* expression microarrays (Figure 5.1C) showed a high level of reproducibility, confirming the reliability of the microarray procedure and of the growth of cultures. Parsing the *wt* and *cid1Δ* $t_{½}$ measurements for species exhibiting long and short $t_{½}$ measurements recapitulated with excellent fidelity previously compiled lists of known long- and short-$t_{½}$ transcripts, confirming that the 4-SU labelling of nascent transcripts had been successful and the quantification method was able to retrieve $t_{½}$ values accurately and reliably. As expected, the *cid1* transcript was reduced to background in the *cid1Δ* RNA sample, and no $t_{½}$ could be determined.

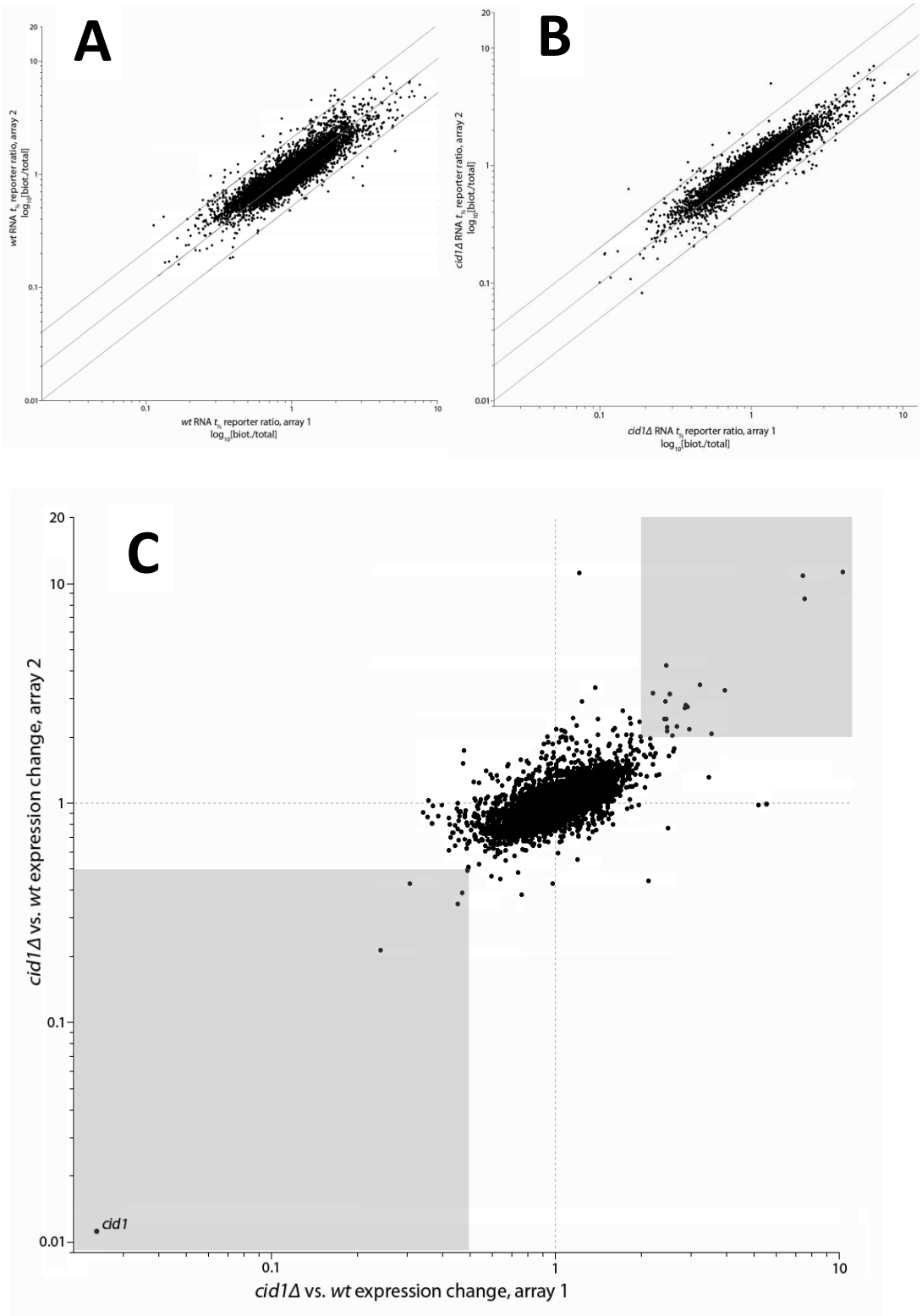


Figure 5.1: The whole-genome RNA $t_{½}$ assay is reliable and reproducible. Two replicates of (A) *wt* total vs. transcribed; (B) *cid1Δ* total vs. transcribed; and (C) *wt:cid1Δ* expression arrays were plotted against each other. Lines in (A) and (B) represent (from top to bottom) 2-fold upregulation of replicate 2 over replicate 1, equality between replicates and 2-fold downregulation of replicate 2 over replicate 1. The shaded areas in (C) contain transcripts exhibiting reproducible >2-fold up- or down-regulation in response to the deletion of *cid1.*

Table 5.1: List of genes with reproducible >2-fold upregulation in a *cid1Δ* strain

Systematic Name	Common Name	Expression Change, *cid1Δ* vs. *wt*		Mean Expression Change	Comments
		Run 1	Run 2		
SPBCPT2R1.08c*	*tlh2*	10.27	11.27	10.77	identical sequence to *tlh1*
SPAC1039.09	*isp5*	7.40	10.86	9.13	induced in sporulation
SPAC212.11*	*tlh1*	7.50	8.55	8.02	identical sequence to *tlh2*
SPCC285.05		3.94	3.27	3.61	
SPBC29B5.02c	*isp4*	2.44	4.25	3.35	induced in sporulation
SPAC1039.09.RC*	*isp5.RC*	3.21	3.45	3.33	antisense to *isp5*
SPAC513.03	*mfm2*	2.87	2.78	2.83	
SPAC11D3.18c		2.91	2.74	2.83	
SPAC27D7.03c	*mei2*	2.51	3.14	2.82	key regulator of meiosis
SPBPB2B2.01		3.53	2.08	2.80	
SPAC922.01	*mmf2; hpf1*	2.84	2.72	2.78	
SPNCRNA.130	*omt3*	2.20	3.17	2.68	ncRNA, adjacent to *mei2*
SPAC869.10c	*put4*	2.43	2.89	2.66	
SPAC186.03		2.95	2.17	2.56	
SPAC11D3.17		2.66	2.23	2.45	
SPAC5H10.01		2.44	2.41	2.42	
SPAC513.04		2.41	2.41	2.41	sequence orphan
SPAC1039.08		2.46	2.21	2.33	
SPBPB21E7.09		2.47	2.14	2.31	adjacent to *SPBPB10D8.01*
SPCC70.12c	*ecl1*	2.58	2.02	2.30	

* No biotin-enrichment data retrieved

Table 5.2: List of genes with reproducible >2-fold down-regulation in a *cid1Δ* strain

Systematic Name	Common Name	Expression Change, *cid1Δ* vs. *wt*		Mean Expression Change	Comments
		Run 1	Run 2		
SPBPB10D8.01		0.49	0.49	0.49	subtelomeric
SPBPB10D8.02c		0.47	0.39	0.43	subtelomeric
SPNCRNA.31	*prl31*	0.45	0.35	0.40	intergenic ncRNA
SPBPB24D3.07c		0.30	0.43	0.37	
SPNCRNA.155		0.24	0.21	0.23	intergenic ncRNA
SPAC19D5.03*	*cid1*	0.02	0.01	0.02	deleted gene

*No biotin-enrichment data retrieved

Table 5.3: List of genes with reproducible >1.4-fold change in RNA $t_{½}$ in a *cid1Δ* strain. Green indicates destabilisation upon *cid1* deletion; red stabilisation

Systematic Name	Common Name	Biotin-RNA enrichment, *cid1Δ* vs. *wt*		Mean Biotin-RNA Enrichment	Expression Change, *cid1Δ* vs. *wt*		Mean Expression Change	Comments
		Run 1	Run 2		Run 1	Run 2		
SPAC212.11.RC	*tlh1.RC*	1.66	1.60	1.63	1.72	1.29	1.51	antisense transcript of *tlh1*
SPATRNAPRO.03		1.49	1.75	1.62		0.85	0.85	tRNA
SPBC337.08c	*ubi4*	1.58	1.43	1.51	0.98	0.91	0.94	
SPNCRNA.199		0.66	0.70	0.68	1.01	1.08	1.05	intergenic ncRNA
SPNCRNA.491		0.65	0.66	0.66	1.79	1.40	1.59	intergenic ncRNA
SPAC212.01c		0.71	0.56	0.64	1.57	0.90	1.24	subtelomeric
SPAC6F6.03c.RC (intron)		0.67	0.60	0.63		1.04	1.04	
SPMIT.09	*atp8*	0.58	0.65	0.62	1.11	1.77	1.44	mitochondrially encoded
SPAC1952.14c	*mrpl25*	0.61	0.58	0.59	1.08	1.35	1.21	
SPATRNAILE.02		0.66	0.24	0.46		1.34	1.34	tRNA
SPAC15A10.07.RC		1.74	0.50	1.12	0.87	1.20	1.04	antisense transcript
SPCC663.07c		1.53	0.71	1.12	0.49	0.67	0.58	pseudogene
SPAC1D4.12.RC	*rad15.RC*	1.46	0.60	1.03	0.72	0.88	0.80	antisense to *rad15*
SPAC26A3.02.RC	*myh1.RC*	0.60	1.66	1.13	1.11	0.88	0.99	antisense to *myh1*

In order to identify transcripts with reproducible RNA $t_{½}$ changes in response to *cid1* deletion, separate lists were compiled of transcripts exhibiting reproducible >2-fold upregulation (Table 5.1) and >2-fold downregulation (Table 5.2) in the *wt*:*cid1Δ* expression array, and of transcripts exhibiting reproducible >1.4-fold changes in $t_{½}$ (Table 5.3). Surprisingly, derived $t_{½}$ values were extremely consistent between *wt* and *cid1Δ* samples, with only ten transcripts showing significant and consistent RNA $t_{½}$ changes in response to *cid1* deletion (three with decreased $t_{½}$, seven with increased) and a further four showing significant but contradictory $t_{½}$ changes (Table 5.3). These species predominantly contained non-coding transcripts (9/14) including antisense transcripts, tRNAs and ncRNAs, many of which are expressed at extremely low levels and consequently may represent false-positive readings. Consistent with this interpretation, none of the transcripts showing significant $t_{½}$ changes also showed >2-fold up- or down-regulation in the *wt*:*cid1Δ* expression arrays (Table 5.3). In a complementary fashion, none of the transcripts showing >2-fold up- or down-regulation in the expression arrays (Tanle 5.1 and Table 5.2, respectively) showed significant changes in RNA $t_{½}$ upon *cid1* deletion, suggesting that these expression differences derive from transcriptional rather than post-transcriptional regulation (though it should be noted that biotin-enrichment data could not be retrieved for two of the most-upregulated transcripts, *tlh1* and *tlh2*). It is thus concluded that, at least under the conditions assayed here, no RNAs can be identified that exhibit significant RNA $t_{½}$ changes and consequent accumulation/depletion upon *cid1Δ* deletion.

5.3 Cid1 does not regulate the *chk1* checkpoint kinase transcript during exponential growth

During analysis of the expression microarray data gathered above, it was noticed that the mRNA encoding Chk1, a key checkpoint kinase, exhibited down-regulation in both replicates tested, but had been excluded from the above analyses due to the down-regulation only crossing the >2-fold threshold in one of two replicates (2.04-fold downregulated in array 1, 1.27-fold in array 2). Chk1 is

phosphorylated by Rad3 upon DNA damage or replication inhibition and mediates downstream inhibition of the cell cycle through inhibitory phosphorylation of the mitotic cyclin-dependent kinase Cdc2 (Tapia-Alveal *et al.* 2009). Chk1 has also been proposed as a downstream component of the Cid1 regulatory pathway (Wang *et al.* 2000). Importantly, regulation of *chk1* transcript stability would be a simple and powerful method by which Cid1 could control S phase arrest and aberrant mitotic progression (Wang *et al.* 2000). However, upon confirmatory quantitation of *chk1* in *wt* and *cid1Δ* strains by poly(A)$^+$ cDNA synthesis and quantitative PCR (qPCR), no significant differences in *chk1* transcript levels were found between strains (Figure 5.2). Parallel quantitation of the *tlh1-4* transcripts (discussed below), the top hit(s) in the microarray, confirmed their highly significant upregulation ($p < 0.005$; Figure 5.2), demonstrating that *cid1Δ*-dependent transcript changes could be recapitulated by qPCR. It is thus concluded that *chk1* does not represent a *cid1* target under vegetative growth conditions and that *cid1* must achieve its checkpoint-regulatory role by an alternative mechanism.

5.4 Genes upregulated in *cid1Δ* cluster to extended subtelomeric regions

The presence of numerous transcripts exhibiting significant up- or down-regulation in the expression arrays without also exhibiting RNA $t_{½}$ changes was surprising, since it suggests that a population of transcripts in *cid1Δ* undergo transcriptional modulation. Observed microarray expression changes were replicated by qPCR for five reproducibly upregulated transcripts (*mfm2, isp5, mei2, tlh1-4* and *cox1*) and one reproducibly downregulated gene (*SPBPB10D8.02c*), whilst no transcript accumulation/depletion could be found for four transcripts showing up- (*ptr2*) or down-regulation (*chk1, ccq1, spt3*) in only one replicate, confirming the veracity of the microarray observations (Figure 5.3). GO enrichment analysis of the reproducibly up-regulated transcripts (Table 5.1; the small number of transcripts showing reproducible downregulation precluded meaningful GO analysis) using the g:Profiler programme

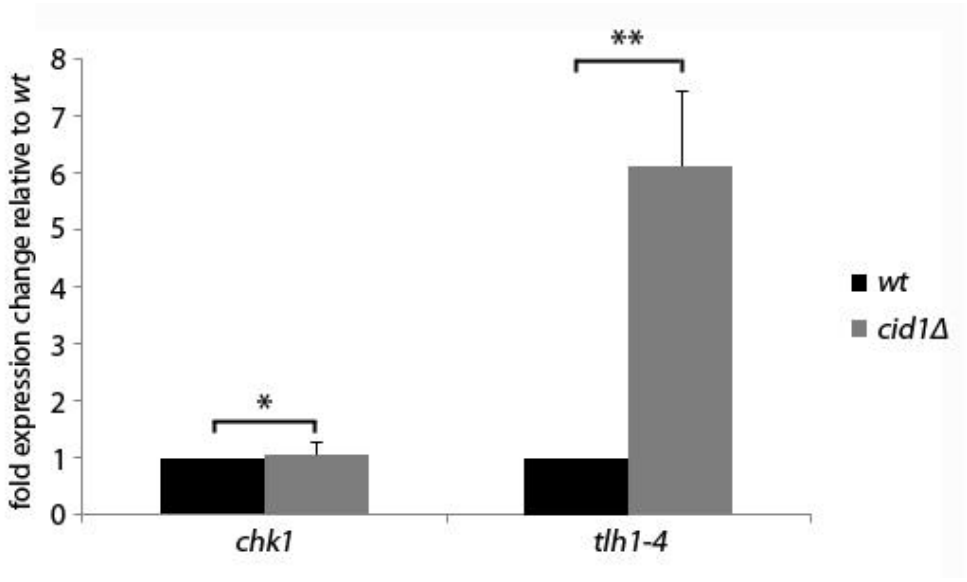


Figure 5.2: *cid1* does not regulate the expression of *chk1*. The abundance of *chk1* and *tlh1-4* transcripts in poly(A)$^+$ cDNA pools from *wt* and *cid1Δ* strains were compared by qPCR. Data represent the average of four experiments ± 1 s.d.. * = $p > 0.5$, ** = $p < 0.005$. *p*-values calculated using Student's *t*-test.

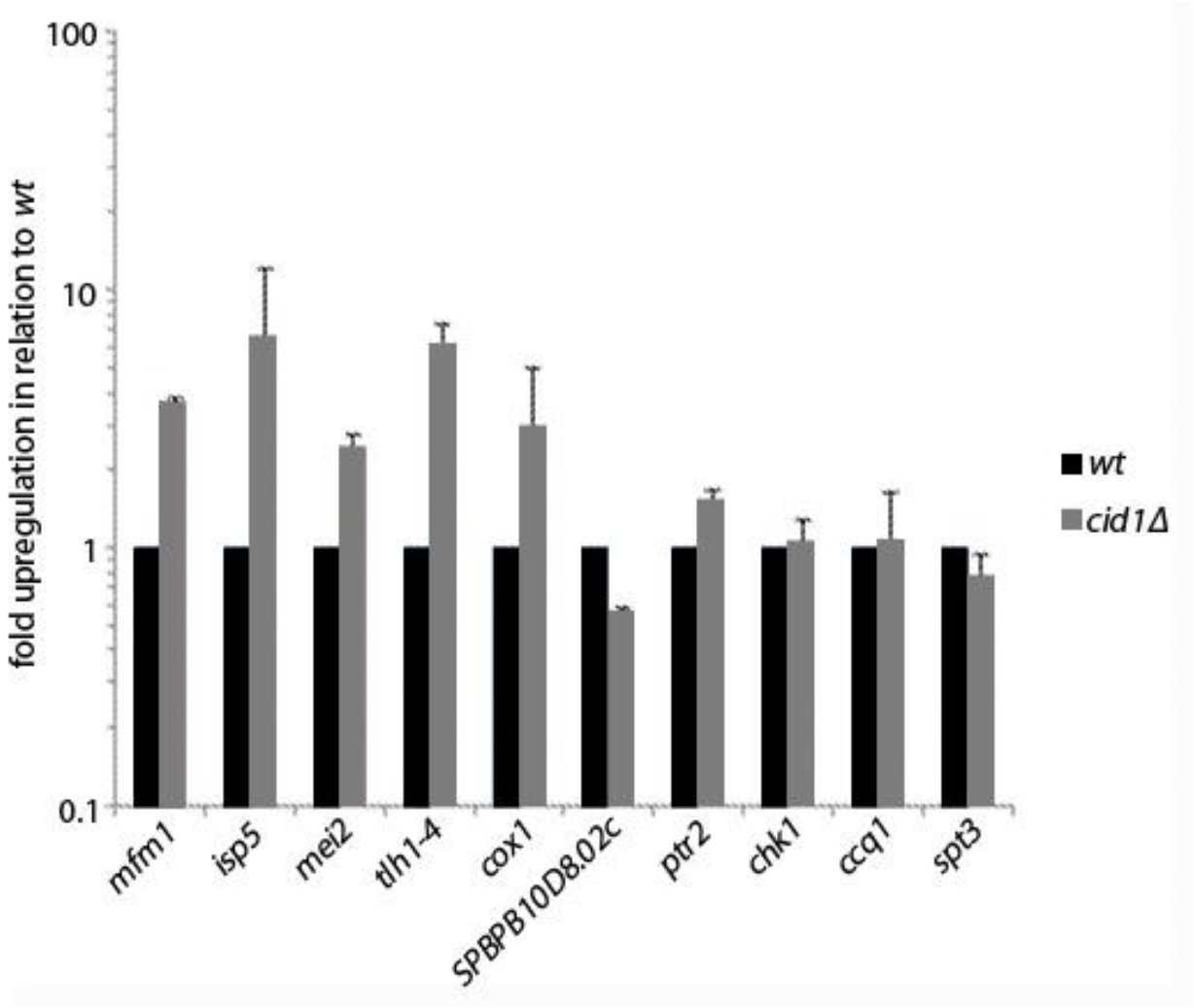


Figure 5.3: Expression changes observed by microarray can be recapitulated using qPCR. Abundance of transcripts of the indicated genes were quantified in poly(A)$^+$ cDNA pools from *wt* and *cid1Δ* strains. Data represent means of between two and five experiments ± 1 s.d.

(Reimand *et al.* 2007, 2011) yielded one significantly enriched GO term, "telomere maintenance via recombination" (p = 0.00138); however, it is noted that this GO annotation applies only to the *tlh1* and *tlh2* transcripts. Expansion of the analysis to the larger lists of transcripts showing >2-fold up- or down-regulation in one array (see Appendix F) yielded significant GO enrichments only for "unannotated" transcripts (p = 1.53×10^{-7}, p = 3.29×10^{-5} respectively). The lack of significant GO term enrichment suggests that the transcripts upregulated upon *cid1* deletion do not belong to a common functional or mechanistic group, and as such the means by which they are upregulated remains unclear.

A striking observation was made, however, when the locations of the up-regulated genes within the *S. pombe* genome were determined. Transcripts upregulated upon *cid1* deletion showed remarkable clustering near to the ends of chromosomes I and II, with 13/20 identified transcripts lying within 150 kb of a telomere (Figure 5.4). Of the seven not found in extended subtelomeric regions, four (*mfm2/SPAC513.04* and *mei2/omt3*) existed as two divergent gene pairs in the body of chromosome I, each containing an important meiotic/conjugation factor (*mfm2*, a precursor of the M-factor pheromone (Davey 1992), and *mei2*, a key initiator of meiotic DNA synthesis and meiosis I (Watanabe and Yamamoto 1994; Yamamoto 1996) respectively) whilst a further one (*isp4*) was located close to the chromosome II centromere. Though the list of genes down-regulated upon *cid1* deletion is small, it is noted that two of the five genes in this list, *SPBPB10D8.01* and *SPBPB10D8.02c*, are adjacent to each other and lie within 150 kb of the left telomere of chromosome II. These results suggest that, instead of targeting a functionally or sequentially defined set of transcripts, *cid1* regulates the transcription of multiple genes in extended subtelomeric regions on chromosomes I and II.

A search for other microarray datasets exhibiting a similar telomeric position-dependent transcript regulation retrieved arrays for both point and deletion mutants of *clr1*, *clr3* and *clr4*, three factors central to the formation of heterochromatin (Hansen *et al.* 2005). Furthermore, the list of transcripts

showing >2-fold regulation in *cid1Δ* overlaps significantly with similar gene lists derived from point and deletion mutants of these genes (Figure 5.5). Noticeably, the *clr* family microarrays, as in the *cid1Δ* microarrays presented here, showed extremely strong upregulation of the *tlh1-4* transcripts, representing the most up-regulated gene in all of these arrays. Recent work has also described subtelomeric enrichment of transcripts upregulated upon deletion of *cid14* which has recently been implicated in the nucleation of RNAi-independent heterochromatin (Reyes-Turcu *et al.* 2011; Zhang *et al.* 2011).

tlh1 and *tlh2* (and two other genes, *tlh3* and *tlh4*, not currently present in the *S. pombe* genome assembly; Wood *et al.* 2002) are located as the final genes on the Watson and Crick strands of chromosomes I and II and are orientated so as to face out into telomeric sequence (Mandell *et al.* 2005). These genes produce large (7.6 kb) transcripts of identical sequence and are likely maintained by inter-recombination. Given that this identical sequence results in cross-reactivity of the *tlh* genes in microarrays and qPCR and precludes separate characterisation of the genes, these transcripts are hereafter collectively referred to as *tlh1-4*. These genes encode a large RecQ-family helicase (Mandell *et al.* 2005); however, Tlh1-4 proteins have thus far eluded detection (Dr. Li Phing Liew, unpubl. obs.). The *tlh1-4* reading frames also contain repeat sequences highly similar to (and presumably the ancestors of) the *dg* and *dh* repeats required for centromeric heterochromatin formation, suggesting that these genes might also function as sites for subtelomeric heterochromatin nucleation (Hansen *et al.* 2006). On the basis of the various results described above, an hypothesis was drawn in which the regulation of *tlh1-4* by *cid1*, or the regulation of some other aspect of telomeric RNA metabolism, could modulate the transcription of subtelomeric mRNAs by affecting the formation of subtelomeric heterochromatin.

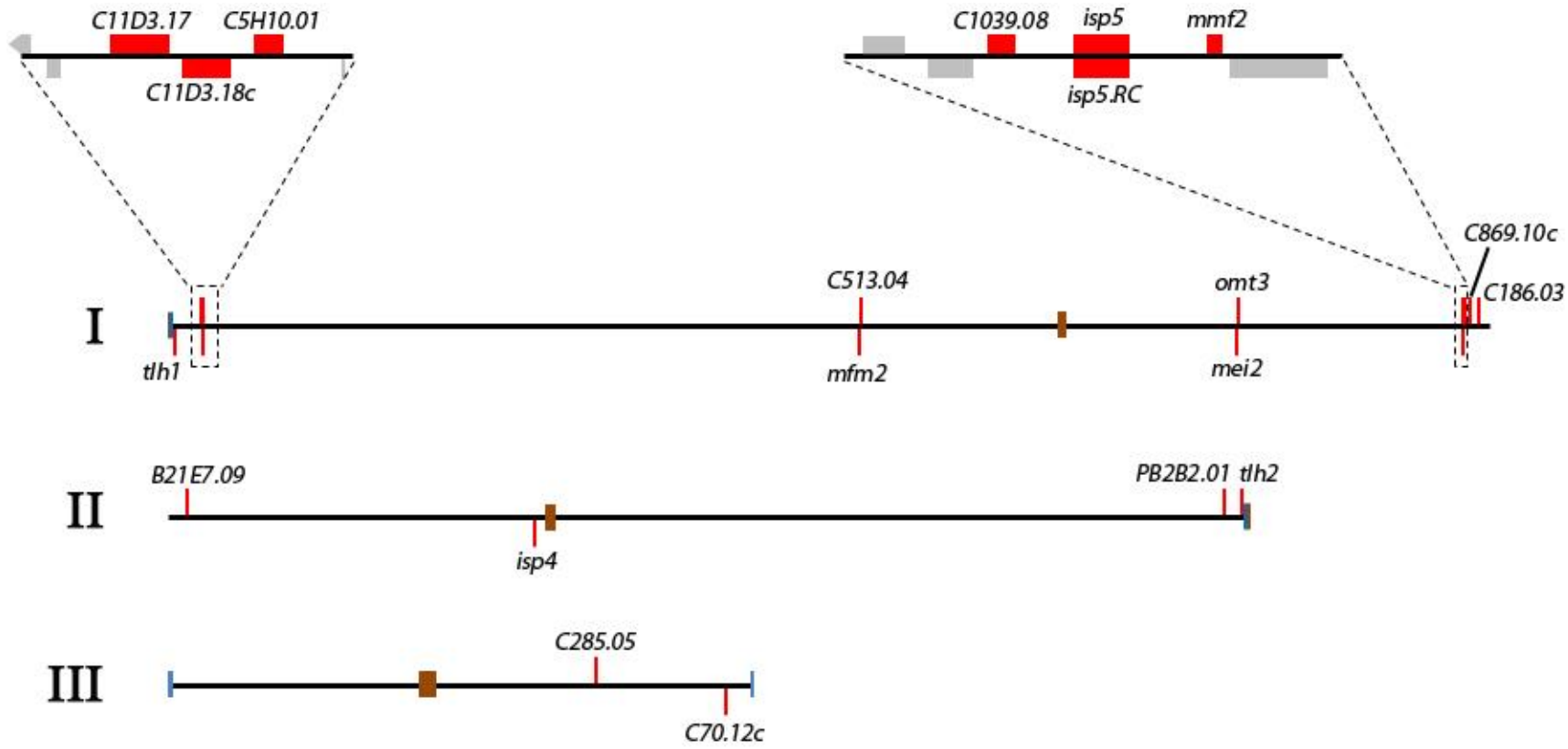


Figure 5.4: Genomic location of genes exhibiting reproducible >2-fold upregulation in a *cid1Δ* strain. Genomic locations are drawn to scale on the three chromosomes of *S. pombe* using genome annotations retrieved from the *S. pombe* genome database PomBase (Wood *et al.* 2012).

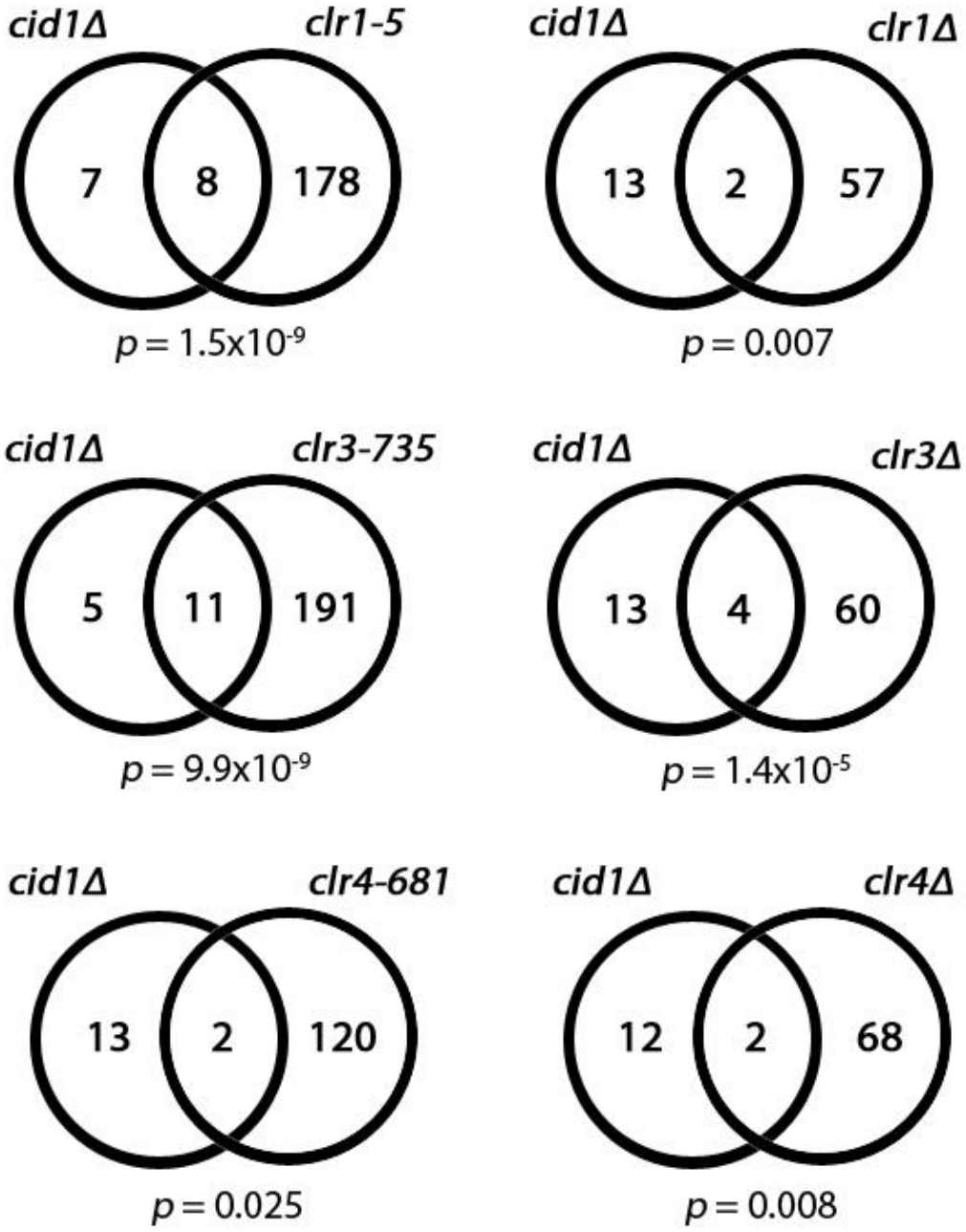


Figure 5.5: *cid1Δ*-regulated transcript lists overlap with those of *clr1, clr3* and *clr4* deletions and point mutants. Lists of regulated genes in *clr1, clr3* and *clr4* point and deletion mutants were retrieved from Hansen *et al*. (2005) and the extent of overlap with the *cid1* regulated gene lists derived in this study. *p*-values were calculated using Fisher's Exact Test.

5.5 Discussion

5.5.1 Summary

Transcriptome-wide attempts to identify RNAs exhibiting $t_{½}$ changes in response to *cid1* deletion led to the surprising finding that no transcripts exhibited correlated RNA $t_{½}$ and total abundance changes, suggesting that, at least under the conditions tested here, *cid1* is dispensable for the stability of RNAs *in vivo*. However, analysis of the expression profiles from *wt* and *cid1Δ* cells revealed clustering of genes encoding transcripts upregulated upon *cid1* deletion to extended subtelomeric regions; in particular, *cid1Δ* caused the robust upregulation of the conserved *tlh1-4* transcripts, which represent the terminal genes of chromosomes I and II and sites of subtelomeric heterochromatin nucleation (Hansen *et al.* 2006). The similarity of the *cid1Δ* expression profile to those of several heterochromatin factor deletions suggests that *cid1* could potentially be mediating this effect through the regulation of subtelomeric heterochromatin. Some possible mechanisms by which *cid1* could regulate the formation and/or integrity of subtelomeric heterochromatin are discussed below.

5.5.2 cid1 *is dispensable for the regulation of RNA* $t_{½}$ in vivo

The finding that deletion of *cid1* does not affect the $t_{½}$ of transcripts during exponential growth is surprising given the evidence that uridylylation of mRNAs induces their degradation via decapping (Mullen and Marzluff 2008; Rissland and Norbury 2009; Morozov *et al.* 2010). While it is possible that Cid1-mediated uridylylation of mRNAs has other, non-degradative functions *in vivo*, a more likely explanation is that targets of *cid1* have not been represented in this assay. There are several possible reasons why *cid1*-dependent transcripts may have been missed. Firstly, transcripts whose half-lives are significantly shorter than the duration of the 4-SU pulse (15' in these experiments) will undergo near-complete labelling over the course of the pulse, resulting in identical signals for the "total RNA" and "transcribed RNA" pools (see **2.6.1**); as a result, modest changes in the half-lives of these species will not

be detectable with this assay. In line with this hypothesis, it is noted that RNA $t_{½}$ measurements were not obtainable for the *tlh1-4* transcripts in this assay (Tables 5.1, 5.3). Secondly, any potential restriction of biologically relevant Cid1 activity to a particular cellular state or in response to particular stimuli – for example, activity upon induction of S phase arrest – would be missed by this study, which exclusively quantifies transcripts during exponential growth in minimal medium. Consistent with such a hypothesis, the earlier observation of *cid1*-dependent destabilisation of the *urg1* transcript (Rissland and Norbury 2009) was made following uracil withdrawal, rather than under the constant uracil supplementation used here. Unfortunately, attempts to measure RNA $t_{½}$ during S phase arrest are likely to be unsuccessful due to technical limitations of the 4-SU labelling protocol. Finally, a number of biologically relevant RNA species are not represented in the microarrays used in this assay or are not defined in the genome annotations used here (Wood *et al.* 2002). Of particular relevance to this study, these species will include small RNAs and mRNA degradation products, amongst others. Which of these explanations, if any, is responsible for the lack of observed $t_{½}$ variation in *cid1Δ* cells is unclear; however, these observations suggest that the physiologically relevant activities of *cid1* likely extend beyond its uridylylation activity on the mRNAs thus far identified.

5.5.3 cid1 *as a possible regulator of subtelomeric heterochromatin?*

While the RNA $t_{½}$ assays presented above did not reveal any transcripts regulated by *cid1*, the expression arrays conducted in parallel revealed a number of upregulated transcripts exhibiting striking proximity to the telomeres of chromosomes I and II, with 13/20 transcripts lying within 150 kb of these structures (Figure 5.4). These observations suggest that *cid1* is able to regulate the transcriptional activity of these genes in a genome position-specific fashion, a hitherto unprecedented observation.

Heterochromatin formation at the telomeres of chromosomes I and II is initiated at the *dg/dh*-like repeats contained within the subtelomeric *tlh1-4* genes; chromosome III, by contrast, terminates in

tandem rDNA repeats which initiate heterochromatin by a different method (Mandell *et al.* 2005; Hansen *et al.* 2006). *tlh1-4*-dependent heterochromatin formation is controlled redundantly by both the RNAi machinery (Noma *et al.* 2004; Sadaie *et al.* 2004; Hansen *et al.* 2006) and by RNAi-independent recruitment of Swi6 by the telomere-binding protein Taz1 (Kanoh *et al.* 2005); once heterochromatin is established at the *tlh1-4* loci, it is likely spread into the subtelomeres via the secondary synthesis of siRNAs by RDRC and the spreading of RITS/CLRC along the chromosome (see section **1.5;** Goto and Nakayama 2011). Deletion of the *tlh1-4* repeat loci compromises heterochromatin formation by both these pathways, suggesting that these repeats are essential to heterochromatin nucleation at the subtelomeres (Noma *et al.* 2004; Kanoh *et al.* 2005). These observations, combined with the similarity of the *cid1Δ* expression profiles to those of deletions/mutations in known heterochromatin factors *clr1, clr3* and *clr4* (Figure 5.5) suggest the possibility that *cid1* may regulate subtelomeric transcription by controlling the initiation of heterochromatin at the *tlh1-4* loci and/or its subsequent propagation along the DNA fibre.

Though these observations are intriguing, no mechanistic information is so far available on how these effects might be achieved; indeed, the finding that the cytoplasmic, RNA metabolic enzyme Cid1 is able to regulate subtelomeric transcript levels was highly surprising. The central role of the RNAi machinery and of RNA metabolism in the establishment and propagation of heterochromatin in *S. pombe* (see **1.5** above; Goto and Nakayama, 2011) nevertheless suggest several hypotheses regarding how this role may be achieved. The most parsimonious explanation for the effect of *cid1Δ* on subtelomeric heterochromatin formation is an indirect one in which Cid1 regulates the stability, translation or transcription of the mRNA of a heterochromatin establishment factor such as *clr4*. However, it is noted here that the expression arrays conducted with the *cid1Δ* strain showed no evidence of expression changes for any known heterochromatin regulatory factors (data not shown) and thus that, while it is

possible that a regulatory effect exists which is not revealed in the assays conducted herein, on the basis of current evidence no indirect effects can be elucidated.

The central role of the *tlh1-4* transcripts in the establishment of subtelomeric heterochromatin by both Swi6- and RNAi-dependent pathways (Noma *et al*., 2004; Sadaie *et al*., 2004; Kanoh *et al*., 2005; Hansen *et al*., 2006) suggests that degradative activity of Cid1 on *tlh1-4* mRNA, or on a metabolite of this transcript, could alter the degradation or the processing of this transcript into siRNA and consequently control the initiation of heterochromatin at the *dg/dh*-like repeats contained in the *tlh1-4* genes. Interestingly, the cytoplasmic dsRNA exoribonuclease *eri1*, initially identified as an siRNA exoribonuclease (Kennedy *et al*., 2004; Yang *et al.*, 2006; Iida *et al*., 2006) has recently been implicated in the maturation and production of siRNAs and consequently in the establishment of certain kinds of RNAi (Duchaine *et al*., 2006; Lee *et al*., 2006; Pavelec *et al*., 2009); Duchaine *et al*. (2006) further propose that this activity may be mediated via acting on repeat-containing transcripts prior to siRNA formation.

Finally, it is noted that uridylylation has been implicated in the metabolism of siRNAs in a range of other species, and that these metabolic activities have been shown to regulate heterochromatin formation and chromosomal segregation during both mitosis and meiosis (Li *et al*., 2005; Claycomb *et al*., 2009; van Wolfswinkel *et al*., 2009; Ibrahim *et al*., 2010; Ameres *et al*., 2010, 2011). An intriguing hypothesis is that *cid1Δ* may mediate its subtelomeric effects through the uridylylation of siRNAs, either before or after loading onto the Ago1-RISC complex. Such activity would result in a loss of silencing of both the *tlh1-4* transcript and the subsequent propagation of heterochromatin into the subtelomeres. However, it is important to note that such an activity may also be expected to compromise heterochromatin formation non-specifically in the cell and consequently result in pleiotropic effects on pericentromeric and mating-type locus transcripts as well.

Though these hypotheses remain purely speculative at this point, they provide important guidance for the future testing of *cid1* activity at the subtelomeres. Detailed characterisation of the chromatin status of *tlh1-4* and other subtelomeric transcripts would be expected to provide insight into the mechanism(s) by which *cid1Δ* regulates their expression. Characterisation of the RNAs whose uridylylation by Cid1 mediate the repression of subtelomeric transcription is also essential; approaches such as RNA immunoprecipitation (RIP) and deep-sequencing of sRNAs in *wt* and *cid1Δ* strains are expected to provide important insights on these fronts. Finally, detailed genetic and molecular analysis of *cid1Δ* in combination with other heterochromatin factor mutants should provide guidance towards a more complete understanding of this exciting and hitherto unappreciated mechanism by which *cid1* may regulate RNA accumulation and the propagation of subtelomeric heterochromatin.

Chapter 6
Concluding Remarks

The results presented in this thesis have significantly expanded the understanding of the extent and potential functions of RNA 3'-terminal oligouridylylation in *S. pombe*. It is now known that at least two separate ncPAPs in *S. pombe* are PUPs *in vitro* and that both are capable of uridylylating *act1* mRNA *in vivo*. These observations align *S. pombe* with other organisms encoding multiple PUPs (Table 1.1), in particular with humans where two cytoplasmic PUPs, ZCCHC6 and ZCCHC11, are known to have overlapping but distinct activities on a range of different RNA (Heo *et al.* 2009; Jones *et al.* 2009; Schmidt *et al.* 2011). The definitive identification of *cid16* as the second mRNA uridylyltransferase should be of considerable use to the investigation of both *cid1* and *cid16*, as it will now be possible to investigate the activities and functions of these two uridylyltransferases independently. One question of particular interest will be whether *cid1* and *cid16* act redundantly on some or all targets, and if so whether additional functions for these genes may be uncovered by their simultaneous deletion. The current list of *cid1* and *cid16* targets is likely far from complete; however, it is noted here that both Cid1 and Cid16 have been shown to add oligo(U) tails to *act1* mRNA, suggesting these enzymes are capable of targeting at least some overlapping transcripts *in vivo*.

While the identification of *cid16* as an mRNA uridylyltransferase is a major step forward in the characterisation of this gene, many questions regarding its targets, regulatory activities and functions *in vivo* remain. It is hoped that the analyses presented in **Chapter 4** will provide guidance for future investigations of this gene. Of particular interest will be the extension of the N-terminal truncation studies to investigate whether this large and apparently highly-structured region plays roles *in vivo* that

have not been revealed by the *in vitro* analysis conducted here; further experiments regarding the N-terminal extension, such as mass spectrometric and/or deep-sequencing identification of protein and/or RNA species that interact with this region should further assist in this characterisation.

The observation that two transcripts upregulated upon deletion of *cid16*, namely *spn6* and *meu6*, are meiotically regulated and are present in convergent gene pairs with key regulators of the RNAi and heterochromatin machinery is both surprising and intriguing. These transcripts are known to be biologically important for the correct regulation of RNAi factor homeostasis during mitosis (Gullerova *et al.* 2011); their robust meiotic regulation, as well as that of many other transcripts found in convergent orientation with RNAi and/or heterochromatin factors (Table 4.3) suggests the possibility that *cid16* may play a role in a hitherto unappreciated pathway of meiotic RNAi regulation. Considerable further investigation will be required to establish whether this pathway is active and biologically relevant, and to dissect the mechanism, if any, through which it may be regulated by *cid16*; however, the RNAi machinery is known to be important for meiotic chromosome segregation in *S. pombe* (Win *et al.* 2006) and PUPs have been implicated in the regulation of meiotic chromatin via RNAi pathways in *C. elegans* (van Wolfswinkel *et al.* 2009), suggesting that such a pathway may have precedent elsewhere.

Several recent papers have demonstrated that the *cid14*-mediated nuclear RNA quality control pathway plays important roles in the regulation of centromeric heterochromatin in *S. pombe*, and is capable of initiating heterochromatin formation independently of the RNAi machinery (Reyes-Turcu *et al.* 2011; Zhang *et al.* 2011). Results presented in **Chapter 5** provide the striking observation that deletion of *cid1* results in the upregulation of numerous subtelomeric transcripts, providing the first evidence for a possible role of Cid1, and indeed for uridylylation generally, in the regulation of telomeric chromatin dynamics in *S. pombe*. The mechanism(s) by which *cid1* regulates the expression of these subtelomeric transcripts, and indeed even the broad pathway in which it acts to achieve this regulation, is currently

far from clear. However, these observations open the way for a wide range of future experiments that can begin to answer these questions and, it is hoped, provide novel insights into thus far unappreciated mechanisms of *cid1* activity. The possible activity of *cid1* in the regulation of both mRNA and heterochromatin-related targets is reminiscent of the multiple activities of uridylyltransferases in other species including Mut68 in *Chlamydomonas* (Ibrahim *et al.* 2006; Ibrahim *et al.* 2010) and ZCCHC11 in humans (Heo *et al.* 2009; Jones *et al.* 2009; Schmidt *et al.* 2011) and suggests that the functions of RNA uridylylation in *S. pombe* may be more wide-ranging than has previously been appreciated.

Chapter 7
References

Adler, B. K., Harris, M. E., Bertrand, K. I. and Hajduk, S. L. (1991) Modification of *Trypanosoma brucei* mitochondrial rRNA by posttranscriptional 3' polyuridine tail formation. *Molecular and Cellular Biology* **11**(12): 5878-5884.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

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

Ambros, V. and Horvitz, H. R. (1984) Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science* **226**(4673): 409-416.

Ameres, S. L., Horwich, M. D., Hung, J.-H., Xu, J., Ghildiyal, M., Weng, Z. and Zamore, P. D. (2010) Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**(5985): 1534-1539.

Ameres, S. L., Hung, J.-H., Xu, J., Weng, Z. and Zamore, P. D. (2011) Target RNA-directed tailing and trimming purifies the sorting of endo-siRNAs between the two *Drosophila* Argonaute proteins. *RNA* **17**(1): 54-63.

Andrade, J. M., Pobre, V., Silva, I. J., Domingues, S. and Arraiano, C. M. (2009) The role of 3'-5' exoribonucleases in RNA degradation. *Prog Mol Biol Transl Sci* **85**: 187-229.

Andrews, N. C. and Baltimore, D. (1986) Purification of a terminal uridylyltransferase that acts as host factor in the *in vitro* poliovirus replicase reaction. *PNAS* **83**(2): 221-225.

Anindya, R., Chittori, S. and Savithri, H. S. (2005) Tyrosine 66 of Pepper vein banding virus genome-linked protein is uridylylated by RNA-dependent RNA polymerase. *Virology* **336**(2): 154-162.

Ansel, K. M., Pastor, W. A., Rath, N., Lapan, A. D., Glasmacher, E., Wolf, C., Smith, L. C., Papadopoulou, N., Lamperti, E. D., Tahiliani, M., Ellwart, J. W., Shi, Y., Kremmer, E., Rao, A. and Heissmeyer, V. (2008) Mouse Eri1 interacts with the ribosome and catalyzes 5.8S rRNA processing. *Nature Structural and Molecular Biology* **15**(5): 523-530.

Aphasizhev, R. (2005) RNA uridylyltransferases. *Cellular and Molecular Life Sciences* **62**(19-20): 2194-2203.

Aphasizhev, R. and Aphasizheva, I. (2008) Terminal RNA uridylyltransferases of trypanosomes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1779**(4): 270-280.

Aphasizhev, R., Aphasizheva, I., Nelson, R. E., Gao, G., Simpson, A. M., Kang, X., Falick, A. M., Sbicego, S. and Simpson, L. (2003) Isolation of a U-insertion/deletion editing complex from Leishmania tarentolae mitochondria. *EMBO Journal* **22**(4): 913-924.

Aphasizhev, R., Aphasizheva, I. and Simpson, L. (2003) A tale of two TUTases. *PNAS* **100**(19): 10617-10622.

Aphasizhev, R., Aphasizheva, I. and Simpson, L. (2004) Multiple terminal uridylyltransferases of trypanosomes. *FEBS Letters* **572**(1-3): 15-18.

Aphasizhev, R., Sbicego, S., Peris, M., Jang, S. H., Aphasizheva, I., Simpson, A. M., Rivlin, A. and Simpson, L. (2002) Trypanosome mitochondrial 3' terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing. *Cell* **108**(5): 637-648.

Aphasizheva, I. and Aphasizhev, R. (2010) RET1-catalyzed uridylylation shapes the mitochondrial transcriptome in Trypanosoma brucei. *Molecular and Cellular Biology* **30**(6): 1555-1567.

Aphasizheva, I., Aphasizhev, R. and Simpson, L. (2004) RNA-editing terminal uridylyl transferase 1: identification of functional domains by mutational analysis. *Journal of Biological Chemistry* **279**(23): 24123-24130.

Aphasizheva, I., Ringpis, G.-E., Weng, J., Gershon, P. D., Lathrop, R. H. and Aphasizhev, R. (2009) Novel TUTase associates with an editosome-like complex in mitochondria of *Trypanosoma brucei*. *RNA* **15**(7): 1322-1337.

Aravin, A. A., van der Heijden, G. W., Castaneda, J., Vagin, V. V., Hannon, G. J. and Bortvin, A. (2009) Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genetics* **5**(12): e1000764.

Aravind, L. and Koonin, E. V. (1999) DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Research* **27**(7): 1609-1618.

Aslett, M. and Wood, V. (2006) Gene ontology annotation status of the fission yeast genome: preliminary coverage approaches 100%. *Yeast* **23**(13): 913-919.

Baccarini, A., Chauhan, H., Gardner, T. J., Jayaprakash, A. D., Sachidanandam, R. and Brown, B. D. (2011) Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells. *Current Biology* **21**(5): 369-376.

Badis, G., Saveanu, C., Fromont-Racine, M. and Jacquier, A. (2004) Targeted mRNA degradation by deadenylation-independent decapping. *Molecular Cell* **15**(1): 5-15.

Bahler, J., Wu, J. Q., Longtine, M. S., Shah, N. G., McKenzie, A., Steever, A. B., Wach, A., Philippsen, P. and Pringle, J. R. (1998) Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**(10): 943-951.

Bakalara, N., Simpson, A. M. and Simpson, L. (1989) The *Leishmania* kinetoplast-mitochondrion contains terminal uridylyltransferase and RNA ligase activities. *Journal of Biological Chemistry* **264**(31): 18679-18686.

Baker, E. N. and Hubbard, R. E. (1984) Hydrogen bonding in globular proteins. *Progress in Biophysics and Molecular Biology* **44**: 97-179.

Balzer, E., Heine, C., Jiang, Q., Lee, V. M. and Moss, E. G. (2010) LIN28 alters cell fate succession and acts independently of the *let-7* microRNA during neurogliogenesis in vitro. *Development* **137**(6): 891-900.

Barbrook, A. C., Dorrell, R. G., Burrows, J., Plenderleith, L. J., Nisbet, R. E. and Howe, C. J. (2012) Polyuridylylation and processing of transcripts from multiple gene minicircles in chloroplasts of the dinoflagellate Amphidinium carterae. *Plant Molecular Biology* **79**(4-5): 347-357.

Barnard, D. C., Ryan, K., Manley, J. L. and Richter, J. D. (2004) Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. *Cell* **119**(5): 641-651.

Batista, P. J., Ruby, J. G., Claycomb, J. M., Chiang, R., Fahlgren, N., Kasschau, K. D., Chaves, D. A., Gu, W., Vasale, J. J., Duan, S., Conte, D., Jr., Luo, S., Schroth, G. P., Carrington, J. C., Bartel, D. P. and Mello, C. C. (2008) PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Molecular Cell* **31**(1): 67-78.

Bayne, E. H., White, S. A., Kagansky, A., Bijos, D. A., Sanchez-Pulido, L., Hoe, K. L., Kim, D. U., Park, H. O., Ponting, C. P., Rappsilber, J. and Allshire, R. C. (2010) Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity. *Cell* **140**(5): 666-677.

Benoit, P., Papin, C., Kwak, J. E., Wickens, M. and Simonelig, M. (2008) PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in Drosophila. *Development* **135**(11): 1969-1979.

Berezikov, E., Robine, N., Samsonova, A., Westholm, J. O., Naqvi, A., Hung, J. H., Okamura, K., Dai, Q., Bortolamiol-Becet, D., Martin, R., Zhao, Y., Zamore, P. D., Hannon, G. J., Marra, M. A., Weng, Z., Perrimon, N. and Lai, E. C. (2011) Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. *Genome Research* **21**(2): 203-215.

Berndt, H., Harnisch, C., Rammelt, C., Stohr, N., Zirkel, A., Dohm, J. C., Himmelbauer, H., Tavanez, J. P., Huttelmaier, S. and Wahle, E. (2012) Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* **18**(5): 958-972.

Billi, A. C., Alessi, A. F., Khivansara, V., Han, T., Freeberg, M., Mitani, S. and Kim, J. K. (2012) The *Caenorhabditis elegans* HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs. *PLoS Genetics* **8**(4): e1002617.

Blum, B., Bakalara, N. and Simpson, L. (1990) A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* **60**(2): 189-198.

Blum, B. and Simpson, L. (1990) Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. *Cell* **62**(2): 391-397.

Booth, B. L., Jr. and Pugh, B. F. (1997) Identification and characterization of a nuclease specific for the 3' end of the U6 small nuclear RNA. *Journal of Biological Chemistry* **272**(2): 984-991.

Bourc'his, D. and Voinnet, O. (2010) A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* **330**(6004): 617-622.

Boutet, S., Vazquez, F., Liu, J., Béclin, C., Fagard, M., Gratias, A., Morel, J.-B., Crété, P., Chen, X. and Vaucheret, H. (2003) *Arabidopsis* HEN1: A genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Current Biology* **13**(10): 843-848.

Bru, C., Courcelle, E., Carrère, S., Beausse, Y., Dalmar, S. and Kahn, D. (2005) The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Research* **33**(Database Issue): D212-D215.

Buhler, M., Haas, W., Gygi, S. P. and Moazed, D. (2007) RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* **129**(4): 707-721.

Buhler, M., Spies, N., Bartel, D. P. and Moazed, D. (2008) TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway. *Nature: Structural and Molecular Biology* **15**(10): 1015-1023.

Buker, S. M., Iida, T., Buhler, M., Villen, J., Gygi, S. P., Nakayama, J. and Moazed, D. (2007) Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. *Nature: Structural and Molecular Biology* **14**(3): 200-207.

Büssing, I., Slack, F. J. and Grosshans, H. (2008) let-7 microRNAs in development, stem cells and cancer. *Trends in Molecular Medicine* **14**(9): 400-409.

Cam, H. P. (2010) Roles of RNAi in chromatin regulation and epigenetic inheritance. *Epigenomics* **2**(5): 613-626.

Cam, H. P., Sugiyama, T., Chen, E. S., Chen, X., FitzGerald, P. C. and Grewal, S. I. (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nature Genetics* **37**(8): 809-819.

Campbell, D. A., Spithill, T. W., Samaras, N., Simpson, A. M. and Simpson, L. (1989) Sequence of a cDNA for the ND1 gene from Leishmania major: potential uridine addition in the polyadenosine tail. *Molecular and Biochemical Parasitology* **36**(2): 197-199.

Cao, D., Leffert, J. J., McCabe, J., Kim, B. and Pizzorno, G. (2005) Abnormalities in uridine homeostatic regulation and pyrimidine nucleotide metabolism as a consequence of the deletion of the uridine phosphorylase gene. *Journal of Biological Chemistry* **280**(22): 21169-21175.

Carnes, J., Trotter, J. R., Peltan, A., Fleck, M. and Stuart, K. (2008) RNA editing in *Trypanosoma brucei* requires three different editosomes. *Molecular and Cellular Biology* **28**(1): 122-130.

Castaño, I. B., Brzoska, P. M., Sadoff, B. U., Chen, H. and Christman, M. F. (1996) Mitotic chromosome condensation in the rDNA requires TRF4 and DNA topoisomerase I in *Saccharomyces cerevisiae*. *Genes and Development* **10**(20): 2564-2576.

Castaño, I. B., Heath-Pagliuso, S., Sadoff, B. U., Fitzhugh, D. J. and Christman, M. F. (1996) A novel family of TRF (DNA topoisomerase l-related function) genes required for proper nuclear segregation. *Nucleic Acids Research* **24**(12): 2404-2410.

Chan, S., Choi, E. A. and Shi, Y. (2011) Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdisciplinary Reviews: RNA* **2**(3): 321-335.

Chen, C.-C. G., Simard, M. J., Tabara, H., Brownell, D. R., McCollough, J. A. and Mello, C. C. (2005) A member of the polymerase β nucleotidyltransferase superfamily Is required for RNA interference in *C. elegans*. *Current Biology* **15**(4): 378-383.

Chen, C. Y. and Shyu, A. B. (2011) Mechanisms of deadenylation-dependent decay. *WIREs: RNA* **2**(2): 167-183.

Chen, D., Toone, W. M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N. and Bahler, J. (2003) Global transcriptional responses of fission yeast to environmental stress. *Molecular Biology of the Cell* **14**(1): 214-229.

Chen, X., Liu, J., Cheng, Y. and Jia, D. (2002) HEN1 functions pleiotropically in *Arabidopsis* development and acts in C function in the flower. *Development* **129**(5): 1085-1094.

Chiang, H. R., Schoenfeld, L. W., Ruby, J. G., Auyeung, V. C., Spies, N., Baek, D., Johnston, W. K., Russ, C., Luo, S., Babiarz, J. E., Blelloch, R., Schroth, G. P., Nusbaum, C. and Bartel, D. P. (2010) Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes and Development* **24**(10): 992-1009.

Choi, Y. S., Patena, W., Leavitt, A. D. and McManus, M. T. (2012) Widespread RNA 3'-end oligouridylation in mammals. *Bioinformatics* **18**: 394-401.

Chowdhury, A., Mukhopadhyay, J. and Tharun, S. (2007) The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA* **13**(7): 998-1016.

Claycomb, J. M., Batista, P. J., Pang, K. M., Gu, W. F., Vasale, J. J., van Wolfswinkel, J. C., Chaves, D. A., Shirayama, M., Mitani, S., Ketting, R. F., Conte, D. and Mello, C. C. (2009) The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* **139**(1): 123-134.

Colmenares, S. U., Buker, S. M., Buhler, M., Dlakic, M. and Moazed, D. (2007) Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Molecular Cell* **27**(3): 449-461.

Couttet, P., Fromont-Racine, M., Steel, D., Pictet, R. and Grange, T. (1997) Messenger RNA deadenylylation precedes decapping in mammalian cells. *PNAS* **94**(11): 5628-5633.

Couvillion, M. T., Lee, S. R., Hogstad, B., Malone, C. D., Tonkin, L. A., Sachidanandam, R., Hannon, G. J. and Collins, K. (2009) Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of *Tetrahymena thermophila*. *Genes & Development* **23**(17): 2016-2032.

Crawford, N. M. and Baltimore, D. (1983) Genome-linked protein VPg of poliovirus is present as free VPg and VPg-pUpU in poliovirus-infected cells. *PNAS* **80**(24): 7452-7455.

Dang, Y. and Green, B. R. (2009) Substitutional editing of *Heterocapsa triquetra* chloroplast transcripts and a folding model for its divergent chloroplast 16S rRNA. *Gene* **442**(1-2): 73-80.

Dang, Y. and Green, B. R. (2010) Long transcripts from dinoflagellate chloroplast minicircles suggest "rolling circle" transcription. *Journal of Biological Chemistry* **285**(8): 5196-5203.

Davey, J. (1992) Mating pheromones of the fission yeast *Schizosaccharomyces pombe*: purification and structural characterization of M-factor and isolation and analysis of two genes encoding the pheromone. *EMBO Journal* **11**(3): 951-960.

Desjardins, A., Yang, A., Bouvette, J., Omichinski, J. G. and Legault, P. (2012) Importance of the NCp7-like domain in the recognition of pre-let-7g by the pluripotency factor Lin28. *Nucleic Acids Research* **40**(4): 1767-1777.

Dickinson, H., Tretbar, S., Betat, H. and Morl, M. (2012) The TRAMP complex shows tRNA editing activity in *S. cerevisiae*. *Molecular Biology and Evolution* **29**(5): 1451-1459.

Dölken, L., Perot, J., Cognat, V., Alioua, A., John, M., Soutschek, J., Ruzsics, Z., Koszinowski, U., Voinnet, O. and Pfeffer, S. (2007) Mouse cytomegalovirus microRNAs dominate the cellular small RNA profile during lytic infection and show features of posttranscriptional regulation. *Journal of Virology* **81**(24): 13771-13782.

Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C. C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Gahazal, P. and Koszinowski, U. H. (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**: 1959-1972.

Dominski, Z. and Marzluff, W. F. (1999) Formation of the 3' end of histone mRNA. *Gene* **239**(1): 1-14.

Dominski, Z., Yang, X. C., Kaygun, H., Dadlez, M. and Marzluff, W. F. (2003) A 3' exonuclease that specifically interacts with the 3' end of histone mRNA. *Molecular Cell* **12**(2): 295-305.

Dosztányi, Z., Csizmók, V., Tompa, P. and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**: 3433-3434.

Du, L. and Richter, J. D. (2005) Activity-dependent polyadenylation in neurons. *RNA* **11**(9): 1340-1347.

Duchaine, T. F., Wohlschlegel, J. A., Kennedy, S., Bei, Y., Conte, D., Jr., Pang, K., Brownell, D. R., Harding, S., Mitani, S., Ruvkun, G., Yates, J. R., 3rd and Mello, C. C. (2006) Functional proteomics reveals

the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* **124**(2): 343-354.

Eckmann, C. R., Crittenden, S. L., Suh, N. and Kimble, J. (2004) GLD-3 and control of the mitosis/meiosis decision in the germline of *Caenorhabditis elegans*. *Genetics* **168**(1): 147-160.

Eckmann, C. R., Kraemer, B., Wickens, M. and Kimble, J. (2002) GLD-3, a Bicaudal-C homolog that inhibits FBF to control germline sex determination in *C. elegans*. *Developmental Cell* **3**(5): 697-710.

Ekwall, K., Nimmo, E. R., Javerzat, J. P., Borgstrom, B., Egel, R., Cranston, G. and Allshire, R. (1996) Mutations in the fission yeast silencing factors *clr4+* and *rik1+* disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. *Journal of Cell Science* **109 (Pt 11)**: 2637-2648.

Ernst, N. L., Panicucci, B., Igo Jr, R. P., Panigrahi, A. K., Salavati, R. and Stuart, K. (2003) TbMP57 is a 3' terminal uridylyl yransferase (TUTase) of the *Trypanosoma brucei* editosome. *Molecular Cell* **11**(6): 1525-1536.

Etheridge, R. D., Aphasizheva, I., Gershon, P. D. and Aphasizhev, R. (2008) 3' adenylation determines mRNA abundance and monitors completion of RNA editing in *T. brucei* mitochondria. *EMBO Journal* **27**(11): 1596-1608.

Etheridge, R. D., Clemens, D. M., Gershon, P. D. and Aphasizhev, R. (2009) Identification and characterization of nuclear non-canonical poly(A) polymerases from *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **164**(1): 66-73.

Feng, C., Neumeister, V., Ma, W., Xu, J., Lu, L., Bordeaux, J., Maihle, N. J., Rimm, D. L. and Huang, Y. (2012) Lin28 regulates HER2 and promotes malignancy through multiple mechanisms. *Cell Cycle* **11**(13): 2486-2494.

Fernandez-Valverde, S. L., Taft, R. J. and Mattick, J. S. (2010) Dynamic isomiR regulation in *Drosophila* development. *RNA* **16**(10): 1881-1888.

Flanegan, J. B., Petterson, R. F., Ambros, V., Hewlett, N. J. and Baltimore, D. (1977) Covalent linkage of a protein to a defined nucleotide sequence at the 5'-terminus of virion and replicative intermediate RNAs of poliovirus. *PNAS* **74**(3): 961-965.

Ford, L. P. and Wilusz, J. (1999) 3'-Terminal RNA structures and poly(U) tracts inhibit initiation by a 3'-->5' exonuclease in vitro. *Nucleic Acids Research* **27**(4): 1159-1167.

Furnari, F. B., Adams, M. D. and Pagano, J. S. (1993) Unconventional processing of the 3' termini of the Epstein-Barr Virus DNA polymerase messenger RNA. *PNAS* **90**(2): 378-382.

Gabel, H. W. and Ruvkun, G. (2008) The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi. *Nature Structural and Molecular Biology* **15**(5): 531-533.

Gallwitz, D. and Mueller, G. C. (1969) Histone synthesis *in vitro* on HeLa cell microsomes. *Journal of Biological Chemistry* **244**(21): 5947-5952.

Galzitskaya, O. V., Garbuzynskiy, S. O. and Lobanov, M. Y. (2006) FoldUnfold: web server for the prediction of disordered regions in protein chain. *Bioinformatics* **22**(23): 2948-2949.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*. Walker, J. M., Humana Press**:** 571-607.

Gerber, K., Wimmer, E. and Paul, A. V. (2001) Biochemical and genetic studies of the initiation of human rhinovirus 2 RNA replication: identification of a *cis*-replicating element in the coding sequence of $2A^{pro}$. *Journal of Virology* **75**(22): 10979-10990.

Glahder, J. A. and Norrild, B. (2011) Involvement of hGLD-2 in cytoplasmic polyadenylation of human p53 mRNA. *Acta Pathologica, Microbiologica et Immunologica Scandinavica* **119**(11): 769-775.

Goldstrohm, A. C. and Wickens, M. (2008) Multifunctional deadenylase complexes diversify mRNA control. *Nature Reviews: Molecular Cell Biology* **9**(4): 337-344.

Gonzales, M. L., Mellman, D. L. and Anderson, R. A. (2008) CKIα is associated with and phosphorylates Star-PAP and is also required for expression of select Star-PAP target messenger RNAs. *Journal of Biological Chemistry* **283**(18): 12665-12673.

Goodfellow, I., Chaudhry, Y., Richardson, A., Meredith, J., Almond, J. W., Barclay, W. and Evans, D. J. (2000) Identification of a *cis*-acting replication element within the poliovirus coding region. *Journal of Virology* **74**(10): 4590-4600.

Goto, D. B. and Nakayama, J. I. (2011) RNA and epigenetic silencing: insight from fission yeast. *Development, Growth and Differentiation*.

Gradwohl, G., Menissier de Murcia, J., Molinete, M., Simonin, F. and de Murcia, G. (1989) Expression of functional zinc finger domain of human poly(ADP-ribose)polymerase in *E. coli*. *Nucleic Acids Research* **17**(17): 7112.

Gradwohl, G., Menissier de Murcia, J. M., Molinete, M., Simonin, F., Koken, M., Hoeijmakers, J. H. and de Murcia, G. (1990) The second zinc-finger domain of poly(ADP-ribose) polymerase determines specificity for single-stranded breaks in DNA. *PNAS* **87**(8): 2990-2994.

Grewal, S. I. (2010) RNAi-dependent formation of heterochromatin and its diverse functions. *Current Opinion in Genetics and Development* **20**(2): 134-141.

Grewal, S. I. and Klar, A. J. (1997) A recombinationally repressed region between *mat2* and *mat3* loci shares homology to centromeric repeats and regulates directionality of mating-type switching in fission yeast. *Genetics* **146**(4): 1221-1238.

Grzechnik, P. and Kufel, J. (2008) Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Molecular Cell* **32**(2): 247-258.

Gu, W., Shirayama, M., Conte Jr, D., Vasale, J., Batista, P. J., Claycomb, J. M., Moresco, J. J., Youngman, E. M., Keys, J., Stoltz, M. J., Chen, C.-C. G., Chaves, D. A., Duan, S., Kasschau, K. D., Fahlgren, N., Yates Iii, J. R., Mitani, S., Carrington, J. C. and Mello, C. C. (2009) Distinct Argonaute-mediated

22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Molecular Cell* **36**(2): 231-244.

Guleria, P., Mahajan, M., Bhardwaj, J. and Yadav, S. K. (2011) Plant small RNAs: biogenesis, mode of action and their roles in abiotic stresses. *Genomics Proteomics Bioinformatics* **9**(6): 183-199.

Gullerova, M., Moazed, D. and Proudfoot, N. J. (2011) Autoregulation of convergent RNAi genes in fission yeast. *Genes and Development* **25**(6): 556-568.

Gullerova, M. and Proudfoot, N. J. (2008) Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. *Cell* **132**(6): 983-995.

Hagan, J. P., Piskounova, E. and Gregory, R. I. (2009) Lin28 recruits the TUTase Zcchc11 to inhibit *let-7* maturation in mouse embryonic stem cells. *Nature: Structural and Molecular Biology* **16**(10): 1021-U1033.

Halic, M. and Moazed, D. (2010) Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. *Cell* **140**(4): 504-516.

Hansen, K. R., Burns, G., Mata, J., Volpe, T. A., Martienssen, R. A., Bahler, J. and Thon, G. (2005) Global effects on gene expression in fission yeast by silencing and RNA interference machineries. *Molecular and Cellular Biology* **25**(2): 590-601.

Hansen, K. R., Ibarra, P. T. and Thon, G. (2006) Evolutionary-conserved telomere-linked helicase genes of fission yeast are repressed by silencing factors, RNAi components and the telomere-binding protein Taz1. *Nucleic Acids Research* **34**(1): 78-88.

Heintz, N., Sive, H. L. and Roeder, R. G. (1983) Regulation of human histone gene expression: kinetics of accumulation and changes in the rate of synthesis and in the half-lives of individual histone mRNAs during the HeLa cell cycle. *Molecular and Cellular Biology* **3**(4): 539-550.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J. and Kim, V. N. (2008) Lin28 mediates the terminal uridylation of *let-7* precursor microRNA. *Molecular Cell* **32**(2): 276-284.

Heo, I., Joo, C., Kim, Y. K., Ha, M., Yoon, M. J., Cho, J., Yeom, K. H., Han, J. and Kim, V. N. (2009) TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**(4): 696-708.

Hirai, H., Lee, D. I., Natori, S. and Sekimizu, K. (1988) Uridylation of U6 RNA in a nuclear extract in Ehrlich ascites tumor cells. *Journal of Biochemistry* **104**(6): 991-994.

Holm, L. and Sander, C. (1995) DNA polymerase β belongs to an ancient nucleotidyltransferase superfamily. *Trends in Biochemical Sciences* **20**: 345-347.

Hong, E. J., Villen, J., Gerace, E. L., Gygi, S. P. and Moazed, D. (2005) A cullin E3 ubiquitin ligase complex associates with Rik1 and the Clr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation. *RNA Biology* **2**(3): 106-111.

Horn, P. J., Bastie, J. N. and Peterson, C. L. (2005) A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. *Genes and Development* **19**(14): 1705-1714.

Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P. and Zamore, P. D. (2007) The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Current Biology* **17**(14): 1265-1272.

Houseley, J., Kotovic, K., El Hage, A. and Tollervey, D. (2007) Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. *EMBO Journal* **26**(24): 4996-5006.

Houseley, J. and Tollervey, D. (2006) Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO Reports* **7**(2): 205-211.

Houwing, S., Berezikov, E. and Ketting, R. F. (2008) Zili is required for germ cell differentiation and meiosis in zebrafish. *EMBO Journal* **27**(20): 2702-2711.

Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D. V., Blaser, H., Raz, E., Moens, C. B., Plasterk, R. H., Hannon, G. J., Draper, B. W. and Ketting, R. F.

(2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**(1): 69-82.

Huang, Y., Ji, L., Huang, Q., Vassylyev, D. G., Chen, X. and Ma, J. B. (2009) Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature* **461**(7265): 823-827.

Huisinga, K. L. and Elgin, S. C. (2009) Small RNA-directed heterochromatin formation in the context of development: what flies might learn from fission yeast. *Biochimica et Biophysica Acta* **1789**(1): 3-16.

Ibrahim, F., Rohr, J., Jeong, W.-J., Hesson, J. and Cerutti, H. (2006) Untemplated oligoadenylation promotes degradation of RISC-cleaved transcripts. *Science* **314**(5807): 1893.

Ibrahim, F., Rymarquis, L. A., Kim, E.-J., Becker, J., Balassa, E., Green, P. J. and Cerutti, H. (2010) Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in *Chlamydomonas*. *PNAS* **107**(8): 3906-3911.

Iida, T., Kawaguchi, R. and Nakayama, J.-i. (2006) Conserved ribonuclease, Eri1, negatively regulates heterochromatin assembly in fission yeast. *Current Biology* **16**(14): 1459-1464.

Iida, T., Nakayama, J. and Moazed, D. (2008) siRNA-mediated heterochromatin establishment requires HP1 and is associated with antisense transcription. *Molecular Cell* **31**(2): 178-189.

Iliopoulos, D., Hirsch, H. A. and Struhl, K. (2009) An epigenetic switch involving NF-kappaB, Lin28, let-7 microRNA, and IL6 links inflammation to cell transformation. *Cell* **139**(4): 693-706.

Irvine, D. V., Zaratiegui, M., Tolia, N. H., Goto, D. B., Chitwood, D. H., Vaughn, M. W., Joshua-Tor, L. and Martienssen, R. A. (2006) Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313**(5790): 1134-1137.

Jain, R. and Shuman, S. (2010) Bacterial Hen1 is a 3' terminal RNA ribose 2'-O-methyltransferase component of a bacterial RNA repair cassette. *RNA* **16**(2): 316-323.

Janouškovec, J., Horák, A., Oborník, M., Lukeš, J. and Keeling, P. J. (2010) A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *PNAS* **107**(24): 10949-10954.

Jia, S., Kobayashi, R. and Grewal, S. I. (2005) Ubiquitin ligase component Cul4 associates with Clr4 histone methyltransferase to assemble heterochromatin. *Nature: Cell Biology* **7**(10): 1007-1013.

Jia, S., Noma, K. and Grewal, S. I. (2004) RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science* **304**(5679): 1971-1976.

Jin, J., Jing, W., Lei, X. X., Feng, C., Peng, S., Boris-Lawrie, K. and Huang, Y. (2011) Evidence that Lin28 stimulates translation by recruiting RNA helicase A to polysomes. *Nucleic Acids Research* **39**(9): 3724-3734.

Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* **292**: 195-202.

Jones, M. R., Quinton, L. J., Blahna, M. T., Neilson, J. R., Fu, S. N., Ivanov, A. R., Wolf, D. A. and Mizgerd, J. P. (2009) Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nature: Cell Biology* **11**(9): 1157-U1258.

Kadaba, S., Wang, X. Y. and Anderson, J. T. (2004) Nuclear RNA surveillance in *Saccharomyces cerevisiae*: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA* **12**(3): 508-521.

Kagansky, A., Folco, H. D., Almeida, R., Pidoux, A. L., Boukaba, A., Simmer, F., Urano, T., Hamilton, G. L. and Allshire, R. C. (2009) Synthetic heterochromatin bypasses RNAi and centromeric repeats to establish functional centromeres. *Science* **324**(5935): 1716-1719.

Kamminga, L. M., Luteijn, M. J., den Broeder, M. J., Redl, S., Kaaij, L. J. T., Roovers, E. F., Ladurner, P., Berezikov, E. and Ketting, R. F. (2010) Hen1 is required for oocyte development and piRNA stability in zebrafish. *EMBO Journal* **29**(21): 3688-3700.

Kamminga, L. M., van Wolfswinkel, J. C., Luteijn, M. J., Kaaij, L. J., Bagijn, M. P., Sapetschnig, A., Miska, E. A., Berezikov, E. and Ketting, R. F. (2012) Differential impact of the HEN1 homolog HENN-1 on 21U and 26G RNAs in the germline of *Caenorhabditis elegans*. *PLoS Genetics* **8**(7): e1002702.

Kanoh, J., Sadaie, M., Urano, T. and Ishikawa, F. (2005) Telomere binding protein Taz1 establishes Swi6 heterochromatin independently of RNAi at telomeres. *Current Biology* **15**(20): 1808-1819.

Kao, C. Y. and Read, L. K. (2007) Targeted depletion of a mitochondrial nucleotidyltransferase suggests the presence of multiple enzymes that polymerize mRNA 3' tails in *Trypanosoma brucei* mitochondria. *Molecular Biochemistry and Parasitology* **154**(2): 158-169.

Kashiwabara, S., Nakanishi, T., Kimura, M. and Baba, T. (2008) Non-canonical poly(A) polymerase in mammalian gametogenesis. *Biochim Biophys Acta* **1779**(4): 230-238.

Kataoka, K. and Mochizuki, K. (2011) Programmed DNA elimination in *Tetrahymena*: a small RNA-mediated genome surveillance mechanism. *Advances in Experimental Medicine and Biology* **722**: 156-173.

Kato, H., Goto, D. B., Martienssen, R. A., Urano, T., Furukawa, K. and Murakami, Y. (2005) RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309**(5733): 467-469.

Katoh, T., Sakaguchi, Y., Miyauchi, K., Suzuki, T., Kashiwabara, S.-i., Baba, T. and Suzuki, T. (2009) Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes & Development* **23**(4): 433-438.

Kawaoka, S., Izumi, N., Katsuma, S. and Tomari, Y. (2011) 3' end formation of PIWI-interacting RNAs *in vitro*. *Molecular Cell* **43**(6): 1015-1022.

Kaygun, H. and Marzluff, W. F. (2005) Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. *Nature: Structural and Molecular Biology* **12**(9): 794-800.

Kaygun, H. and Marzluff, W. F. (2005) Translation termination is involved in histone mRNA degradation when DNA replication is inhibited. *Molecular and Cellular Biology* **25**(16): 6879-6888.

Kennedy, S., Wang, D. and Ruvkun, G. (2004) A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**(6975): 645-649.

Kim, J. H. and Richter, J. D. (2006) Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. *Molecular Cell* **24**(2): 173-183.

Kim, J. H. and Richter, J. D. (2007) RINGO/cdk1 and CPEB mediate poly(A) tail stabilization and translational regulation by ePAB. *Genes & Development* **21**(20): 2571-2579.

Kim, K. W., Wilson, T. L. and Kimble, J. (2010) GLD-2/RNP-8 cytoplasmic poly(A) polymerase is a broad-spectrum regulator of the oogenesis program. *PNAS* **107**(40): 17445-17450.

Kirino, Y. and Mourelatos, Z. (2007) The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA* **13**(9): 1397-1401.

Kirino, Y. and Mourelatos, Z. (2007) Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nature: Structural and Molecular Biology* **14**(4): 347-348.

Klattenhoff, C. and Theurkauf, W. (2008) Biogenesis and germline functions of piRNAs. *Development* **135**(1): 3-9.

Kunkel, G. R., Maser, R. L., Calvet, J. P. and Pederson, T. (1986) U6 small nuclear RNA is transcribed by RNA polymerase III. *PNAS* **83**(22): 8575-8579.

Kurth, H. M. and Mochizuki, K. (2009) 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA* **15**(4): 675-685.

Kwak, J. E., Drier, E., Barbee, S. A., Ramaswami, M., Yin, J. C. P. and Wickens, M. (2008) GLD2 poly(A) polymerase is required for long-term memory. *PNAS* **105**(38): 14644-14649.

Kwak, J. E., Wang, L. T., Ballantyne, S., Kimble, J. and Wickens, M. (2004) Mammalian GLD-2 homologs are poly(A) polymerases. *PNAS* **101**(13): 4407-4412.

Kwak, J. E. and Wickens, M. (2007) A family of poly(U) polymerases. *RNA* **13**(6): 860-867.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A. and Tollervey, D. (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**(5): 713-724.

Laishram, R. S. and Anderson, R. A. (2010) The poly A polymerase Star-PAP controls 3'-end cleavage by promoting CPSF interaction and specificity toward the pre-mRNA. *EMBO Journal* **29**(24): 4132-4145.

Laishram, R. S., Barlow, C. A. and Anderson, R. A. (2011) CKI isoforms α and ε regulate Star-PAP target messages by controlling Star-PAP poly(A) polymerase activity and phosphoinositide stimulation. *Nucleic Acids Research* **39**(18): 7961-7973.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A. O., Landthaler, M., Lin, C., Socci, N. D., Hermida, L., Fulci, V., Chiaretti, S., Foa, R., Schliwka, J., Fuchs, U., Novosel, A., Muller, R. U., Schermer, B., Bissels, U., Inman, J., Phan, Q., Chien, M., Weir, D. B., Choksi, R., De Vita, G., Frezzetti, D., Trompeter, H. I., Hornung, V., Teng, G., Hartmann, G., Palkovits, M., Di Lauro, R., Wernet, P., Macino, G., Rogler, C. E., Nagle, J. W., Ju, J., Papavasiliou, F. N., Benzing, T., Lichter, P., Tam, W., Brownstein, M. J., Bosio, A., Borkhardt, A., Russo, J. J., Sander, C., Zavolan, M. and Tuschl, T. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**(7): 1401-1414.

Lange, H., Sement, F. M., Canaday, J. and Gagliardi, D. (2009) Polyadenylation-assisted RNA degradation processes in plants. *Trends in Plant Sciences* **14**(9): 497-504.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21): 2947-2948.

Lee, R. C., Hammell, C. M. and Ambros, V. (2006) Interacting endogenous and exogenous RNAi pathways in *Caenorhabditis elegans*. *RNA* **12**(4): 589-597.

Lee, S. R. and Collins, K. (2006) Two classes of endogenous small RNAs in *Tetrahymena thermophila*. *Genes & Development* **20**(1): 28-33.

Lee, S. R. and Collins, K. (2007) Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nature: Structural and Molecular Biology* **14**(7): 604-610.

Lee, S. R., Talsky, K. B. and Collins, K. (2009) A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis. *RNA* **15**(7): 1363-1374.

Lee, Y. F., Nomoto, A., Detjen, B. M. and Wimmer, E. (1977) A protein covalently linked to poliovirus genome RNA. *PNAS* **74**(1): 59-63.

Lehrbach, N. J., Armisen, J., Lightfoot, H. L., Murfitt, K. J., Bugaut, A., Balasubramanian, S. and Miska, E. A. (2009) LIN-28 and the poly(U) polymerase PUP-2 regulate *let-7* microRNA processing in *Caenorhabditis elegans*. *Nature Structural and Molecular Biology* **16**(10): 1016-U1027.

Lei, X. X., Xu, J., Ma, W., Qiao, C., Newman, M. A., Hammond, S. M. and Huang, Y. (2012) Determinants of mRNA recognition and translation regulation by Lin28. *Nucleic Acids Research* **40**(8): 3574-3584.

Lejeune, E. and Allshire, R. C. (2011) Common ground: small RNA programming and chromatin modifications. *Current Opinion in Cell Biology* **23**(3): 258-265.

Lemieux, C., Marguerat, S., Lafontaine, J., Barbezier, N., Bahler, J. and Bachand, F. (2011) A pre-mRNA degradation pathway that selectively targets intron-containing genes requires the nuclear poly(A)-binding protein. *Molecular Cell* **44**(1): 108-119.

Li, F., Goto, D. B., Zaratiegui, M., Tang, X., Martienssen, R. and Cande, W. Z. (2005) Two novel proteins, Dos1 and Dos2, interact with Rik1 to regulate heterochromatic RNA interference and histone modification. *Current Biology* **15**(16): 1448-1457.

Li, F., Martienssen, R. and Cande, W. Z. (2011) Coordination of DNA replication and histone modification by the Rik1-Dos2 complex. *Nature* **475**(7355): 244-248.

Li, J. J., Yang, Z. Y., Yu, B., Liu, J. and Chen, X. M. (2005) Methylation protects miRNAs and siRNAs from a 3 '-end uridylation activity in *Arabidopsis*. *Current Biology* **15**(16): 1501-1507.

Li, W., Laishram, Rakesh S., Ji, Z., Barlow, Christy A., Tian, B. and Anderson, Richard A. (2012) Star-PAP control of BIK expression and apoptosis is regulated by nuclear PIPKα and PKCδ signaling. *Molecular Cell* **45**(1): 25-37.

Lightfoot, H. L., Bugaut, A., Armisen, J., Lehrbach, N. J., Miska, E. A. and Balasubramanian, S. (2011) A LIN28-dependent structural change in *pre-let-7g* directly inhibits dicer processing. *Biochemistry* **50**(35): 7514-7521.

Lingel, A., Simon, B., Izaurralde, E. and Sattler, M. (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nature Structural and Molecular Biology* **11**(6): 576-577.

Lund, E. and Dahlberg, J. E. (1992) Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. *Science* **255**(5042): 327-330.

Lunde, B. M., Magler, I. and Meinhart, A. (2012) Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res*.

Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science* **252**: 1162-1164.

Ma, J.-B., Ye, K. and Patel, D. J. (2004) Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**(6989): 318-322.

Machín, A., Martín-Alonso, J. M. and F., P. (2001) Identification of the amino acid residue involved in rabbit hemorrhagic disease virus VPg uridylylation. *Journal of Biological Chemistry* **276**(30): 27787-27792.

MacNeill, S. A., Moreno, S., Reynolds, N., Nurse, P. and Fantes, P. A. (1996) The fission yeast Cdc1 protein, a homologue of the small subunit of DNA polymerase delta, binds to Pol3 and Cdc27. *EMBO Journal* **15**(17): 4613-4628.

Mandell, J. G., Goodrich, K. J., Bahler, J. and Cech, T. R. (2005) Expression of a RecQ helicase homolog affects progression through crisis in fission yeast lacking telomerase. *Journal of Biological Chemistry* **280**(7): 5249-5257.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng, C. and Bryant, S. H. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* **39**(Database issue): D225-229.

Martinson, H. G. (2011) An active role for splicing in 3'-end formation. *WIREs: RNA* **2**(4): 459-470.

Mata, J. and Bähler, J. (2006) Global roles of Ste11p, cell type, and pheromone in the control of gene expression during early sexual differentiation in fission yeast. *PNAS* **103**(42): 15517-15522.

Mata, J., Lyne, R., Burns, G. and Bähler, J. (2002) The transcriptional program of meiosis and sporulation in fission yeast. *Nature: Genetics* **32**(1): 143-147.

Matsuyama, A., Arai, R., Yashiroda, Y., Shirai, A., Kamata, A., Sekido, S., Kobayashi, Y., Hashimoto, A., Hamamoto, M., Hiraoka, Y., Horinouchi, S. and Yoshida, M. (2006) ORFeome cloning and global analysis of protein localization in the fission yeast *Schizosaccharomyces pombe*. *Nature: Biotechnology* **24**(7): 841-847.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B. and Matzke, A. J. (2009) RNA-mediated chromatin-based silencing in plants. *Current Opinion in Cell Biology* **21**(3): 367-376.

Maundrell, K. (1993) Thiamine-repressible expression vectors pREP and pRIP for fission yeast. *Gene* **123**(1): 127-130.

Mayr, F., Schütz, A., Döge, N. and Heinemann, U. (2012) The Lin28 cold-shock domain remodels *pre-let-7* microRNA. *Nucleic Acids Research* **40**(15): 7492-7506.

Mazen, A., Menissier-de Murcia, J., Molinete, M., Simonin, F., Gradwohl, G., Poirier, G. and de Murcia, G. (1989) Poly(ADP-ribose)polymerase: a novel finger protein. *Nucleic Acids Research* **17**(12): 4689-4698.

McInerny, C. J., Kersey, P. J., Creanor, J. and Fantes, P. A. (1995) Positive and negative roles for *cdc10* in cell cycle gene expression. *Nucleic Acids Research* **23**(23): 4761-4768.

McManus, M. T., Adler, B. K., Pollard, V. W. and Hajduk, S. L. (2000) *Trypanosoma brucei* guide RNA poly(U) tail formation is stabilized by cognate mRNA. *Molecular and Cellular Biology* **20**(3): 883-891.

Mellman, D. L. and Anderson, R. A. (2009). A novel gene expression pathway regulated by nuclear phosphoinositides. *Advances in Enzyme Regulation, Vol 49*. Weber, G., Forrest Weber, C. E. and Cocco, L. **49:** 11-28.

Mellman, D. L., Gonzales, M. L., Song, C. H., Barlow, C. A., Wang, P., Kendziorski, C. and Anderson, R. A. (2008) A PtdIns4,5P(2)-regulated nuclear poly(A) polymerase controls expression of select mRNAs. *Nature* **451**(7181): 1013-1017.

Meneghini, M. D., Wu, M. and Madhani, H. D. (2003) Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112**(5): 725-736.

Militello, K. T. and Read, L. K. (2000) UTP-dependent and -independent pathways of mRNA turnover in *Trypanosoma brucei* mitochondria. *Molecular and Cellular Biology* **20**(7): 2308-2316.

Millar, A. A. and Waterhouse, P. M. (2005) Plant and animal microRNAs: similarities and differences. *Functional and Integrated Genetics* **5**(3): 129-135.

Minoda, Y., Saeki, K., Aki, D., Takaki, H., Sanada, T., Koga, K., Kobayashi, T., Takaesu, G. and Yoshimura, A. (2006) A novel zinc finger protein, ZCCHC11, interacts with TIFA and modulates TLR signaling. *Biochemical and Biophysical Research Communications* **344**(3): 1023-1030.

Molinete, M., Vermeulen, W., Burkle, A., Menissier-de Murcia, J., Kupper, J. H., Hoeijmakers, J. H. and de Murcia, G. (1993) Overproduction of the poly(ADP-ribose) polymerase DNA-binding domain blocks alkylation-induced DNA repair synthesis in mammalian cells. *EMBO Journal* **12**(5): 2109-2117.

Molnár, A., Schwach, F., Studholme, D. J., Thuenemann, E. C. and Baulcombe, D. C. (2007) miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature* **447**(7148): 1126-1129.

Montgomery, T. A., Rim, Y.-S., Zhang, C., Dowen, R. H., Phillips, C. M., Fischer, S. E. J. and Ruvkun, G. (2012) PIWI associated siRNAs and piRNAs specifically require the *Caenorhabditis elegans* HEN1 ortholog *henn-1*. *PLoS Genetics* **8**(4): e1002616.

Moreno, S., Klar, A., Nurse, P., Christine, G. and Gerald, R. F. (1991). Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods in Enzymology*, Academic Press. **Volume 194:** 795-823.

Morin, R. D., O'Connor, M. D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A. L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Eaves, C. J. and Marra, M. A. (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Research* **18**(4): 610-621.

Morozov, I. Y., Jones, M. G., Razak, A. A., Rigden, D. J. and Caddick, M. X. (2010) CUCU modification of mRNA promotes decapping and transcript degradation in *Aspergillus nidulans*. *Molecular and Cellular Biology* **30**(2): 460-469.

Morozov, I. Y., Jones, M. G., Spiller, D. G., Rigden, D. J., Dattenbock, C., Novotny, R., Strauss, J. and Caddick, M. X. (2010) Distinct roles for Caf1, Ccr4, Edc3 and CutA in the co-ordination of transcript deadenylation, decapping and P-body formation in Aspergillus nidulans. *Molecular Microbiology* **76**(2): 503-516.

Morris, T. D., Weber, L. A., Hickey, E., Stein, G. S. and Stein, J. L. (1991) Changes in the stability of a human H3 histone mRNA during the HeLa cell cycle. *Molecular and Cellular Biology* **11**(1): 544-553.

Morrissey, J. P., Deardorff, J. A., Hebron, C. and Sachs, A. B. (1999) Decapping of stabilized, polyadenylated mRNA in yeast *pab1* mutants. *Yeast* **15**(8): 687-702.

Motamedi, M. R., Verdel, A., Colmenares, S. U., Gerber, S. A., Gygi, S. P. and Moazed, D. (2004) Two RNAi Complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**(6): 789-802.

Muhlrad, D. and Parker, R. (2005) The yeast EDC1 mRNA undergoes deadenylation-independent decapping stimulated by Not2p, Not4p, and Not5p. *EMBO Journal* **24**(5): 1033-1045.

Mullen, T. E. and Marzluff, W. F. (2008) Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5 ' to 3 ' and 3 ' to 5 '. *Genes & Development* **22**(1): 50-65.

Müller, B., Blackburn, J., Feijoo, C., Zhao, X. and Smythe, C. (2007) DNA-activated protein kinase functions in a newly observed S phase checkpoint that links histone mRNA abundance with DNA replication. *Journal of Cell Biology* **179**(7): 1385-1398.

Munoz-Tello, P., Gabus, C. and Thore, S. (2012) Functional implications from the Cid1 poly(U) polymerase crystal structure. *Structure* **20**(6): 977-986.

Nag, A. and Jack, T. (2010) Sculpting the flower; the role of microRNAs in flower development. *Current Topics in Developmental Biology* **91**: 349-378.

Nagaike, T., Suzuki, T., Katoh, T. and Ueda, T. (2005) Human mitochondrial mRNAs are stabilized with polyadenylation regulated by mitochondria-specific poly(A) polymerase and polynucleotide phosphorylase. *Journal of Biological Chemistry* **280**(20): 19721-19727.

Nakamura, R., Takeuchi, R., Takata, K., Shimanouchi, K., Abe, Y., Kanai, Y., Ruike, T., Ihara, A. and Sakaguchi, K. (2008) TRF4 is involved in polyadenylation of snRNAs in *Drosophila melanogaster*. *Molecular and Cellular Biology* **28**(21): 6620-6631.

Nakanishi, T., Kubota, H., Ishibashi, N., Kumagai, S., Watanabe, H., Yamashita, M., Kashiwabara, S., Miyado, K. and Baba, T. (2006) Possible role of mouse poly(A) polymerase mGLD-2 during oocyte maturation. *Dev Biol* **289**(1): 115-126.

Nakanishi, T., Kumagai, S., Kimura, M., Watanabe, H., Sakurai, T., Kimura, M., Kashiwabara, S. and Baba, T. (2007) Disruption of mouse poly(A) polymerase mGLD-2 does not alter polyadenylation status in oocytes and somatic cells. *Biochem Biophys Res Commun* **364**(1): 14-19.

Nam, Y., Chen, C., Gregory, Richard I., Chou, James J. and Sliz, P. (2011) Molecular basis for interaction of *let-7* microRNAs with Lin28. *Cell* **147**(5): 1080-1091.

Nasmyth, K. and Nurse, P. (1981) Cell division cycle mutants altered in DNA replication and mitosis in the fission yeast *Schizosaccharomyces pombe*. *Molecular and General Genetics* **182**(1): 119-124.

Nayak, A., Goodfellow, I. G. and Belsham, G. J. (2005) Factors required for the uridylylation of the foot-and-mouth disease virus 3B1, 3B2, and 3B3 peptides by the RNA-dependent RNA polymerase ($3D^{pol}$) *in vitro*. *Journal of Virology* **79**(12): 7698-7706.

Nelson, M. J., Dang, Y., Filek, E., Zhang, Z., Yu, V. W., Ishida, K. and Green, B. R. (2007) Identification and transcription of transfer RNA genes in dinoflagellate plastid minicircles. *Gene* **392**(1-2): 291-298.

Newman, M. A., Mani, V. and Hammond, S. M. (2011) Deep sequencing of microRNA precursors reveals extensive 3' end modification. *RNA* **17**(10): 1795-1803.

Newman, M. A., Thomson, J. M. and Hammond, S. M. (2008) Lin-28 interaction with the *let-7* precursor loop mediates regulated microRNA processing. *RNA* **14**(8): 1539-1549.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D. and Grewal, S. I. (2004) RITS acts in *cis* to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nature Genetics* **36**(11): 1174-1180.

Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., Miyauchi, K. and Suzuki, T. (2007) The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nature: Structural and Molecular Biology* **14**(4): 349-350.

Oki, M., Ma, L., Wang, Y., Hatanaka, A., Miyazato, C., Tatebayashi, K., Nishitani, H., Uchida, H. and Nishimoto, T. (2007) Identification of novel suppressors for Mog1 implies its involvement in RNA metabolism, lipid metabolism and signal transduction. *Gene* **400**(1-2): 114-121.

Olsen, A., Vantipalli, M. C. and Lithgow, G. J. (2006) Checkpoint proteins control survival of the postmitotic cells in *Caenorhabditis elegans*. *Science* **312**(5778): 1381-1385.

Onishi, M., Koga, T., Hirata, A., Nakamura, T., Asakawa, H., Shimoda, C., Bähler, J., Wu, J. Q., Takegawa, K., Tachikawa, H., Pringle, J. R. and Fukui, Y. (2010) Role of septins in the orientation of forespore membrane extension during sporulation in fission yeast. *Molecular and Cellular Biology* **30**: 2057-2074.

Osorio, H., Carvalho, E., del Valle, M., Gunther Sillero, M. A., Moradas-Ferreira, P. and Sillero, A. (2003) H2O2, but not menadione, provokes a decrease in the ATP and an increase in the inosine levels

in *Saccharomyces cerevisiae*. An experimental and theoretical approach. *European Journal of Biochemistry* **270**(7): 1578-1589.

Page, R. D. M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357-358.

Pandey, N. B. and Marzluff, W. F. (1987) The stem-loop structure at the 3' end of histone mRNA is necessary and sufficient for regulation of histone mRNA stability. *Molecular and Cellular Biology* **7**(12): 4557-4559.

Panigrahi, A. K., Ernst, N. L., Domingo, G. J., Fleck, M., Salavati, R. and Stuart, K. D. (2006) Compositionally and functionally distinct editosomes in *Trypanosoma brucei*. *RNA* **12**(6): 1038-1049.

Panigrahi, A. K., Schnaufer, A., Ernst, N. L., Wang, B., Carmean, N., Salavati, R. and Stuart, K. (2003) Identification of novel components of *Trypanosoma brucei* editosomes. *RNA* **9**(4): 484-492.

Park, W., Li, J., Song, R., Messing, J. and Chen, X. (2002) CARPEL FACTORY, a dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Current Biology* **12**(17): 1484-1495.

Paul, A. V., Rieder, E., Kim, D. W., van Boom, J. H. and Wimmer, E. (2000) Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. *Journal of Virology* **74**(22): 10359-10370.

Paul, A. V., van Boom, J. H., Filippov, D. and Wimmer, E. (1998) Protein-primed RNA synthesis by purified poliovirus RNA polymerase. *Nature* **393**(6682): 280-284.

Paul, A. V., Yin, J., Mugavero, J., Rieder, E., Liu, Y. and Wimmer, E. (2003) A "slide-back" mechanism for the initiation of protein-primed RNA synthesis by the RNA polymerase of poliovirus. *Journal of Biological Chemistry* **278**: 43951-43960.

Pavelec, D. M., Lachowiec, J., Duchaine, T. F., Smith, H. E. and Kennedy, S. (2009) Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in *Caenorhabditis elegans*. *Genetics* **183**(4): 1283-1295.

Pélisson, A., Sarot, E., Payen-Groschêne, G. and Bucheton, A. (2007) A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense *gypsy* transcripts in somatic cells f the *Drosophila* ovary. *Journal of Virology* **81**(4): 1951-1960.

Peng, S., Chen, L. L., Lei, X. X., Yang, L., Lin, H., Carmichael, G. G. and Huang, Y. (2011) Genome-wide studies reveal that Lin28 enhances the translation of genes important for growth and survival of human embryonic stem cells. *Stem Cells* **29**(3): 496-504.

Petersen, T. N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**(10): 785-786.

Pfeffer, S., Zavolan, M., Grasser, F. A., Chien, M., Russo, J. J., Ju, J., John, B., Enright, A. J., Marks, D., Sander, C. and Tuschl, T. (2004) Identification of virus-encoded microRNAs. *Science* **304**(5671): 734-736.

Piskounova, E., Polytarchou, C., Thornton, J. E., LaPierre, R. J., Pothoulakis, C., Hagan, J. P., Iliopoulos, D. and Gregory, R. I. (2011) Lin28A and Lin28B inhibit *let-7* microRNA biogenesis by distinct mechanisms. *Cell* **147**(5): 1066-1079.

Polesskaya, A., Cuvellier, S., Naguibneva, I., Duquet, A., Moss, E. G. and Harel-Bellan, A. (2007) Lin-28 binds *IGF-2* mRNA and participates in skeletal myogenesis by increasing translation efficiency. *Genes and Development* **21**(9): 1125-1138.

Proudfoot, N. J., Furger, A. and Dye, M. J. (2002) Integrating mRNA processing with transcription. *Cell* **108**(4): 501-512.

Puustinen, P. and Mäkinen, K. (2004) Uridylylation of the potyvirus VPg by viral replicase NIb correlates with the nucleotide binding capacity of VPg. *Journal of Biological Chemistry* **279**(37): 38103-38110.

Qiu, C., Ma, Y., Wang, J., Peng, S. and Huang, Y. (2010) Lin28-mediated post-transcriptional regulation of Oct4 expression in human embryonic stem cells. *Nucleic Acids Research* **38**(4): 1240-1248.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. (2005) InterProScan: protein domains identifier. *Nucleic Acids Research* **33**(suppl 2): W116-W120.

Radford, H. E., Meijer, H. A. and de Moor, C. H. (2008) Translational control by cytoplasmic polyadenylation in *Xenopus* oocytes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1779**(4): 217-229.

Rammelt, C., Bilen, B., Zavolan, M. and Keller, W. (2011) PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *RNA* **17**(9): 1737-1746.

Read, R. L., Martinho, R. G., Wang, S. W., Carr, A. M. and Norbury, C. J. (2002) Cytoplasmic poly(A) polymerases mediate cellular responses to S phase arrest. *PNAS* **99**(19): 12079-12084.

Reddy, R., Henning, D., Das, G., Harless, M. and Wright, D. (1987) The capped U6 small nuclear RNA is transcribed by RNA polymerase III. *Journal of Biological Chemistry* **262**(1): 75-81.

Reimand, J., Arak, T. and Vilo, J. (2011) g:Profiler - a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Research* **39**: W307-W315.

Reimand, J., Kull, M., Peterson, H., Hansen, J. and Vilo, J. (2007) g:Profiler - a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research* **35**: W193-W200.

Ren, G., Chen, X. and Yu, B. (2012) Uridylation of miRNAs by *hen1 suppressor1* in *Arabidopsis*. *Current Biology* **22**(8): 695-700.

Reyes-Turcu, F. E., Zhang, K., Zofall, M., Chen, E. and Grewal, S. I. (2011) Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. *Nature: Structural and Molecular Biology* **18**(10): 1132-1138.

Rinke, J. and Steitz, J. A. (1985) Association of the lupus antigen La with a subset of U6 snRNA molecules. *Nucleic Acids Res* **13**(7): 2617-2629.

Rissland, O. S. (2008) RNA uridylation as a novel mechanism of gene regulation. D.Phil. in Pathology, University of Oxford.

Rissland, O. S., Mikulasova, A. and Norbury, C. J. (2007) Efficient RNA polyuridylation by noncanonical poly(A) polymerases. *Molecular and Cellular Biology* **27**(10): 3612-3624.

Rissland, O. S. and Norbury, C. J. (2008) The Cid1 poly(U) polymerase. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1779**(4): 286-294.

Rissland, O. S. and Norbury, C. J. (2009) Decapping is preceded by 3 ' uridylation in a novel pathway of bulk mRNA turnover. *Nature Structural and Molecular Biology* **16**(6): 616-624.

Robert, V. J. P., Sijen, T., van Wolfswinkel, J. and Plasterk, R. H. A. (2005) Chromatin and RNAi factors protect the *C. elegans* germline against repetitive sequences. *Genes & Development* **19**(7): 782-787.

Rohayem, J., Robel, I., Jager, K., Scheffler, U. and Rudolph, W. (2006) Protein-primed and de novo initiation of RNA synthesis by norovirus $3D^{pol}$. *Journal of Virology* **80**(14): 7060-7069.

Roth, K. M., Byam, J., Fang, F. and Butler, J. S. (2009) Regulation of NAB2 mRNA 3'-end formation requires the core exosome and the Trf4p component of the TRAMP complex. *RNA* **15**(6): 1045-1058.

Rott, R., Zipor, G., Portnoy, V., Liveanu, V. and Schuster, G. (2003) RNA polyadenylation and degradation in cyanobacteria are similar to the chloroplast but different from *Escherichia coli*. *Journal of Biological Chemistry* **278**(18): 15771-15777.

Rouhana, L., Wang, L., Buter, N., Kwak, J. E., Schiltz, C. A., Gonzalez, T., Kelley, A. E., Landry, C. F. and Wickens, M. (2005) Vertebrate GLD2 poly(A) polymerases in the germline and the brain. *RNA* **11**(7): 1117-1130.

Rouhana, L. and Wickens, M. (2007) Autoregulation of GLD-2 cytoplasmic poly(A) polymerase. *RNA* **13**(2): 188-199.

Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., Ge, H. and Bartel, D. P. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**(6): 1193-1207.

Rustici, G., Mata, J., Kivinen, K., Lio, P., Penkett, C. J., Burns, G., Hayles, J., Brazma, A., Nurse, P. and Bahler, J. (2004) Periodic gene expression program of the fission yeast cell cycle. *Nature Genetics* **36**(8): 809-817.

Ryan, Colm J., Roguev, A., Patrick, K., Xu, J., Jahari, H., Tong, Z., Beltrao, P., Shales, M., Qu, H., Collins, Sean R., Kliegman, Joseph I., Jiang, L., Kuo, D., Tosti, E., Kim, H.-S., Edelmann, W., Keogh, M.-C., Greene, D., Tang, C., Cunningham, P., Shokat, Kevan M., Cagney, G., Svensson, J. P., Guthrie, C., Espenshade, Peter J., Ideker, T. and Krogan, Nevan J. (2012) Hierarchical modularity and the evolution of genetic interactomes across species. *Molecular Cell* **46**(5): 691-704.

Ryan, C. M. and Read, L. K. (2005) UTP-dependent turnover of *Trypanosoma brucei* mitochondrial mRNA requires UTP polymerization and involves the RET1 TUTase. *RNA* **11**(5): 763-773.

Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E. E., Nitsch, R. and Wulczyn, F. G. (2008) A feedback loop comprising *lin-28* and *let-7* controls *pre-let-7* maturation during neural stem-cell commitment. *Nature: Cell Biology* **10**(8): 987-993.

Sadaie, M., Iida, T., Urano, T. and Nakayama, J. (2004) A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast. *EMBO Journal* **23**(19): 3825-3835.

Sadoff, B. U., Heath-Pagliuso, S., Castaño, I. B., Zhu, Y. F., Kieff, F. S. and Christman, M. F. (1995) Isolation of mutants of *Saccharomyces cerevisiae* requiring DNA topoisomerase I. *Genetics* **141**(2): 465-479.

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H. and Siomi, M. C. (2007) Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes and Development* **21**(13): 1603-1608.

Saitoh, S., Chabes, A., McDonald, W. H., Thelander, L., Yates, J. R. and Russell, P. (2002) Cid13 is a cytoplasmic poly(A) polymerase that regulates ribonucleotide reductase mRNA. *Cell* **109**(5): 563-573.

San Paolo, S., Vaňáčová, S., Schenk, L., Scherrer, T., Blank, D., Keller, W. and Gerber, A. P. (2009) Distinct roles of non-canonical poly(A) polymerases in RNA metabolism. *PLoS Genetics* **5**(7): e1000555.

Sanchez, R. and Marzluff, W. F. (2002) The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. *Molecular and Cellular Biology* **22**(20): 7093-7104.

Sashital, D. G. and Doudna, J. A. (2010) Structural insights into RNA interference. *Current Opinion in Structural Biology* **20**(1): 90-97.

Schmid, M., Kuchler, B. and Eckmann, C. R. (2009) Two conserved regulatory cytoplasmic poly(A) polymerases, GLD-4 and GLD-2, regulate meiotic progression in *C. elegans*. *Genes & Development* **23**(7): 824-836.

Schmidt, M. J., West, S. and Norbury, C. J. (2011) The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. *RNA* **17**(1): 39-44.

Schnaufer, A., Ernst, N. L., Palazzo, S. S., O'Rear, J., Salavati, R. and Stuart, K. (2003) Separate insertion and deletion subcomplexes of the *Trypanosoma brucei* RNA editing complex. *Molecular Cell* **12**(2): 307-319.

Seitz, H., Ghildiyal, M. and Zamore, P. D. (2008) Argonaute loading improves the 5' precision of both microRNAs and their miRNA* strands in flies. *Current Biology* **18**(2): 147-151.

Shanker, S., Job, G., George, O. L., Creamer, K. M., Shaban, A. and Partridge, J. F. (2010) Continuous requirement for the Clr4 complex but not RNAi for centromeric heterochromatin assembly in fission yeast harboring a disrupted RITS complex. *PLoS Genetics* **6**(10): e1001174.

Sharma, N., O'Donnell, B. J. and Flanegan, J. B. (2005) 3'-Terminal sequence in poliovirus negative-strand templates is the primary *cis*-acting element required for VPgpUpU-primed positive-strand initiation. *Journal of Virology* **79**(6): 3565-3577.

Shcherbik, N., Wang, M., Lapik, Y. R., Srivastava, L. and Pestov, D. G. (2010) Polyadenylation and degradation of incomplete RNA polymerase I transcripts in mammalian cells. *EMBO Reports* **11**(2): 106-111.

Shen, B. Z. and Goodman, H. M. (2004) Uridine addition after microRNA-directed cleavage. *Science* **306**(5698): 997-997.

Shin, C. Y., Kundel, M. and Wells, D. G. (2004) Rapid, activity-induced increase in tissue plasminogen activator is mediated by metabotropic glutamate receptor-dependent mRNA translation. *Journal of Neuroscience* **24**(42): 9425-9433.

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

Simon, B., Kirkpatrick, J. P., Eckhardt, S., Reuter, M., Rocha, E. A., Andrade-Navarro, M. A., Sehr, P., Pillai, R. S. and Carlomagno, T. (2011) Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein. *Structure* **19**(2): 172-180.

Slomovic, S., Portnoy, V., Yehudai-Resheff, S., Bronshtein, E. and Schuster, G. (2008) Polynucleotide phosphorylase and the archaeal exosome as poly(A)-polymerases. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **1779**(4): 247-255.

Song, M. G. and Kiledjian, M. (2007) 3' terminal oligo U-tract-mediated stimulation of decapping. *RNA* **13**(12): 2356-2365.

Souret, F. F., Kastenmayer, J. P. and Green, P. J. (2004) AtXRN4 degrades mRNA in *Arabidopsis* and its substrates include selected miRNA targets. *Molecular Cell* **15**(2): 173-183.

Stagno, J., Aphasizheva, I., Aphasizhev, R. and Luecke, H. (2007) Dual role of the RNA substrate in selectivity and catalysis by terminal uridylyl transferases. *PNAS* **104**(37): 14634-14639.

Stagno, J., Aphasizheva, I., Bruystens, J., Luecke, H. and Aphasizhev, R. (2010) Structure of the mitochondrial editosome-like complex associated TUTase 1 reveals divergent mechanisms of UTP selection and domain organization. *Journal of Molecular Biology* **399**(3): 464-475.

Stagno, J., Aphasizheva, I., Rosengarth, A., Luecke, H. and Aphasizhev, R. (2007) UTP-bound and apo structures of a minimal RNA uridylyltransferase. *Journal of Molecular Biology* **366**(3): 882-899.

Steil, B. P. and Barton, D. J. (2008) Poliovirus *cis*-acting replication element-dependent VPg uridylylation lowers the K_m of the initiating nucleoside triphosphate for viral RNA replication. *Journal of Virology* **82**(19): 9400-9408.

Stevenson, A. L. and Norbury, C. J. (2006) The Cid1 family of non-canonical poly(A) polymerases. *Yeast* **23**(13): 991-1000.

Suh, N., Jedamzik, B., Eckmann, C. R., Wickens, M. and Kimble, J. (2006) The GLD-2 poly(A) polymerase activates *gld-1* mRNA in the *Caenorhabditis elegans* germ line. *PNAS* **103**(41): 15108-15112.

Sullivan, K. D., Mullen, T. E., Marzluff, W. F. and Wagner, E. J. (2009) Knockdown of SLBP results in nuclear retention of histone mRNA. *RNA* **15**(3): 459-472.

Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A. and Mello, C. C. (1999) The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**(2): 123-132.

Takegami, T., Kuhn, R. J., Anderson, C. W. and Wimmer, E. (1983) Membrane-dependent uridylylation of the genome-linked protein VPg of poliovirus. *PNAS* **80**(24): 7447-7451.

Talsky, K. B. and Collins, K. (2010) Initiation by a eukaryotic RNA-dependent RNA polymerase requires looping of the template end and is influenced by the template-tailing activity of an associated uridyltransferase. *Journal of Biological Chemistry* **285**(36): 27614-27623.

Talsky, K. B. and Collins, K. (2012) Strand-asymmetric endogenous *Tetrahymena* small RNA production requires a previously uncharacterized uridylyltransferase protein partner. *RNA* **18**(8): 1553-1562.

Tapia-Alveal, C., Calonge, T. M. and O'Connell, M. J. (2009) Regulation of *chk1*. *Cell Division* **4**: 8.

Tazi, J., Forne, T., Jeanteur, P., Cathala, G. and Brunel, C. (1993) Mammalian U6 small nuclear RNA undergoes 3' end modifications within the spliceosome. *Molecular and Cellular Biology* **13**(3): 1641-1650.

Terns, M. P., Lund, E. and Dahlberg, J. E. (1992) 3'-end-dependent formation of U6 small nuclear ribonucleoprotein particles in *Xenopus laevis* oocyte nuclei. *Molecular and Cellular Biology* **12**(7): 3032-3040.

Thon, G., Hansen, K. R., Altes, S. P., Sidhu, D., Singh, G., Verhein-Hansen, J., Bonaduce, M. J. and Klar, A. J. (2005) The Clr7 and Clr8 directionality factors and the Pcu4 cullin mediate heterochromatin formation in the fission yeast *Schizosaccharomyces pombe*. *Genetics* **171**(4): 1583-1595.

Thon, G. and Verhein-Hansen, J. (2000) Four chromo-domain proteins of *Schizosaccharomyces pombe* differentially repress transcription at various chromosomal locations. *Genetics* **155**(2): 551-568.

Thornton, J. E., Chang, H.-M., Piskounova, E. and Gregory, R. I. (2012) Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7). *RNA*.

Thornton, J. E. and Gregory, R. I. (2012) How does Lin28 *let-7* control development and disease? *Trends in Cell Biology* **22**(9): 1-9.

Tian, Y., Simanshu, D. K., Ma, J. B. and Patel, D. J. (2011) Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *PNAS* **108**(3): 903-910.

Tkaczuk, K. L., Obarska, A. and Bujnicki, J. M. (2006) Molecular phylogenetics and comparative modeling of HEN1, a methyltransferase involved in plant microRNA biogenesis. *BMC Evolutionary Biology* **6**: 6.

Tomecki, R., Dmochowska, A., Gewartowski, K., Dziembowski, A. and Stepien, P. P. (2004) Identification of a novel human nuclear-encoded mitochondrial poly(A) polymerase. *Nucleic Acids Research* **32**(20): 6001-6014.

Tomecki, R., Drazkowska, K. and Dziembowski, A. (2010) Mechanisms of RNA degradation by the eukaryotic exosome. *Chembiochem* **11**(7): 938-945.

Trippe, R., Guschina, E., Hossbach, M., Urlaub, H., Lührmann, R. and Benecke, B.-J. (2006) Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. *RNA* **12**(8): 1494-1504.

Trippe, R., Richly, H. and Benecke, B.-J. (2003) Biochemical characterization of a U6 small nuclear RNA-specific terminal uridylyltransferase. *European Journal of Biochemistry* **270**(5): 971-980.

Trippe, R., Sandrock, B. and Benecke, B.-J. (1998) A highly specific terminal uridylyl transferase modifies the 3'-end of U6 small nuclear RNA. *Nucleic Acids Research* **26**(13): 3119-3126.

Tusnady, G. E. and Simon, I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**(9): 849-850.

Vadla, B., Kemper, K., Alaimo, J., Heine, C. and Moss, E. G. (2012) *lin-28* controls the succession of cell fate choices via two distinct activities. *PLoS Genetics* **8**(3): e1002588.

Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V. and Zamore, P. D. (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**(5785): 320-324.

van Wolfswinkel, J. C., Claycomb, J. M., Batista, P. J., Mello, C. C., Berezikov, E. and Ketting, R. F. (2009) CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. *Cell* **139**(1): 135-148.

Van Wynsberghe, P. M., Kai, Z. S., Massirer, K. B., Burton, V. H., Yeo, G. W. and Pasquinelli, A. E. (2011) LIN-28 co-transcriptionally binds primary *let-7* to regulate miRNA maturation in *Caenorhabditis elegans*. *Nature: Structural and Molecular Biology* **18**(3): 302-308.

Vaňáčová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G. and Keller, W. (2005) A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biology* **3**(6): 986-997.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I. and Moazed, D. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**(5658): 672-676.

Viswanathan, S. R., Daley, G. Q. and Gregory, R. I. (2008) Selective blockade of microRNA processing by Lin28. *Science* **320**(5872): 97-100.

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. and Martienssen, R. A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**(5588): 1833-1837.

Wang, G. and Reinke, V. (2008) A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Current Biology* **18**(12): 861-867.

Wang, L. T., Eckmann, C. R., Kadyk, L. C., Wickens, M. and Kimble, J. (2002) A regulatory cytoplasmic poly(A) polymerase in *Caenorhabditis elegans*. *Nature* **419**(6904): 312-316.

Wang, S.-W., Toda, T., MacCallum, R., Harris, A. L. and Norbury, C. (2000) Cid1, a fission yeast protein required for S-M checkpoint control when DNA polymerase delta or epsilon is inactivated. *Molecular and Cellular Biology* **20**(9): 3234-3244.

Wang, S. W., Norbury, C., Harris, A. L. and Toda, T. (1999) Caffeine can override the S-M checkpoint in fission yeast. *Journal of Cell Science* **112**(6): 927-937.

Wang, S. W., Stevenson, A. L., Kearsey, S. E., Watt, S. and Bahler, J. (2008) Global role for polyadenylation-assisted nuclear RNA degradation in posttranscriptional gene silencing. *Molecular and Cellular Biology* **28**(2): 656-665.

Wang, Y. and Morse, D. (2006) Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate *Lingulodinium*. *Nucleic Acids Research* **34**(2): 613-619.

Wang, Z. H., Castaño, I. B., De Las Penas, A., Adams, C. and Christman, M. F. (2000) Pol kappa: A DNA polymerase required for sister chromatid cohesion. *Science* **289**(5480): 774-779.

Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology* **337**: 635-645.

Watanabe, Y. and Yamamoto, M. (1994) *S. pombe mei2*$^{+}$ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species *meiRNA*. *Cell* **78**(3): 487-498.

Wei, K. F., Wu, L. J., Chen, J., Chen, Y. F. and Xie, D. X. (2012) Structural evolution and functional diversification analyses of argonaute protein. *J Cell Biochem* **113**(8): 2576-2585.

Wei, Y. F., Robins, P., Carter, K., Caldecott, K., Pappin, D. J., Yu, G. L., Wang, R. P., Shell, B. K., Nash, R. A., Schar, P. and et al. (1995) Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Molecular and Cellular Biology* **15**(6): 3206-3216.

Wery, M., Ruidant, S., Schillewaert, S., Lepore, N. and Lafontaine, D. L. J. (2009) The nuclear poly(A) polymerase and exosome cofactor Trf5 is recruited cotranscriptionally to nucleolar surveillance. *RNA* **15**(3): 406-419.

White, T. C. and Borst, P. (1987) RNA end-labeling and RNA ligase activities can produce a circular rRNA in whole cell extracts from trypanosomes. *Nucleic Acids Research* **15**(8): 3275-3290.

Will, C. L. and Lührmann, R. (2011) Spliceosome structure and function. *Cold Spring Harbour: Perspectives on Biology* **3**(7).

Wilusz, C. J. and Wilusz, J. (2008) New ways to meet your (3') end-oligouridylation as a step on the path to destruction. *Genes and Development* **22**(1): 1-7.

Win, T. Z., Draper, S., Read, R. L., Pearce, J., Norbury, C. J. and Wang, S. W. (2006) Requirement of fission yeast Cid14 in polyadenylation of rRNAs. *Molecular and Cellular Biology* **26**(5): 1710-1721.

Win, T. Z., Stevenson, A. L. and Wang, S. W. (2006) Fission yeast Cid12 has dual functions in chromosome segregation and checkpoint control. *Molecular and Cellular Biology* **26**(12): 4435-4447.

Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E. J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R. G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck,

A., Lehrach, H., Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S. J., Xiang, Z., Hunt, C., Moore, K., Hurst, S. M., Lucas, M., Rochet, M., Gaillardin, C., Tallada, V. A., Garzon, A., Thode, G., Daga, R. R., Cruzado, L., Jimenez, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J. L., Moreno, S., Armstrong, J., Forsburg, S. L., Cerutti, L., Lowe, T., McCombie, W. R., Paulsen, I., Potashkin, J., Shpakovski, G. V., Ussery, D., Barrell, B. G. and Nurse, P. (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**(6874): 871-880.

Wood, V., Harris, M. A., McDowall, M. D., Rutherford, K., Vaughan, B. W., Staines, D. M., Aslett, M., Lock, A., Bähler, J., Kersey, P. J. and Oliver, S. G. (2012) PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Research* **40**: D695-D699.

Woodland, H. R. and Pestell, R. Q. W. (1972) Determination of the nucleoside triphosphate contents of eggs and oocytes of *Xenopus laevis*. *Biochemistry Journal* **127**: 597-605.

Wu, H., Neilson, J. R., Kumar, P., Manocha, M., Shankar, P., Sharp, P. A. and Manjunath, N. (2007) miRNA profiling of naive, effector and memory CD8 T cells. *PLoS One* **2**(10): e1020.

Wu, L., Wells, D., Tay, J., Mendis, D., Abbott, M. A., Barnitt, A., Quinlan, E., Heynen, A., Fallon, J. R. and Richter, J. D. (1998) CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of alpha-CaMKII mRNA at synapses. *Neuron* **21**(5): 1129-1139.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J. C., Dufour, M. E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., Libri, D. and Jacquier, A. (2005) Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**(5): 725-737.

Wyman, S. K., Knouf, E. C., Parkin, R. K., Fritz, B. R., Lin, D. W., Dennis, L. M., Krouse, M. A., Webster, P. J. and Tewari, M. (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl

transferases contributes to miRNA transcriptome complexity. *Genome Research* **21**(9): 1450-1461.

Xu, B., Zhang, K. and Huang, Y. (2009) Lin28 modulates cell growth and associates with a subset of cell cycle regulator mRNAs in mouse embryonic stem cells. *RNA* **15**(3): 357-361.

Xu, B. S. and Huang, Y. Q. (2009) Histone H2a mRNA interacts with Lin28 and contains a Lin28-dependent posttranscriptional regulatory element. *Nucleic Acids Research* **37**(13): 4256-4263.

Yamada, T., Fischle, W., Sugiyama, T., Allis, C. D. and Grewal, S. I. (2005) The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. *Molecular Cell* **20**(2): 173-185.

Yamamoto, M. (1996) Regulation of meiosis in fission yeast. *Cell Structure and Function* **21**(5): 431-436.

Yang, X. C., Purdy, M., Marzluff, W. F. and Dominski, Z. (2006) Characterization of 3'hExo, a 3' exonuclease specifically interacting with the 3' end of histone mRNA. *Journal of Biological Chemistry* **281**(41): 30447-30454.

Yang, Z., Ebright, Y. W., Yu, B. and Chen, X. (2006) HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Research* **34**(2): 667-675.

Yates, L. A., Fleurdépine, S., Rissland, O. S., De Colibus, L., Harlos, K., Norbury, C. J. and Gilbert, R. J. C. (2012) Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nature Structural and Molecular Biology* **19**(8): 782-787.

Yogo, Y. and Wimmer, E. (1972) Polyadenylic acid at the 3'-terminus of poliovirus RNA. *PNAS* **69**(7): 1877-1882.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., Steward, R. and Chen, X. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**(5711): 932-935.

Yu, J. Y., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, II and Thomson, J. A. (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**(5858): 1917-1920.

Zabel, P., Dorssers, L., Wernars, K. and Van Kammen, A. (1981) Terminal uridylyl transferase of *Vigna unguiculata*: purification and characterization of an enzyme catalyzing the addition of a single UMP residue to the 3'-end of an RNA primer. *Nucleic Acids Research* **9**(11): 2433-2453.

Zaratiegui, M., Castel, S. E., Irvine, D. V., Kloc, A., Ren, J., Li, F., de Castro, E., Marin, L., Chang, A. Y., Goto, D., Cande, W. Z., Antequera, F., Arcangioli, B. and Martienssen, R. A. (2011) RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II. *Nature* **479**(7371): 135-138.

Zhang, K., Fischer, T., Porter, R. L., Dhakshnamoorthy, J., Zofall, M., Zhou, M., Veenstra, T. and Grewal, S. I. (2011) Clr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA. *Science* **331**(6024): 1624-1627.

Zhang, Z., Green, B. R. and Cavalier-Smith, T. (1999) Single gene circles in dinoflagellate chloroplast genomes. *Nature* **400**(6740): 155-159.

Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T. T., Meyers, B. C., Mo, B. and Chen, X. (2012) The *Arabidopsis* nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Current Biology* **22**(8): 689-694.

Zofall, M., Fischer, T., Zhang, K., Zhou, M., Cui, B., Veenstra, T. D. and Grewal, S. I. (2009) Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. *Nature* **461**(7262): 419-422.

Chapter 8
Appendices

Appendix A: *Schizosaccharomyces pombe* strains used in this study

Strain	Genotype	Source
CN520	*h$^+$ 972 (wt)*	Hayatu Raji
CN521	*h$^-$ 972 (wt)*	Hayatu Raji
CN624	*h$^+$ leu1-32 ura4-D18 ade6-M210*	This study
CN627	*h$^-$ leu1-32 ura4-D18 ade6-M210*	This study
CN514	*h$^+$ lsm1::kanMX cid1::LEU2 leu1-32*	This study
CN533	*h$^-$ lsm1::kanMX cid1::LEU2 leu1-32*	This study
CN601	*h$^-$ lsm1::kanMX cid1::LEU2 cid11::ura4$^+$ leu1-32 ura4-D18*	This study
CN581	*h$^-$ lsm1::kanMX cid1::LEU2 cid12::ura4$^+$ leu1-32 ura4-D18*	This study
CN600	*h$^-$ lsm1::kanMX cid1::LEU2 cid13::ura4$^+$ leu1-32 ura4-D18 ade6-(M210/M216)*	This study
CN597	*h$^-$ lsm1::kanMX cid1::LEU2 cid14::ura4$^+$ leu1-32 ura4-D18*	This study
CN598	*h$^+$ lsm1::kanMX cid1::LEU2 cid14::ura4$^+$ leu1-32 ura4-D18 ade6-M210*	This study
CN584	*h$^+$ lsm1::kanMX cid1::LEU2 cid16::ura4$^+$ leu1-32 ura4-D18 his7-366*	This study
CN632	*h$^+$ cid1::kanMX*	This study
CN711	*h$^-$ cid1::LEU2 cid12::ura4$^+$ leu1-32 ura4-D18*	This study
CN622	*h$^-$ cid1::ura4$^+$ cid13::LEU2 leu1-32 ura4-D18*	Hayatu Raji
CN770	*h$^?$ cid1::LEU2 cid14::ura4$^+$ leu1-32 ura4-D18*	This study
CN749	*h$^+$ cid1::kanMX cid16::ura4$^+$ leu1-32 ura4-D18*	This study
CN635	*h$^+$ cid1::kanMX leu1-32*	This study
CN759	*h$^+$ cid1::kanMX leu1-32 (pREP1-NTAP)*	This study
CN649	*h$^+$ cid1::kanMX leu1-32(pREP1-NTAP-cid1)*	This study
CN768	*h$^+$ cid1::kanMX leu1-32 (pREP1-NTAP-cid11 (gDNA sequence))*	This study
CN772	*h$^+$ cid1::kanMX leu1-32 (pREP1-NTAP-cid11 (cDNA sequence))*	This study
CN760	*h$^+$ cid1::kanMX leu1-32 (pREP1-NTAP-cid12)*	This study
CN761	*h$^+$ cid1::kanMX leu1-32 (pREP1-NTAP-cid13)*	This study
CN762	*h$^+$ cid1::kanMX leu1-32 (pREP1-NTAP-cid14)*	This study
CN763	*h$^+$ cid1::kanMX leu1-32 (pREP1-NTAP-cid16)*	This study
CN766	*h$^-$ cid16::kanMX*	This study
CN767	*h$^+$ cid16::kanMX*	This study
CN769	*h$^-$ cid1::kanMX cid16::kanMX*	This study
CN774	*h$^+$ cid1::kanMX leu1-32 (pREP1-NTAP-cid16-C371)*	This study
CN776	*h$^+$ cid1::kanMX leu1-32 (pREP1-NTAP-cid16-C330)*	This study
CN771	*h$^+$ SPAC13F5.07c::kanMX leu1-32 ura4-D18 ade6-(M210/M216)*	Tim Humphrey
CN778	*h$^+$ SPAC13F5.07c::kanMX leu1-32 ura4-D18 ade6-(M210/M216) (pREP1-NTAP-cid16)*	This study
CN716	*h$^+$ ago1::kanMX*	Monika Gullerova

Strain	Genotype	Source
CN721	*h⁻ ago1::kanMX cid1::kanMX*	This study
CN718	*h⁺ dcr1::kanMX*	Monika Gullerova
CN736	*h⁺ dcr1::kanMX cid1::kanMX*	This study
CN730	*h⁺ h2a.z::kanMX otr1R::ura4⁺ leu1-32 ura4-DS/E ade6-M210 his2⁻*	Shiv Grewal
CN733	*h⁺ h2a.z::kanMX cid1::kanMX leu1-32 ura4-(DSE/D18) ade6-M210*	This study
CN740	*h⁺ eri1::kanMX otr1R::ura4⁺ leu1-32 ura4-DS/E ade6-M210*	Jun-ichi Nakayama
CN744	*h⁻ eri1::hphMX cid1::kanMX otr1R::ura4⁺*	This study
CN741	*h⁺ eri1::hphMX rdp1::kanMX otr1R::ura4⁺ leu1-32 ura4-DS/E ade6-M210*	Jun-ichi Nakayama
CN752	*h⁻ rdp1::kanMX leu1-32 ura4-D18*	This study
CN743	*h⁻ eri1::hphMX*	This study

Appendix B: Vectors used in this study

Code	Plasmid Details	Restriction Enzymes	Source
CN417	pFA6α-kanMX6	N/A	(Bahler *et al.* 1998)
CN407	pREP1-NTAP	N/A	(Maundrell 1993)
CN318	pREP1-NTAP-tCid1	N/A	Olivia Rissland
CN598	pREP1-NTAP-Cid11(gDNA)	SalI, NotI	This study
CN599	pREP1-NTAP-Cid11(cDNA)	SalI, NotI	This study
CN600	pREP1-NTAP-Cid12	BamHI, NotI	This study
CN601	pREP1-NTAP-Cid13	NdeI, NotI	This study
CN602	pREP1-NTAP-Cid14	BamHI, NotI	This study
CN603	pREP1-NTAP-Cid16	NdeI, NotI	This study
CN604	pREP1-NTAP-Cid16-C371	SalI, NotI	This study
CN605	pREP1-NTAP-Cid16-C330	SalI, NotI	This study
CN505	pGEX6P-1-tCid1	N/A	Olivia Rissland
CN608	pGEX6P-1-tCid1-H336K	N/A	This study
CN609	pGEX6P-1-tCid1-H336R	N/A	This study
CN610	pGEX6P-1-tCid1-H336N	N/A	This study
CN611	pGEX6P-1-tCid1-C159V	N/A	This study
CN612	pGEX6P-1-tCid1-N171T	N/A	This study
CN613	pGEX6P-1-tCid1-A196L	N/A	This study
CN614	pGEX6P-1-tCid1-P229R	N/A	This study
CN615	pGEX6P-1-tCid1-N337D	N/A	This study
CN616	pGEX6P-1-tCid1-N171T/H336N/N337D	N/A	This study
CN617	pGEX6P-1-tCid1-R340N	N/A	This study
CN618	pGEX6P-1-tCid1-H336R/R340N	N/A	This study

Appendix C: Primers used in this study

Code	Primer Name	Sequence (5'→3')	Purpose
DS1	Lsm1 5' flank	GACGACCTCAGAGCACTTGC	Genotype testing
DS2	Lsm1 3' flank	GGGAATAAACAGCGGTCGAG	Genotype testing
DS17	kanMX internal rev	CGAAACGTGAGTCTTTTCC	Genotype testing
DS47	cid1 5' flank	GCCACTTTCACCTTTTGAGC	Genotype testing
DS48	cid1 int rev	CCAAACTTCCAAAGGCTACC	Genotype testing
DS49	cid11 5' flank	CCAAGCACACAACTCTACACAC	Genotype testing
DS50	cid11 int rev	AGCCAAATTGTTCTCAGTGG	Genotype testing
DS51	cid12 5' flank	TTACAAGGCACTCGCACG	Genotype testing
DS52	cid12 int rev	GGACTCTGGCACTAGAATGG	Genotype testing
DS53	cid13 5' flank	ACTCACGCCTTTACTAGCTGAC	Genotype testing
DS54	cid13 int rev	CTGTTGAAATGCAAACGACC	Genotype testing
DS55	cid14 5' flank	CCCCAAATTGAATTACCGAG	Genotype testing
DS56	cid14 int rev	TGCTTCCTCTGGTAAATCAGG	Genotype testing
DS57	cid16 5' flank	CGTTTAAAAGCGGCATACC	Genotype testing
DS58	cid16 int rev	AGCTAACATCGTCATCCAGTG	Genotype testing
DS59	LEU2 int rev	GTCACCTGGCAAAACGACG	Genotype testing
DS66	cid1 3' flank	GTTGTACATAAAACAAAGGGGG	Genotype testing
DS67	cid11 3' flank	TACTTCCTTGGTGCATTGC	Genotype testing
DS68	cid12 3' flank	GAAACATTACCACATGCGG	Genotype testing
DS69	cid13 3' flank	CGTCACCAAACCTTCAGC	Genotype testing
DS70	cid14 3' flank	ACTGGTATTCGAGTTTACTGGC	Genotype testing
DS71	cid16 3' flank	TTTGAAAGTGAGTCAGAGAGGG	Genotype testing
DS72	ura4^{+} 3' end fwd	CGTAATCCTGTTGTCGAAGC	Genotype testing
DS73	kanMX 3' end fwd	CGTATGTGAATGCTGGTCG	Genotype testing
DS74	chk1 958F qPCR	TTCAATGACACAGCCAGCTT	Microarray validation
DS75	chk1 1069R qPCR	TTCAAGGATTTCAGTAACGTCAA	Microarray validation
DS76	C188.07 1982F qPCR	CTGCTGAGCAAAGCAAAGAG	Microarray validation
DS77	C188.07 2102R qPCR	CCAGAGAAGAGACCTTAGCGA	Microarray validation
DS78	spt3 642F qPCR	TGGCTTCTTGACCTTTGAGA	Microarray validation
DS79	spt3 731R qPCR	TGCATGAGAGCCTCCACTAC	Microarray validation
DS80	PB10D8.02c 1129F qPCR	CCTTCTCGACTTAGCAAGGG	Microarray validation

Appendix C (continued): Primers used in this study

Code	Primer Name	Sequence (5'→3')	Purpose
DS81	PB10D8.02c 1253R qPCR	TAACTCAAGGATGGTTGGCA	Microarray validation
DS82	mei2 2170F qPCR	CCCAATAATGCACGTCGTAA	Microarray validation
DS83	mei2 2231R qPCR	ACATTTGCTTGCAGTTGGAG	Microarray validation
DS84	mfm1 18F qPCR	CTCCGTTTCTTCCTCCTCTG	Microarray validation
DS85	mfm1 106R qPCR	TGCAATGACACACATGTAAGG	Microarray validation
DS86	isp5 1641F qPCR	CTTGGATACTGGCTTCAGCA	Microarray validation
DS87	isp5 1719R qPCR	CGCAGAAAGATAGGACGGAT	Microarray validation
DS88	ptr2 1328F qPCR	CCATGATTTACGCTGCTGTT	Microarray validation
DS89	ptr2 1431R qPCR	AAACGTAAGCAGGGATTTGG	Microarray validation
DS90	tlh 5518F qPCR	GCCGACAGTGATATGGCTTA	Microarray validation
DS91	tlh 5611R qPCR	CTGCAATAGTCCTTCGGTATTG	Microarray validation
DS92	cox1 mito 730F qPCR	TTGGTCATCCAGAGGTTTATATTT	Microarray validation
DS93	cox1 mito 825R qPCR	CATAGCCCAAAGCATACCTTC	Microarray validation
DS94	MX6 int rev	ATGGGCTAAATGTACGGGC	Microarray validation
DS96	cid1-MX6 5'	TAAGTTTGCCACTTTCACCTTTTGAGCAGTAGACGAATATGAACATTTCTTCTGCACAATTTATTCCTGG TGTTCACACCGGATCCCCGGGTTAATTAA	HR transformation
DS97	cid1-MX6 3'	TGATTACTATGAATCTAGGGGAATGAATATTGATGCGCTTTTTAAGACTTCTTTGTAAAATCACTCAGA ATTGTCACCAGAATTCGAGCTCGTTTAAAC	HR transformation
DS122	h2a.z 5' flank fw	GATGTTTAAAGGTACAGCCCC	Genotype testing
DS123	h2a.z 3' flank rev	GATGGAGGGAAGAGCATTTC	Genotype testing
DS126	h2a.z internal rev	AAGGCAGATTCGTGAACG	Genotype testing
DS127	h2a.z internal fw	GAAAAATATCCCGAGGAGG	Genotype testing
DS158	5' Rqh2 Northern	TTCAGAAATTGCTAAAGTCGC	Northern probe
DS159	3' Rqh2 Northern	GTCAGTATAATTTGATGATGG	Northern probe
DS166	cid1 3' end fw	GATTCATTATTTGAGGAGGCC	Genotype testing
DS170	cid12 5' BamHI inframe	GGGGGATCCCATGGGTAAAGTCCTGTT	Cid family cloning
DS171	cid12 3' NotI inframe	GGAGCGGCCGCTTATCCGCCAGCTTG	Cid family cloning
DS173	cid13 3' NotI inframe	GGGGCGGCCGCTCAACTCTTGAATGCTT	Cid family cloning
DS174	cid14 5' BamHI inframe	GGGGGATCCCATGGGTAAAAAAAGCG	Cid family cloning
DS175	cid14 3' NotI inframe	GGAGCGGCCGCCTAAAAACGTTTGCG	Cid family cloning
DS177	cid16 3' NotI inframe	GGGGCGGCCGCTTATTGAATCAAGGGATC	Cid family cloning

Appendix C (continued): Primers used in this study

Code	Primer Name	Sequence (5'→3')	Purpose
DS178	cid13 5' NdeI inframe	GGGCATATGTCGATGGACAACGCTAATTG	Cid family cloning
DS179	Cid16 5' NdeI inframe	GGCCATATGTCGATGCTATTTGCCAAAT	Cid family cloning
DS180	pREP1 nmt1 pro seq fw	ATTCGGCAATGTGCAGC	Clone sequencing
DS181	pREP1 nmt1 term seq rev	GTAATATGCAGCTTGAATGGG	Clone sequencing
DS182	cid11 3' end fw	ATTACCCTATCCGATGGAGG	Genotype testing
DS183	cid12 3' end fw	CACCTTTGCAATTACGGAAG	Genotype testing
DS184	cid13 3' end fwd	GGTCCATTCCAGTACTATCCTG	Genotype testing
DS185	cid14 3' end fwd	CAAGCTTTGACATCTACGGG	Genotype testing
DS186	cid16 3' end fw	AACTGAATTTAAGGATTGCCC	Genotype testing
DS190	cid1 N171T fw	CAATCGTCTAGCTATTCATACTACGCTTTTACTTTCTTC	cid1 mutagenesis
DS191	cid1 N171T rev	GAAGAAAGTAAAAGCGTAGTATGAATAGCTAGACGATTG	cid1 mutagenesis
DS192	cid1 H336K fw	GATCCTTTCGAGATTTCAAAGAATGTGGGTAGGACAG	cid1 mutagenesis
DS193	cid1 H336K rev	CTGTCCTACCCACATTCTTTGAAATCTCGAAAGGATC	cid1 mutagenesis
DS194	cid1 H336N fw	GATCCTTTCGAGATTTCAAATAATGTGGGTAGGACAG	cid1 mutagenesis
DS195	cid1 H336N rev	CTGTCCTACCCACATTATTTGAAATCTCGAAAGGATC	cid1 mutagenesis
DS196	cid1 H336R fw	CCTTTCGAGATTTCACGTAATGTGGGTAGGACAG	cid1 mutagenesis
DS197	cid1 H336R rev	CTGTCCTACCCACATTACGTGAAATCTCGAAAGG	cid1 mutagenesis
DS198	NTAP qPCR fw	CAAGCACCGAAAGCTGATAA	Overexpression qPCR
DS199	NTAP qPCR rev	CTTTGGCTTGGGTCATCTTTT	Overexpression qPCR
DS200	cid12 qPCR fw	ACGACATTGGAAAGCAATCA	Overexpression qPCR
DS201	cid12 qPCR rev	CCGTAATTGCAAAGGTGCTA	Overexpression qPCR
DS202	cid13 qPCR fw	AATGTCGACGATGCTTGTTC	Overexpression qPCR
DS203	cid13 qPCR rev	TACAGCTGGTGTCCGAAGAG	Overexpression qPCR
DS204	cid14 qPCR fw	GAAACAGATGCTGCAAAGGA	Overexpression qPCR
DS205	cid14 qPCR rev	TGCTCTCCCGTAGATGTCA	Overexpression qPCR
DS206	cid16 qPCR fw	CTGGGCTTCGAATAGACTCA	Overexpression qPCR
DS207	cid16 qPCR rev	GGTTCCGGAATCTGTTGTAGA	Overexpression qPCR
DS208	cid1 qPCR fw	GGTAGCCTTTGGAAGTTTGG	Overexpression qPCR
DS209	cid1 qPCR rev	TTTCCTTCAAATCCTTCAGCTA	Overexpression qPCR
DS212	cid1 H336N-N337D fwd	GATCCTTTCGAGATTTCAAATGATGTGGGTAGGACAGTTAGC	cid1 mutagenesis
DS213	cid1 H336N-N337D rev	GCTAACTGTCCTACCCACATCATTTGAAATCTCGAAAGGATC	cid1 mutagenesis

Appendix C (continued): Primers used in this study

Code	Primer Name	Sequence (5'→3')	Purpose
DS214	cid1 N337D fwd	CCTTTTCGAGATTTCACATGATGTGGGTAGGACAGTTAGC	cid1 mutagenesis
DS215	cid1 N337D rev	GCTAACTGTCCTACCCACATCATGTGAAATCTCGAAAGG	cid1 mutagenesis
DS216	cid1 A196L fwd	CCTTCTTGTTAAGCATTGGCTCAAACGGAAGCAAATC	cid1 mutagenesis
DS217	cid1 A196L rev	GATTTGCTTCCGTTTGAGCCAATGCTTAACAAGAAGG	cid1 mutagenesis
DS218	cid1 P229R fwd	CACGTTATCAAGCCTCGCGTCTTTCCTAATTTACTG	cid1 mutagenesis
DS219	cid1 P229R rev	CAGTAAATTAGGAAAGACGCGAGGCTTGATAACGTG	cid1 mutagenesis
DS220	cid1 C159V fwd	GGGGCTTCGTTTCAAGTTGATATTGGATTTAACAATCGTC	cid1 mutagenesis
DS221	cid1 C159V rev	GACGATTGTTAAATCCAATATCAACTTGAAACGAAGCCCC	cid1 mutagenesis
DS225	cid1 H336R R340N fwd	CTTTCGAGATTTCACGTAATGTGGGTAACACAGTTAGCAGTTCTG	cid1 mutagenesis
DS226	cid1 H336R R340N rev	CAGAACTGCTAACTGTGTTACCCACATTACGTGAAATCTCGAAAG	cid1 mutagenesis
DS227	cid1 R340N fwd	CACATAATGTGGGTAACACAGTTAGCAGTTCTGGATTGTATCG	cid1 mutagenesis
DS228	cid1 R340N rev	CGATACAATCCAGAACTGCTAACTGTGTTACCCACATTATGTG	cid1 mutagenesis
DS230	cid16-C371 SalI fwd	CCCCCCGTCGACTCTTCCTAGAGGAAATTGGC	cid16 truncations
DS231	cid16-C330 SalI fwd	CCCCCCGTCGACTCTTTTTTGAGAACGTACCTTG	cid16 truncations
DS234	Oligo-dC15	CCCCCCCCCCCCCCC	UTP specificity test
DS235	Pseudo-dG23	GGGGGAGGGGGAGGGGGAGGGGG	UTP specificity test
DS239	cid11 SalI fwd-2	TGGTGGGTCGACTATGGAGTTATTGACATTTACTTG	Cid family cloning
DS240	cid11 NotI rev-2	AAAAAAGCGGCCGCGAATTAATTTCCGGACGCAG	Cid family cloning
DS241	cid14 BamHI fwd pGEX	GGGGGATCCGAGATGGGTAAAAAAAGCG	Cid family cloning
DS242	cid12 BamHI fwd pGEX	GGGGGATCCAGAATGGGTAAAGTCCTGTTAG	Cid family cloning
DS243	pREP1-NTAP MCS fw	GAATTTCATAGCCGTCTCAGC	Clone sequencing
DS244	cid11 5' end fwd	GCTGCGCACTATTTTAAGTACC	Cid family cloning
DS245	cid11 3' end rev	ATTGCAGGTAGGATTGGGG	Cid family cloning
DS250	cid11 SalI fwd pGEX	TGGTGGGTCGACATATGGAGTTATTGACATTTACTTG	Cid family cloning
DS255	qPCR dcr1 fwd	AATTTGGGTCTCCAAAGCAC	Microarray validation
DS256	qPCR dcr1 rev	AAGCTTCTTTCAGGCCAAAG	Microarray validation
DS257	qPCR clr4 fwd	ATTTGATTACGCTGGTGCAA	Microarray validation
DS258	qPCR clr4 rev	AGCCAGCCACGACAATTAG	Microarray validation
DS259	qPCR ctr4 fwd	ACACCCATGTCTGGAATGAA	Microarray validation
DS260	qPCR ctr4 rev	CCTTTGCGGTTGTAGTGGTA	Microarray validation
DS261	qPCR dmc1 fwd	AAGCCATTGCAGAACGATTT	Microarray validation

Appendix C (continued): Primers used in this study

Code	Primer Name	Sequence (5'→3')	Purpose
DS262	qPCR dmc1 rev	AAACGGTATTGGCCATCTTC	Microarray validation
DS263	qPCR isp3 fwd	CCCAAATGCTGTGAGAAGAA	Microarray validation
DS264	qPCR isp3 rev	CATCACCAAAGCTACCCAGA	Microarray validation
DS265	qPCR meu6 fwd	ACGTCTAAGAAAGCCGGAAA	Microarray validation
DS266	qPCR meu6 rev	GGGAAGGATCTGCCTGTAAA	Microarray validation
DS267	qPCR meu14 fwd	CCTCAGTGGAATTCTCGGAT	Microarray validation
DS268	qPCR meu14 rev	AGCTGGGCCTTGTAAGATTC	Microarray validation
DS269	qPCR mfm2 fwd	CTAGGGAATACAAACTTGTAAAGTA	Microarray validation
DS270	qPCR mfm2 rev	AACGATGCAATTAGAAATCACAA	Microarray validation
DS271	qPCR pho1 fwd	CGAGGAGGTATTCCCTTTGA	Microarray validation
DS272	qPCR pho1 rev	ACGAACGGGAGAATTCAAGT	Microarray validation
DS273	qPCR rep2 fwd	CTTCTGATGCTCGTGATGCT	Microarray validation
DS274	qPCR rep2 rev	TAAGAGCCGTTGAATGGTTG	Microarray validation
DS275	qPCR spn6 fwd	AAGCTTCTTTCAGGCCAAAG	Microarray validation
DS276	qPCR spn6 rev	AATTTGGGTCTCCAAAGCAC	Microarray validation
DS277	cid11 exon1 XhoI rev	GGGGGGCTCGAGTGGCATGAAACAAATTGTCTG	cid11 cDNA cloning
DS278	cid11 exon2 XhoI fwd	CCCCCCCTCGAGTAAAGTAATTGAGCCAATTG	cid11 cDNA cloning
DS285	cid11 intron flank fwd	CTCAACATGTCAATTGGCTC	Genotype testing
DS286	cid11 intron flank rev	CAGACAATTTGTTTCATGCC	Genotype testing
LP1	ura4-kanR 5'	TCAGCATTCTTTCTCTAAATAGGAATTTGTTACTTAATGGAGAAAAAAATGTTTCGATTTACCTAGTGTA TTTGTTTGTACGGATCCCCGGGTTAATTAA	HR transformation
LP2	ura4-kanR 3'	CCAACCAAAAAAATTTTACATTAGTCTTTTTTTAATGCTGAGAAAGTCTTTGCTGATATGCCTTCCAACC AGCTTCTCTAGAATTCGAGCTCGTTTAAAC	HR transformation
MJ103	pik1_1861F	TTAGTGACCGGCGACAATAG	qPCR normalisation
MJ104	pik1_1974R	TCCTTGGGCTAGTTCAGCTT	qPCR normalisation
OR114	M13 reverse	CAGGAAACAGCTATGAC	cRACE sequencing
OR115	M13 forward	GTAAACGACGGCCAG	cRACE sequencing
OR126	3' T7 actin	CCACTATGTATCCCGGTATTGC	act1 cRACE
OR128	Seq actin	GCTCCTCTTACTTTTTTGTAACG	act1 cRACE
OR243	Upstream 5' T7 actin	TAATACGACTCACTATAGGGACTCAAAGTCCAAAGCGAC	act1 cRACE
OR312	Actin cRACE upstream	GAGGGGAATACAGCTCTAG	act1 cRACE

Appendix C (continued): Primers used in this study

Code	Primer Name	Sequence (5'→3')	Purpose
OR244	Oligo$(dT)_{20}$	TTTTTTTTTTTTTTTTTTTT	cDNA synthesis
SF31	pGexFwd	GGGCTGGCAAGCCACGTTTGGTG	Clone sequencing
SF32	pGexRev	CCGGGAGCTGCATGTGTCAGAGG	Clone sequencing

Appendix D: *act1* cRACE sequences retrieved in this study

Appendix D.1: act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ *cells*

Clone	5' End	3' End	Tail
DS0001	***TTACCGTTTACC… (-59)***	***…CTCCCGTTAA (+1553)***	***A_6U_3***
DS0002	***TTACCGTTTACC… (-59)***	***…CTCCCGTTAA (+1553)***	***A_8U***
DS0003	***TTACCGTTTACC… (-59)***	***…GGTATCAATT (+1271)***	***U***
DS0004	***TTACCGTTTACC… (-59)***	***…TCTTCTGATA (+1189)***	***A_8U_2***
DS0005	***TACCGTTTACCA… (-58)***	***…ACCTGAACTT (+1695)***	***–***
DS0006	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{19}***
DS0007	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{15}***
DS0008	ACCGTTTACCAA… (-57)	…TGGT ATATGA (+1792)	A_9
DS0009	ACCGTTTACCAA… (-57)	…CATGGTATAT (+1790)	A_3
DS0010	ACCGTTTACCAA… (-57)	...AAGATTTAAT (+1774)	–
DS0011	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{17}
DS0012	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_7
DS0013	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_6
DS0014	ACCGTTTACCAA… (-57)	…TTTCTTTAAC (+1442)	–
DS0015	ACCGTTTACCAA… (-57)	…AATATTTCCA (+1389)	–
DS0016*	ACCGTTTACCAA… (-57)	…AGCCTTTTAC (+1363)	–
DS0017	ACCGTTTACCAA… (-57)	…CTTTGGGTAT (+1266)	–
DS0018	ACCGTTTACCAA… (-57)	…ACCTTTTTCC (+1234)	–
DS0020	ACCGTTTACCAA… (-57)	…TTTTTACCTA (+1217)	A_{23}
DS0021	ACCGTTTACCAA… (-57)	…TTTTTTACCT (+1216)	A_{23}
DS0022*	ACCGTTTACCAA… (-57)	…TATAAATTTC (+1201)	–
DS0023†	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	CA_{20}
DS0024	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{17}
DS0025	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_7
DS0026	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_6
DS0027*	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_4
DS0028	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	–
DS0029	ACCGTTTACCAA… (-57)	…TCTGATATAT (+1192)	–
DS0030*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{18}
DS0031	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{16}
DS0032	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{13}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.1 (continued): act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ *cells*

Clone	5' End	3' End	Tail
DS0033	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{34}
DS0034	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{33}
DS0035†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_6GA_{26}
DS0036	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{30}
DS0037*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{22}
DS0038	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{20}
DS0039	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{18}
DS0040	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{14}
DS0041	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{13}
DS0042	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{12}
DS0043	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_6
DS0044	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{21}
DS0045	ACCGTTTACCAA… (-57)	…CTTTTGAATA (+1171)	–
DS0046*	ACCGTTTACCAA… (-57)	…ACTTTTGAAT (+1170)	–
DS0047	ACCGTTTACCAA… (-57)	…CATACTTTTG (+1167)	A_4
DS0048	ACCGTTTACCAA… (-57)	…ACATACTTTT (+1166)	A_4
DS0049*	ACCGTTTACCAA… (-57)	…CGTTTTTTAC (+1158)	–
DS0050	ACCGTTTACCAA… (-57)	…TAACGTTTTT (+1155)	–
DS0051	GTTTACCAACTA… (-54)	…TATATATAAA (+1197)	A_{18}
DS0052	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{28}
DS0053	TTACCAACTACA… (-52)	…CTCCCGTTAA (+1553)	A_{16}
DS0054	AACTACATTTTT… (-47)	…ATGGT ATATG (+1791)	A_{23}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.2: act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ cid11Δ *cells*

Clone	5' End	3' End	Tail
DS0055	***TTACCGTTTACC… (-59)***	***…CTCCCGTTAA (+1553)***	***A_6U_4***
DS0056*	***TTACCGTTTACC… (-59)***	***…CTGATATATA (+1193)***	***A_7U_3***
DS0057	***TACCGTTTACCA… (-58)***	***…CTCCCGTTAA (+1553)***	***A_7***
DS0058	***TACCGTTTACCA… (-58)***	***…TATATATAAA (+1197)***	***A_6***
DS0059	***TACCGTTTACCA… (-58)***	***…AATAACATCG (+1177)***	***–***
DS0060	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{13}
DS0061	ACCGTTTACCAA… (-57)	…AGCCTTTTAC (+1363)	–
DS0062	ACCGTTTACCAA… (-57)	…TCTTTTTTTC (+1332)	–
DS0063	ACCGTTTACCAA… (-57)	…TTGGGTATCA (+1268)	–
DS0064	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_{24}
DS0065	ACCGTTTACCAA… (-57)	…ATAAATTTCA (+1202)	–
DS0066	ACCGTTTACCAA… (-57)	…ATATAAATTT (+1200)	–
DS0067	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{25}
DS0068	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_3
DS0069	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{22}
DS0070	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{12}
DS0071	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{15}
DS0072	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{50}
DS0073	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{18}
DS0074†	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	AGA_3
DS0075*	ACCGTTTACCAA… (-57)	…ACTTTTGAAT (+1170)	–
DS0076	ACCGTTTACCAA… (-57)	…TACTTTTGAA (+1169)	–
DS0077	ACCGTTTACCAA… (-57)	…ACATACTTTT (+1166)	A_4
DS0078	ACCGTTTACCAA… (-57)	…GTTTTTTACA (+1159)	–
DS0079	ACCGTTTACCAA… (-57)	…ACGTTTTTTA(+1157)	–
DS0080	***CCGTTTACCAAC… (-56)***	***…CTGATATATA (+1193)***	***A_6U_6***
DS0081	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{17}
DS0082	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{11}
DS0083	ACCAACTACATT… (-50)	…TATATATAAA (+1197)	A_{35}
DS0084	AACTACATTTTT… (-47)	…TATAAATTTC (+1201)	–
DS0085	ACATTTTTTGTA… (-34)	…GTTTTATATA (+1289)	–
DS0086	TTTGTAACGAAC… (-28)	…TTTTGTAACG (+1150)	–

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.3: act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ cid12Δ *cells*

Clone	5' End	3' End	Tail
DS0087	ACTTACCGTTTA … (-61)	…TTTTTACCTA (+1217)	–
DS0088	***TACCGTTTACCA… (-58)***	***…TATATATAAA (+1197)***	***A_6***
DS0089	***TACCGTTTACCA… (-58)***	***…AATAACATCG (+1177)***	–
DS0090	ACCGTTTACCAA… (-57)	…TATATGATCA (+1795)	A_7
DS0091	ACCGTTTACCAA… (-57)	…TTTTTACGAA (+1563)	A_6
DS0092	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{23}
DS0093*†	ACCGTTTACCAA… (-57)	…ACTCCCGTTA (+1552)	CA_{12}
DS0094	ACCGTTTACCAA… (-57)	…ATTTTACTTA (+1542)	–
DS0095	ACCGTTTACCAA… (-57)	…CCTTTTACAT (+1365)	–
DS0096	ACCGTTTACCAA… (-57)	…GATTACTATC (+1305)	–
DS0097	ACCGTTTACCAA… (-57)	… TCTGATTACT(+1302)	–
DS0098	ACCGTTTACCAA… (-57)	…TTTTATATAC (+1290)	–
DS0099	ACCGTTTACCAA… (-57)	…TGTTTTATAT (+1288)	–
DS0100	ACCGTTTACCAA… (-57)	…GGTATCAATT (+1271)	–
DS0101	ACCGTTTACCAA… (-57)	…TTTACCTATT (+1219)	A_{22}
DS0102	ACCGTTTACCAA… (-57)	…TCTTTTTTTA (+1213)	–
DS0103	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_7
DS0104	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	–
DS0105*	ACCGTTTACCAA… (-57)	…TATAAATTTC (+1201)	–
DS0106	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{19}
DS0107	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{13}
DS0108	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{12}
DS0109	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_5
DS0110	ACCGTTTACCAA… (-57)	…GATATATATA (+1195)	–
DS0111	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{23}
DS0112	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{28}
DS0113	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{22}
DS0114	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{15}
DS0115	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{14}
DS0116	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_6
DS0117	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{41}
DS0118	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{35}
DS0119	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{32}
DS0120†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	$A_{29}C$

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.3 (continued): act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ cid12Δ *cells*

Clone	5' End	3' End	Tail
DS0121*†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	$A_{28}C$
DS0122	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{24}
DS0123	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{23}
DS0124	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{22}
DS0125	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{21}
DS0126	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{18}
DS0127	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{13}
DS0128	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{12}
DS0129	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{11}
DS0130	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{11}
DS0131*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{10}
DS0132	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{10}
DS0133	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{9}
DS0134	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{9}
DS0135*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{7}
DS0136	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{6}
DS0137	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{43}
DS0138	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	–
DS0139	ACCGTTTACCAA… (-57)	…ACATCGATTC (+1181)	–
DS0140	ACCGTTTACCAA… (-57)	…ATAACATCGA (+1178)	–
DS0141†	ACCGTTTACCAA… (-57)	…AATAACATCG (+1177)	G
DS0142	ACCGTTTACCAA… (-57)	…TTGAATAACA (+1174)	–
DS0143*	ACCGTTTACCAA… (-57)	…TTTGAATAAC (+1173)	–
DS0144	ACCGTTTACCAA… (-57)	…CTTTTGAATA (+1171)	–
DS0145	ACCGTTTACCAA… (-57)	…ACTTTTGAAT (+1170)	–
DS0146	ACCGTTTACCAA… (-57)	…ATACTTTTGA (+1168)	–
DS0147	ACCGTTTACCAA… (-57)	…CATACTTTTG (+1167)	–
DS0148	ACCGTTTACCAA… (-57)	…TACATACTTT (+1165)	–
DS0149	ACCGTTTACCAA… (-57)	…TTTACATACT (+1163)	–
DS0150	ACCGTTTACCAA… (-57)	…TTTTACATAC (+1162)	–
DS0151*	ACCGTTTACCAA… (-57)	…TTTTTACATA (+1161)	–
DS0152*	ACCGTTTACCAA… (-57)	…CGTTTTTTAC (+1158)	–
DS0153	ACCGTTTACCAA… (-57)	…CGTTTTTTAC (+1158)	–
DS0154	ACCGTTTACCAA… (-57)	…ACGTTTTTTA (+1157)	–

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.3 (continued): act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ cid12Δ *cells*

Clone	5’ End	3’ End	Tail
DS0155	ACCGTTTACCAA… (-57)	…AACGTTTTTT (+1156)	–
DS0156	ACCGTTTACCAA… (-57)	…TTTGTAACGT (+1151)	–
DS0157	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{37}
DS0158	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{35}
DS0159	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{13}
DS0160	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{32}
DS0161	GTTTACCAACTA… (-54)	…ATTCTTCTGA (+1187)	A_{8}
DS0162	AACTACATTTTT… (-47)	…CTCCCGTTAA (+1553)	A_{16}
DS0163	AACTACATTTTT… (-47)	…TTACAGAAAT (+1389)	–
DS0164	AACTACATTTTT… (-47)	…TATATATAAA (+1197)	A_{28}G
DS0165†	AACTACATTTTT… (-47)	…TATATATAAA (+1197)	A_{4}
DS0166*	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{28}
DS0167	AACTACATTTTT… (-47)	…ATTCTTCTGA (+1187)	–
DS0168	ACATTTTTTGTA… (-43)	…ATAACATCGA (+1178)	–
DS0169	AACGAACCAAAA… (-32)	…GGAACTTTGG (+1262)	–
DS0170	ACGAACCAAAAA… (-31)	…ATTCTTCTGA (+1187)	A_{20}

Italic Bold sequences possess 3’ oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.4: act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ cid13Δ *cells*

Clone	5' End	3' End	Tail
DS0171*	***TTACCGTTTACC… (-59)***	***…CTCCCGTTAA (+1553)***	***A_7***
DS0172*†	***TTACCGTTTACC… (-59)***	***…CTGATATATA (+1193)***	***A_5UCU***
DS0173	***TTACCGTTTACC… (-59)***	***…GATTCTTCTG (+1186)***	***U_{10}***
DS0174*†	***TACCGTTTACCA… (-58)***	***…TATATATAAA (+1197)***	***A_6UA_3***
DS0175*	ACCGTTTACCAA… (-57)	…TGGT ATATGA (+1792)	A_{12}
DS0176	ACCGTTTACCAA… (-57)	…TGGT ATATGA (+1792)	A_{10}
DS0177	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{15}
DS0178*†	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_6UA_3
DS0179*	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_9
DS0180	ACCGTTTACCAA… (-57)	…GTAAATGAAT (+1535)	–
DS0181	ACCGTTTACCAA… (-57)	…CTTATAACTT (+1427)	–
DS0182	ACCGTTTACCAA… (-57)	…TTTTTTTCAT (+1334)	–
DS0183	ACCGTTTACCAA… (-57)	…TTTTTACCTA (+1217)	A_3
DS0184	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_{28}
DS0185	ACCGTTTACCAA… (-57)	…TATAAATTTC (+1201)	–
DS0186	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{19}
DS0187	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{12}
DS0188	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{10}
DS0189	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_5
DS0190	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_5
DS0191	ACCGTTTACCAA… (-57)	…GATATATATA (+1195)	–
DS0192*	ACCGTTTACCAA… (-57)	…TGATATATAT (+1194)	–
DS0193	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_4
DS0194	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{27}
DS0195	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{26}
DS0196	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_7
DS0197	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_6
DS0198	ACCGTTTACCAA… (-57)	…CTTCTGATAT (+1190)	–
DS0199†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_9CA_{15}
DS0200	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{25}
DS0201	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{22}
DS0202	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{21}
DS0203	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{18}
DS0204	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{16}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.4 (continued): act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ cid13Δ *cells*

Clone	5' End	3' End	Tail
DS0205	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{14}
DS0206	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{12}
DS0207	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{11}
DS0208	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{8}
DS0209	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	–
DS0210	ACCGTTTACCAA… (-57)	…GAATAACATC (+1176)	–
DS0211	ACCGTTTACCAA… (-57)	…TGAATAACAT (+1175)	–
DS0212*	ACCGTTTACCAA… (-57)	…TTGAATAACA (+1174)	–
DS0213	ACCGTTTACCAA… (-57)	…ACTTTTGAAT (+1170)	–
DS0214	ACCGTTTACCAA… (-57)	…TTTACATACT (+1163)	–
DS0215	ACCGTTTACCAA… (-57)	…TTTTTTACAT (+1160)	–
DS0216	ACCGTTTACCAA… (-57)	…CGTTTTTTAC (+1158)	–
DS0217	GTTTACCAACTA… (-54)	…TCTTTTTTTC (+1332)	–
DS0218	GTTTACCAACTA… (-54)	…TATAAATTTC (+1201)	–
DS0219	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{22}
DS0220*	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{15}
DS0221	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{13}
DS0222	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{7}
DS0223	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{31}
DS0224	ACCAACTACATT… (-50)	…TTCTGATATA (+1191)	A_{7}
DS0225	AACTACATTTTT… (-47)	…TGGT ATATGA (+1792)	A_{11}
DS0226*	***AACTACATTTTT… (-47)***	***…CTCCCGTTAA (+1553)***	***A_7U_2***
DS0227	AACTACATTTTT… (-47)	…TATATATAAA (+1197)	A_{33}
DS0228*†	AACTACATTTTT… (-47)	…TATATATAAA (+1197)	$A_{29}G$
DS0229	AACTACATTTTT… (-47)	…TATATATAAA (+1197)	A_{6}
DS0230	AACTACATTTTT… (-47)	…TTCTGATATA (+1191)	A_{29}
DS0231	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{43}
DS0232	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{40}
DS0233	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{18}
DS0234	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{18}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.5: act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ cid14Δ *cells*

Clone	5’ End	3’ End	Tail
DS0235	ACCGTTTACCAA… (-57)	…GTGCAATGTA (+1622)	–
DS0236	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_9
DS0237	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_9
DS0238	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_4
DS0239	ACCGTTTACCAA… (-57)	…TATTTTTGTC (+1477)	–
DS0240	ACCGTTTACCAA… (-57)	…ATATTGAAGT (+1373)	–
DS0241	ACCGTTTACCAA… (-57)	…AGCCTTTTAC (+1363)	–
DS0242	ACCGTTTACCAA… (-57)	…CTGTTTTCTT (+1326)	–
DS0243	ACCGTTTACCAA… (-57)	…TATCAATTAT (+1273)	–
DS0244	ACCGTTTACCAA… (-57)	…GGTATCAATT (+1271)	–
DS0245	ACCGTTTACCAA… (-57)	…ATGAGGTATC (+1267)	–
DS0246	ACCGTTTACCAA… (-57)	…TCTCATGAGG (+1254)	–
DS0247	ACCGTTTACCAA… (-57)	…TTTTACCTAT (+1218)	–
DS0248	ACCGTTTACCAA… (-57)	…CTTTTTTTAC (+1214)	–
DS0249	ACCGTTTACCAA… (-57)	…CTTTTTTTAC (+1214)	–
DS0250	ACCGTTTACCAA… (-57)	…AAATTTCAAT (+1204)	–
DS0251	ACCGTTTACCAA… (-57)	…TATAAATTTC (+1201)	–
DS0252	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{14}
DS0253	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{11}
DS0254	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_4
DS0255	ACCGTTTACCAA… (-57)	…GATATATATA (+1195)	–
DS0256*	ACCGTTTACCAA… (-57)	…TGATATATAT (+1194)	–
DS0257	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_4
DS0258	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	–
DS0259*†	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_9CA_8
DS0260	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{12}
DS0261*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{11}
DS0262	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{10}
DS0263	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_9
DS0264	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_8
DS0265*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_7
DS0266	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A
DS0267	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{27}
DS0268	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{22}

Italic Bold sequences possess 3’ oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.5 (continued): act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ cid14Δ *cells*

Clone	5' End	3' End	Tail
DS0269	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{22}
DS0270	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{20}
DS0271	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{15}
DS0272	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{12}
DS0273	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{7}
DS0274	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A
DS0275	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	–
DS0276	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{23}
DS0277	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{18}
DS0278	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{10}
DS0279	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{6}
DS0280	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{2}
DS0282	ACCGTTTACCAA… (-57)	…CGATTCTTCT (+1185)	–
DS0283	ACCGTTTACCAA… (-57)	…TCGATTCTTC (+1184)	–
DS0284	ACCGTTTACCAA… (-57)	…AACATCGATT (+1180)	–
DS0285	ACCGTTTACCAA… (-57)	…TTGAATAACA (+1174)	–
DS0286	ACCGTTTACCAA… (-57)	…ACTTTTGAAT (+1170)	–
DS0287	ACCGTTTACCAA… (-57)	…TACTTTTGAA (+1169)	–
DS0288	ACCGTTTACCAA… (-57)	…AACGTTTTTT (+1156)	–
DS0289	ACCGTTTACCAA… (-57)	…TTTGTAACGT (+1151)	–
DS0290	ACCAACTACATT… (-50)	…TCTTCTGATA (+1189)	A_{23}
DS0291	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{14}
DS0292	AACTACATTTTT… (-47)	…ACGTTTTTTA (+1157)	–
DS0293	CATTTTTTGTAA… (-42)	…TATACTTTCT (+1295)	–

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.6: act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ cid16Δ *cells*

Clone	5' End	3' End	Tail
DS0293	***TACCGTTTACCA… (-58)***	***…TATATATAAA (+1197)***	***A_{10}***
DS0294	***TACCGTTTACCA… (-58)***	***…TATATATAAA (+1197)***	***A_6***
DS0295	ACCGTTTACCAA… (-57)	…TGGAGTGTTT (+1283)	–
DS0296	ACCGTTTACCAA… (-57)	…GGTATCAATT (+1271)	–
DS0297	ACCGTTTACCAA… (-57)	…CTTTGGGTAT (+1266)	–
DS0298	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_2
DS0299	ACCGTTTACCAA… (-57)	…TCTTTTTTA (+1213)	–
DS0300	ACCGTTTACCAA… (-57)	…AATTTCAATC (+1205)	–
DS0301	ACCGTTTACCAA… (-57)	…TATAAATTTC (+1201)	–
DS0302*	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{12}
DS0303	ACCGTTTACCAA… (-57)	…GATATATATA (+1195)	–
DS0304	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_3
DS0305	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{14}
DS0306	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{11}
DS0307	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{10}
DS0308	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_9
DS0309	ACCGTTTACCAA… (-57)	…CTTCTGATAT (+1190)	–
DS0310	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{38}
DS0311	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{28}
DS0312	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{27}
DS0313	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{23}
DS0314	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{22}
DS0315	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{17}
DS0316	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{11}
DS0317	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{10}
DS0318	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_9
DS0319	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_7
DS0320	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_4
DS0321	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	–
DS0322†	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	$A_{18}CA_6$
DS0323†	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	$A_{17}CA_6$
DS0324*	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{22}
DS0325	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{20}
DS0326	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_4

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.6 (continued): act1 *cRACE sequences retrieved from* lsm1Δ cid1Δ cid16Δ *cells*

Clone	5’ End	3’ End	Tail
DS0327*	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	–
DS0328	ACCGTTTACCAA… (-57)	…AACATCGATT (+1180)	–
DS0329	ACCGTTTACCAA… (-57)	…ATAACATCGA (+1178)	–
DS0330	ACCGTTTACCAA… (-57)	…AATAACATCG (+1177)	–
DS0331	ACCGTTTACCAA… (-57)	…TTGAATAACA (+1174)	–
DS0332*	ACCGTTTACCAA… (-57)	…TTGAATAACA (+1174)	–
DS0333	ACCGTTTACCAA… (-57)	…TACATACTTT (+1165)	–
DS0334	ACCGTTTACCAA… (-57)	…TTTTTACATA (+1161)	–
DS0335*	ACCGTTTACCAA… (-57)	…TTTGTAACGT (+1151)	–
DS0336	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{14}
DS0337	GTTTACCAACTA… (-54)	…ATACTTTTGA (+1168)	–
DS0338	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{18}
DS0339	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{21}
DS0340	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{9}
DS0341†	AACTACATTTTT… (-47)	…TTCTTCTGAT (+1188)	GA_{35}
DS0342	AACTACATTTTT… (-47)	…AATAACATCG (+1177)	–
DS0343	AAAAGACAAGAC… (-13)	…CTCCCGTTAA (+1553)	A_{11}

Italic Bold sequences possess 3’ oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.7: act1 *cRACE sequences retrieved from* cid1Δ cid12Δ *cells*

Clone	5' End	3' End	Tail
DS0344†	GTGTTATAACTT… (-69)	…TATATATAAA (+1197)	A_8CA_{14}
DS0345	***TACCGTTTACCA… (-58)***	***…CTCCCGTTAA (+1553)***	A_{10}
DS0346	***TACCGTTTACCA… (-58)***	***…TTTTTACCTA (+1217)***	A_{21}
DS0347	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	A_{10}
DS0348	ACCGTTTACCAA… (-57)	…TGGT ATATGA (+1792)	A_5
DS0349	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{36}
DS0350	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{20}
DS0351	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{15}
DS0352	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_8
DS0353	ACCGTTTACCAA… (-57)	…CCTTTTACAT (+1365)	–
DS0354	ACCGTTTACCAA… (-57)	…CTAGCCTTTT (+1361)	–
DS0355	ACCGTTTACCAA… (-57)	…TTTGGGTATC (+1267)	–
DS0356	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_{22}
DS0357	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_{14}
DS0358	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_{11}
DS0359	ACCGTTTACCAA… (-57)	…TTTTTACCTA (+1217)	A_{29}
DS0360	ACCGTTTACCAA… (-57)	…TTTTTACCTA (+1217)	A_{22}
DS0361	ACCGTTTACCAA… (-57)	…TTTTTACCTA (+1217)	A_{21}
DS0362	ACCGTTTACCAA… (-57)	…TTTTTACCTA (+1217)	A_{19}
DS0363	ACCGTTTACCAA… (-57)	…TTTTTACCTA (+1217)	A_{16}
DS0364	ACCGTTTACCAA… (-57)	…TCTTTTTTTA (+1213)	–
DS0365	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{36}
DS0366	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{35}
DS0367	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{35}
DS0368	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{32}
DS0369	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{24}
DS0370	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{20}
DS0371	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{20}
DS0372†	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	$A_{17}UA_2$
DS0373	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{18}
DS0374	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{14}
DS0375	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{12}
DS0376	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{11}
DS0377	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_7

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.7 (continued): act1 *cRACE sequences retrieved from* cid1Δ cid12Δ *cells*

Clone	5' End	3' End	Tail
DS0378	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_4
DS0379	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_3
DS0380	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{27}
DS0381	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{13}
DS0382	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{29}
DS0383	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{26}
DS0384	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{25}
DS0385	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{24}
DS0386*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{23}
DS0387	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{22}
DS0388	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{18}
DS0389	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{17}
DS0390	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{16}
DS0391	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{15}
DS0392*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{14}
DS0393	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{13}
DS0394*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{11}
DS0395*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{11}
DS0396	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{10}
DS0397	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_9
DS0398*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_8
DS0399	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_7
DS0400	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{50}
DS0401†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	$A_{26}GA_{16}$
DS0402	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{40}
DS0403	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{39}
DS0404*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{38}
DS0405	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{28}
DS0406*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{27}
DS0407	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{24}
DS0408	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{23}
DS0409	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{22}
DS0410†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_2CA_{17}
DS0411*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{20}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.7 (continued): act1 *cRACE sequences retrieved from* cid1Δ cid12Δ *cells*

Clone	5' End	3' End	Tail
DS0412	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{18}
DS0413*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{18}
DS0414	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{17}
DS0415*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{16}
DS0416	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{15}
DS0417	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{15}
DS0418	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{14}
DS0419	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{13}
DS0420†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	ACA_{11}
DS0421	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{11}
DS0422	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{10}
DS0423*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{9}
DS0424	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{8}
DS0425*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{2}
DS0426	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{24}
DS0427	ACCGTTTACCAA… (-57)	…ATACTTTTGA (+1168)	–
DS0428	ACCGTTTACCAA… (-57)	…AACGTTTTTT (+1156)	–
DS0429	CCGTTTACCAAC… (-56)	…ATATAAATTT (+1200)	–
DS0430	GTTTACCAACTA… (-54)	…CTCCCGTTAA (+1553)	A_{19}
DS0431	GTTTACCAACTA… (-54)	…TATATATAAA (+1197)	A_{34}
DS0432	GTTTACCAACTA… (-54)	…CTGATATATA (+1193)	A_{28}
DS0433	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{32}
DS0434††	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	$A_4GA_8GA_{16}$
DS0435	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{28}
DS0436	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{25}
DS0437	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{7}
DS0438†	AACTACATTTTT… (-47)	…TGGT ATATGA (+1792)	A_6GA_{13}
DS0439	AACTACATTTTT… (-47)	…AACTTTGGGT (+1264)	–
DS0440	AACTACATTTTT… (-47)	…TAAATTTCAA (+1203)	A_{14}
DS0441	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{32}
DS0442	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{32}
DS0443	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{26}
DS0444	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{19}
DS0445*	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{18}
DS0446	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{17}
DS0447	AACGAACCAAAA… (-32)	…CGTTTTTTAC (+1158)	–

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.8: act1 *cRACE sequences retrieved from* cid1Δ cid13Δ *cells*

Clone	5' End	3' End	Tail
DS0448	GTTATAACTTAC… (-67)	…TCTTCTGATA (+1189)	A_{21}
DS0449	***TTACCGTTTACC… (-59)***	***…CTCCCGTTAA (+1553)***	***A_6U_3***
DS0450*	***TTACCGTTTACC… (-59)***	***…TTCTGATATA (+1191)***	***A***
DS0451	***TTACCGTTTACC… (-59)***	***…GATTCTTCTG (+1186)***	***U***
DS0452	***TACCGTTTACCA… (-58)***	***…TTCTGATATA (+1191)***	***A_6***
DS0453	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{43}***
DS0454	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{45}
DS0455	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{24}
DS0456	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{24}
DS0457	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{19}
DS0458	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{14}
DS0459††	ACCGTTTACCAA… (-57)	…ACTCCCGTTA (+1552)	CA_8UA_2
DS0460	ACCGTTTACCAA… (-57)	…TTACAGAAAT (+1382)	–
DS0461	ACCGTTTACCAA… (-57)	…CTAGCCTTTT (+1361)	–
DS0462	ACCGTTTACCAA… (-57)	…TTATTGGAGT (+1279)	–
DS0463	ACCGTTTACCAA… (-57)	…TTTGGGTATC (+1267)	–
DS0464	ACCGTTTACCAA… (-57)	…GAACTTTGGG (+1263)	–
DS0465	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_{37}
DS0466	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_{25}
DS0467*	ACCGTTTACCAA… (-57)	…TTTTTACCTA (+1217)	A_{12}
DS0468	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_{25}
DS0469	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_{23}
DS0470	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_{20}
DS0471	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_{17}
DS0472	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_{10}
DS0473	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_7
DS0474	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{37}
DS0475	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{35}
DS0476	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{31}
DS0477	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{28}
DS0478	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{22}
DS0479	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{22}
DS0480	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{20}
DS0481	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{19}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.8 (continued): act1 *cRACE sequences retrieved from* cid1Δ cid13Δ *cells*

Clone	5’ End	3’ End	Tail
DS0482*	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_8
DS0483	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_3
DS0484	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{21}
DS0485	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{32}
DS0486	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{31}
DS0487	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{29}
DS0488	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{25}
DS0489	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{25}
DS0490	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{22}
DS0491	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{20}
DS0492*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{18}
DS0493	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{17}
DS0494	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{15}
DS0495	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{14}
DS0496	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{13}
DS0497	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{11}
DS0498	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{11}
DS0499	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{10}
DS0500	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_7
DS0501	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_7
DS0502*	ACCGTTTACCAA… (-57)	… CTTCTGATAT (+1190)	–
DS0503	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{53}
DS0504	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{51}
DS0505	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{38}
DS0506	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{37}
DS0507	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{33}
DS0508	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{29}
DS0509†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	$A_{24}GA_4$
DS0510*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{27}
DS0511*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{26}
DS0512*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{24}
DS0513	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{22}
DS0514*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{19}
DS0515*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{18}

Italic Bold sequences possess 3’ oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.8 (continued): act1 *cRACE sequences retrieved from* cid1Δ cid13Δ *cells*

Clone	5' End	3' End	Tail
DS0516*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{17}
DS0517	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{17}
DS0518	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{16}
DS0519	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{14}
DS0520	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{12}
DS0521†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	GA_{11}
DS0522	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{9}
DS0523*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{8}
DS0524	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{8}
DS0525	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{7}
DS0526	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{4}
DS0527	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{45}
DS0528	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{43}
DS0529	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{36}
DS0530*	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{34}
DS0531	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{26}
DS0532	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{16}
DS0533	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{10}
DS0534	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{9}
DS0535	ACCGTTTACCAA… (-57)	…TTTTTTACAT (+1160)	–
DS0536	ACCGTTTACCAA… (-57)	…CGTTTTTTAC (+1158)	–
DS0537	GTTTACCAACTA… (-54)	…CTCCCGTTAA (+1553)	A_{13}
DS0538	GTTTACCAACTA… (-54)	…TATATATAAA (+1197)	A_{8}
DS0539	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{46}
DS0540*	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{28}
DS0541†	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_3UA_{47}
DS0542	ACCAACTACATT… (-50)	…TCTTCTGATA (+1189)	A_{19}
DS0543	ACCAACTACATT… (-50)	…TCGATTCTTC (+1184)	–
DS0544	AACTACATTTTT… (-47)	…TATATATAAA (+1197)	A_{30}
DS0545	AACTACATTTTT… (-47)	…TATATATAAA (+1197)	A_{10}
DS0546	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{38}
DS0547	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{37}
DS0548	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{29}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.9: act1 *cRACE sequences retrieved from* cid1Δ cid14Δ *cells*

Clone	5' End	3' End	Tail
DS0549	***TTACCGTTTACC… (-59)***	***…TCTTCTGATA (+1189)***	***A_8U***
DS0550	***TACCGTTTACCA… (-58)***	***…TATATATAAA (+1197)***	***A_7***
DS0551	***TACCGTTTACCA… (-58)***	***…ATTCTTCTGA (+1187)***	***A_4***
DS0552	ACCGTTTACCAA… (-57)	…TATATGATCA (+1795)	A_{21}
DS0553	ACCGTTTACCAA… (-57)	…TATATGATCA (+1795)	A_2
DS0554	ACCGTTTACCAA… (-57)	…ACATACATGG (+1787)	A_{17}
DS0555	ACCGTTTACCAA… (-57)	…TTAATACATA (+1780)	A_{10}
DS0556	ACCGTTTACCAA… (-57)	…GTTTATTAGG (+1758)	–
DS0557	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{38}
DS0558	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{36}
DS0559	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{21}
DS0560	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{17}
DS0561	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{16}
DS0562	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{16}
DS0563	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{12}
DS0564	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{11}
DS0565*	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{10}
DS0566	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_9
DS0567	ACCGTTTACCAA… (-57)	…TTTCTTTAAC (+1442)	–
DS0568	ACCGTTTACCAA… (-57)	…TATTTCCAGT (+1391)	–
DS0569	ACCGTTTACCAA… (-57)	…GTTTTCTTTT (+1328)	–
DS0570	ACCGTTTACCAA… (-57)	…ATACTTTCTG (+1296)	–
DS0571	ACCGTTTACCAA… (-57)	…GGGTATAACT (+1270)	–
DS0572	ACCGTTTACCAA… (-57)	…AACTTTGGGT (+1264)	–
DS0573	ACCGTTTACCAA… (-57)	…GAACTTTGGG (+1263)	–
DS0574†	ACCGTTTACCAA… (-57)	…TATTTAACCT (+1225)	A_7CA_6
DS0575	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_{29}
DS0576	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_7
DS0577	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_7
DS0578†	ACCGTTTACCAA… (-57)	…TACCTATTTA (+1221)	GA_{15}
DS0579	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_{14}
DS0580	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_6
DS0581	ACCGTTTACCAA… (-57)	…TATAAATTTC (+1201)	–
DS0582	ACCGTTTACCAA… (-57)	…ATATATAAAT (+1198)	A_{28}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.9 (continued): act1 *cRACE sequences retrieved from* cid1Δ cid14Δ *cells*

Clone	5' End	3' End	Tail
DS0583	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{35}
DS0584	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{28}
DS0585†	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_5UA_{21}
DS0586	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{23}
DS0587†	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	$A_{19}GA_3$
DS0588	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{17}
DS0589	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{17}
DS0590	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{15}
DS0591	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{12}
DS0592	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{11}
DS0593	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{10}
DS0594	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_7
DS0595	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A
DS0596†	ACCGTTTACCAA… (-57)	…GATATATATA (+1195)	GA_6
DS0597*	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	–
DS0598*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{26}
DS0599*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{23}
DS0600	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{21}
DS0601	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{18}
DS0602	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{17}
DS0603	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{15}
DS0604*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{12}
DS0605	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{10}
DS0606*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A
DS0607	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{47}
DS0608	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{46}
DS0609	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{41}
DS0610	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{40}
DS0611	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{36}
DS0612	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{36}
DS0613	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{34}
DS0614†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	$A_{23}GA_9$
DS0615*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{25}
DS0616*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{24}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.9 (continued): act1 *cRACE sequences retrieved from* cid1Δ cid14Δ *cells*

Clone	5' End	3' End	Tail
DS0617	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{21}
DS0618†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	$A_{12}GA_7$
DS0619†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	$A_{19}GA_3$
DS0620	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{18}
DS0621	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{18}
DS0622*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{17}
DS0623	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{15}
DS0624*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{14}
DS0625	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{13}
DS0626*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_7
DS0627	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_6
DS0628	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_3
DS0629	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{26}
DS0630	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{23}
DS0631	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{10}
DS0632*	ACCGTTTACCAA… (-57)	…TTTGAATAAC (+1173)	–
DS0633*	ACCGTTTACCAA… (-57)	…ACTTTTGAAT (+1170)	–
DS0634	ACCGTTTACCAA… (-57)	…TTTTTTACAT (+1160)	–
DS0635	GTTTACCAACTA… (-54)	…CTCCCGTTAA (+1553)	A_{46}
DS0636	GTTTACCAACTA… (-54)	…CTGATATATA (+1193)	A_{45}
DS0637	GTTTACCAACTA… (-54)	…CTGATATATA (+1193)	–
DS0638	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{11}
DS0639	GTTTACCAACTA… (-54)	…AACGTTTTTT (+1156)	–
DS0640	CAACTACATTTT… (-48)	…CTCCCGTTAA (+1553)	A_{37}
DS0641	CAACTACATTTT… (-48)	…TCTTCTGATA (+1189)	A_{18}
DS0642	AACTACATTTTT… (-47)	…TTCGTCCTGT (+1650)	–
DS0643	AACTACATTTTT… (-47)	…ATAACTTCCC (+1430)	–
DS0644	AACTACATTTTT… (-47)	…ACCTATTTAA (+1222)	A_{23}
DS0645	AACTACATTTTT… (-47)	…TATATATAAA (+1197)	A_{22}
DS0646	AACTACATTTTT… (-47)	…TTCTGATATA (+1191)	A_{26}
DS0647††	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	$A_{22}GA_2CA_{14}$
DS0648	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{26}
DS0649	ACATTTTTTGTA… (-43)	…TCTTCTGATA (+1189)	A_{32}
DS0650	TTTTTTGTAACG… (-40)	…TTTTGTAACG (+1150)	–
DS0651	GTAACGAACCAA… (-34)	…TATATATAAA (+1197)	A_{12}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.10: act1 *cRACE sequences retrieved from* cid1Δ cid16Δ *cells*

Clone	5' End	3' End	Tail
DS0652	GTTATAACTTAC… (-67)	…TTCTGATATA (+1191)	A_{31}
DS0653	ACCGTTTACCAA… (-57)	…TATATGATCA (+1795)	A_{10}
DS0654†	ACCGTTTACCAA… (-57)	…ACATACATGG (+1785)	CA_{14}
DS0655	ACCGTTTACCAA… (-57)	…ACTTTGAAGC (+1710)	–
DS0656	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{34}
DS0657	ACCGTTTACCAA… (-57)	…TTATGTTTAT (+1497)	–
DS0658†	ACCGTTTACCAA… (-57)	…GCCTTTTACA (+1364)	A_5CA_{10}
DS0659*	ACCGTTTACCAA… (-57)	…TTTCTTGCTC (+1317)	A_{24}
DS0660†	ACCGTTTACCAA… (-57)	…TTTCTTGCTC (+1317)	$A_{18}CA$
DS0661	ACCGTTTACCAA… (-57)	…AGTGTTTTAT (+1286)	–
DS0662	ACCGTTTACCAA… (-57)	…GGTATCAATT (+1271)	–
DS0663	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{24}
DS0664	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{17}
DS0665	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{13}
DS0666	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{42}
DS0667	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{34}
DS0668	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{31}
DS0669	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{30}
DS0670	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{26}
DS0671†	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_3UA_{22}
DS0672	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{23}
DS0673	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{21}
DS0674	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{17}
DS0675	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{16}
DS0676	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{14}
DS0677	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{13}
DS0678	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_6
DS0679	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{36}
DS0680††	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	$A_8GA_4GA_{18}$
DS0681	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{32}
DS0682	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{31}
DS0683	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{30}
DS0684	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{27}
DS0685	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{23}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.10: act1 *cRACE sequences retrieved from* cid1Δ cid16Δ *cells*

Clone	5' End	3' End	Tail
DS0686	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{19}
DS0687†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	AGA_{15}
DS0688	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{12}
DS0689	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{10}
DS0690	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{6}
DS0691	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{28}
DS0692	ACCGTTTACCAA… (-57)	…TAATACATAC (+1781)	A_{9}
DS0693	AACTACATTTTT… (-47)	…TCTTTTTTTA (+1213)	A_{50}
DS0694	AACTACATTTTT… (-47)	…TATATATAAA (+1197)	A_{14}
DS0695	AACTACATTTTT… (-47)	…CTGATATATA (+1193)	A_{19}
DS0696	AACTACATTTTT… (-47)	…TTCTGATATA (+1191)	A_{23}
DS0697	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{24}
DS0698	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{22}
DS0699	AACGAACCAAAA… (-32)	…CTCCCGTTAA (+1553)	A_{9}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.11: act1 *cRACE sequences retrieved from induced* cid1Δ (pREP1-NTAP) *cells*

Clone	5' End	3' End	Tail
DS0700	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{40}***
DS0701	ACCGTTTACCAA… (-57)	…TAAATTTCAA (+1203)	A_{12}
DS0702	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{14}
DS0703†	ACCGTTTACCAA… (-57)	…ATATATATAA (+1196)	GA_5
DS0704	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{23}
DS0705	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{10}
DS0706	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{35}
DS0707	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{32}
DS0708	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{21}
DS0709	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{20}
DS0710	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{19}
DS0711*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{17}
DS0712	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{27}
DS0713	***GTTTACCAACTA… (-54)***	***…CGATTCTTCT (+1185)***	***$A_{31}U$***
DS0714	ACCAACTACATT… (-50)	…TAAATTTCAA (+1203)	A_{28}
DS0715	CATTTTTTGTAA… (-42)	…ACCTATTTAA (+1222)	A_{29}
DS0716	TCGCAGCGTTGG… (+14)	…TTCTGATATA (+1191)	A_{25}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.12: act1 *cRACE sequences retrieved from induced* cid1Δ (pREP1-NTAP-cid1^{+}) *cells*

Clone	5' End	3' End	Tail
DS0717	***TTACCGTTTACC… (-59)***	***…TATATATAAA (+1197)***	***A_{26}***
DS0718	***TTACCGTTTACC… (-59)***	***…TATATATAAA (+1197)***	***A_{24}***
DS0719	***TTACCGTTTACC… (-59)***	***…TCTTCTGATA (+1189)***	***A_{29}***
DS0720	***TACCGTTTACCA… (-58)***	***…ACCTATTTAA (+1222)***	***A_{14}***
DS0721	***TACCGTTTACCA… (-58)***	***…TCTTTTTTA (+1213)***	***A_{36}***
DS0722	***TACCGTTTACCA… (-58)***	***…TATATATAAA (+1197)***	***A_{41}***
DS0723	***TACCGTTTACCA… (-58)***	***…TATATATAAA (+1197)***	***A_{26}***
DS0724	***TACCGTTTACCA… (-58)***	***…TATATATAAA (+1197)***	***A_{25}***
DS0725	***TACCGTTTACCA… (-58)***	***…TATATATAAA (+1197)***	***A_{14}***
DS0726	***TACCGTTTACCA… (-58)***	***…CTGATATATA (+1193)***	***A_{26}***
DS0727	***TACCGTTTACCA… (-58)***	***…TTCTGATATA (+1191)***	***A_{37}***
DS0728	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{48}***
DS0729	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{39}***
DS0730	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{39}***
DS0731	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{37}***
DS0732	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{34}***
DS0733	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{34}***
DS0734	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{29}***
DS0735	***TACCGTTTACCA… (-58)***	***…TCTTCTGATA (+1189)***	***A_{26}***
DS0736	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{37}
DS0737†	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	$A_{15}UA_{19}$
DS0738	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{31}
DS0739	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{31}
DS0740	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{26}
DS0741	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{21}
DS0742	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{32}
DS0743	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{22}
DS0744†	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	$A_{20}UA$
DS0745	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{20}
DS0746	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{45}
DS0747	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{43}
DS0748	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{43}
DS0749	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{36}
DS0750	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{36}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.12 (continued): act1 *cRACE sequences retrieved from induced* cid1Δ (pREP1-NTAP-cid1^{+}) *cells*

Clone	5' End	3' End	Tail
DS0751	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{35}
DS0752	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{33}
DS0753†	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_7GA_{25}
DS0754	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{31}
DS0755	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{31}
DS0756	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{33}
DS0757	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_8
DS0758	ACCGTTTACCAA… (-57)	…TTTTTTACAT (+1160)	–
DS0759	ACCGTTTACCAA… (-57)	…CGTTTTTTAC (+1158)	–
DS0760	ACCGTTTACCAA… (-57)	…AACGTTTTTT (+1156)	–
DS0761	ACCGTTTACCAA… (-57)	…GAAGGTCAAG (+984)	–
DS0762	GTTTACCAACTA… (-54)	…CTCCCGTTAA (+1553)	A_{28}
DS0763	GTTTACCAACTA… (-54)	…TTTTTTACCT (+1216)	–
DS0764	***AACTACATTTTT… (-47)***	***…TAAATTTCAA (+1203)***	***$A_{30}U$***
DS0765	***AACTACATTTTT… (-47)***	***…TTCTGATATA (+1191)***	***$A_{37}U$***
DS0766†	***AACTACATTTTT… (-47)***	***…TCTTCTGATA (+1189)***	***$A_{37}UAU_2$***
DS0767	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{35}
DS0768†	***AACTACATTTTT… (-47)***	***…ATTCTTCTGA (+1187)***	***$A_{18}CA_{14}U$***
DS0769	AACTACATTTTT… (-47)	…CATACTTTG (+1167)	–
DS0770	AACTACATTTTT… (-47)	…AACGTTTTTT (+1156)	–

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.13: act1 *cRACE sequences retrieved from induced* cid1Δ (pREP1-NTAP-cid12^{+}) *cells*

Clone	5' End	3' End	Tail
DS0771	***TACCGTTTACCA… (-58)***	***…TTCTGATATA (+1191)***	***A_{18}***
DS0772	ACCGTTTACCAA… (-57)	…CTCCCGTTAA (+1553)	A_{23}
DS0773†	ACCGTTTACCAA… (-57)	…TTTCTTGCTC (+1317)	A_{18}CA
DS0774	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_{24}
DS0775	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{34}
DS0776	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{31}
DS0777*	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{23}
DS0778	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{22}
DS0779	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{19}
DS0780†	ACCGTTTACCAA… (-57)	…GATATATATA (+1195)	GA_6
DS0781	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{28}
DS0782	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{27}
DS0783	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{23}
DS0784*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{22}
DS0785*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{21}
DS0786	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{19}
DS0787	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{18}
DS0788	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{17}
DS0789	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{41}
DS0790	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{31}
DS0791	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{31}
DS0792	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{28}
DS0793	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{23}
DS0794	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{23}
DS0795	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{20}
DS0796	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{19}
DS0797*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{19}
DS0798	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{17}
DS0799	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{13}
DS0800	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{11}
DS0801	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{10}
DS0802	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{19}
DS0803	ACCGTTTACCAA… (-57)	…AAGCTCCTCT (+1136)	–
DS0804	GTTTACCAACTA… (-54)	…TATATATAAA (+1197)	A_{19}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.13 (continued): act1 *cRACE sequences retrieved from induced* cid1Δ (pREP1-NTAP-cid12^{+}) *cells*

Clone	5' End	3' End	Tail
DS0805	***GTTTACCAACTA… (-54)***	***…TATATATAAA (+1197)***	***$A_{16}U$***
DS0806	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{17}
DS0807	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{14}
DS0808	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{13}
DS0809	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{12}
DS0810*	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{6}
DS0811	AACTACATTTTT… (-47)	…TAAATTTCAA (+1203)	A_{14}
DS0812†	AACTACATTTTT… (-47)	…CTGATATATA (+1193)	A_{18}CA
DS0813	AACTACATTTTT… (-47)	…TCTTCTGATA (+1189)	A_{32}
DS0814	CGAACCAAAAAA… (-30)	…TCTTCTGATA (+1189)	A_{28}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.14: act1 *cRACE sequences retrieved from* cid1Δ (pREP1-NTAP-cid13^{+}) *cells*

Clone	5' End	3' End	Tail
DS0815	GTTATAACTTAC… (-67)	…TTCTGATATA (+1191)	A_{32}
DS0816†	ACCGTTTACCAA… (-57)	…TTTCTTGCTC (+1317)	A_{16}CA
DS0817	ACCGTTTACCAA… (-57)	…GGAGTGTTTT (+1284)	–
DS0818	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{23}
DS0819	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{27}
DS0820	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{22}
DS0821	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{15}
DS0822	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{22}
DS0823	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{20}
DS0824	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{15}
DS0825	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{22}
DS0826	ACCGTTTACCAA… (-57)	…ACATACTTTT (+1166)	–
DS0827	ACCGTTTACCAA… (-57)	…CGTTTTTTAC (+1158)	–

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.15: act1 *cRACE sequences retrieved from* cid1Δ (pREP1-NTAP-cid14+) *cells*

Clone	5' End	3' End	Tail
DS0828	ACCGTTTACCAA… (-57)	…AACTTTGGGT (+1265)	–
DS0829	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_8
DS0830	ACCGTTTACCAA… (-57)	…GATATATATA (+1195)	–
DS0831	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{25}
DS0832	ACCGTTTACCAA… (-57)	…CTGATATATA (+1193)	A_{21}
DS0833	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{23}
DS0834*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{22}
DS0835	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{14}
DS0836*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{24}
DS0837	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{20}
DS0838	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{19}
DS0839	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{17}
DS0840	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{15}
DS0841	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{13}
DS0842	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{10}
DS0843	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{19}
DS0844*	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{15}
DS0845	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{24}
DS0846	GTTTACCAACTA… (-54)	…TAAATTTCAA (+1203)	A_6
DS0847	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{12}
DS0848	AACTACATTTTT… (-47)	…TATATATAAA (+1197)	A_{27}

Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix D.16: act1 *cRACE sequences retrieved from* cid1Δ (pREP1-NTAP-cid16^{+}) *cells*

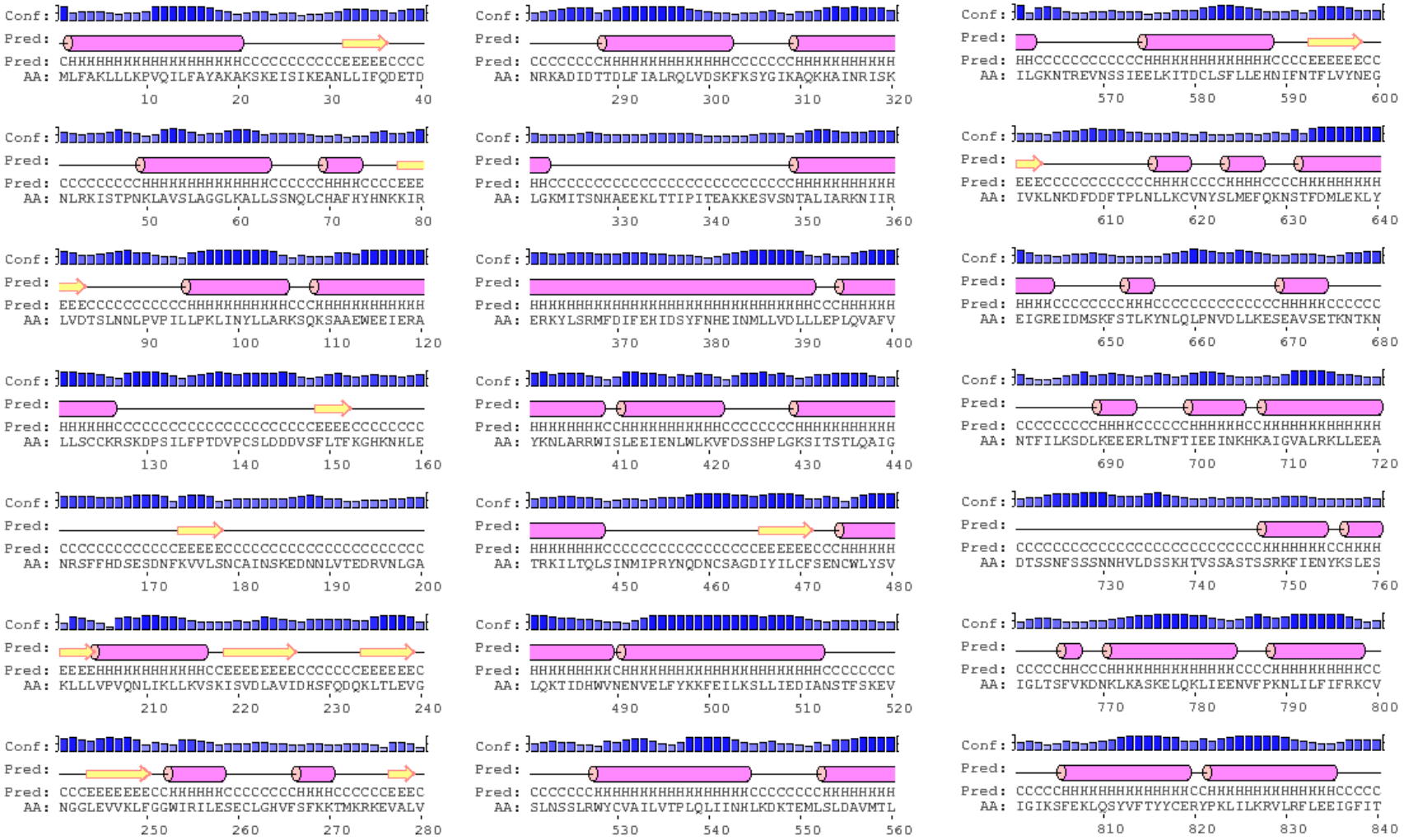
Clone	5' End	3' End	Tail
DS0849	GTGTTATAACTT… (-69)	…TCTTCTGATA (+1189)	A_{22}
DS0850	ACCGTTTACCAA… (-57)	…TATATGATCA (+1795)	A_{8}
DS0851*	ACCGTTTACCAA… (-57)	…TATATGATCA (+1795)	A_{7}
DS0852	ACCGTTTACCAA… (-57)	…CGTTTTCTTT (+1439)	–
DS0853	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_{34}
DS0854	ACCGTTTACCAA… (-57)	…ACCTATTTAA (+1222)	A_{16}
DS0855†	ACCGTTTACCAA… (-57)	…CTTTTTTACC (+1215)	AGA_{20}
DS0856	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{22}
DS0857*	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{20}
DS0858	ACCGTTTACCAA… (-57)	…TATATATAAA (+1197)	A_{18}
DS0859	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{22}
DS0860	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{22}
DS0861*	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{21}
DS0862	ACCGTTTACCAA… (-57)	…TTCTGATATA (+1191)	A_{9}
DS0863	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{28}
DS0864*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{24}
DS0865	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{23}
DS0866	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{20}
DS0867*	ACCGTTTACCAA… (-57)	…TCTTCTGATA (+1189)	A_{17}
DS0868	ACCGTTTACCAA… (-57)	…ATTCTTCTGA (+1187)	A_{26}
DS0869	GTTTACCAACTA… (-54)	…TTCTGATATA (+1191)	A_{15}
DS0870	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{13}
DS0871*	GTTTACCAACTA… (-54)	…TCTTCTGATA (+1189)	A_{13}

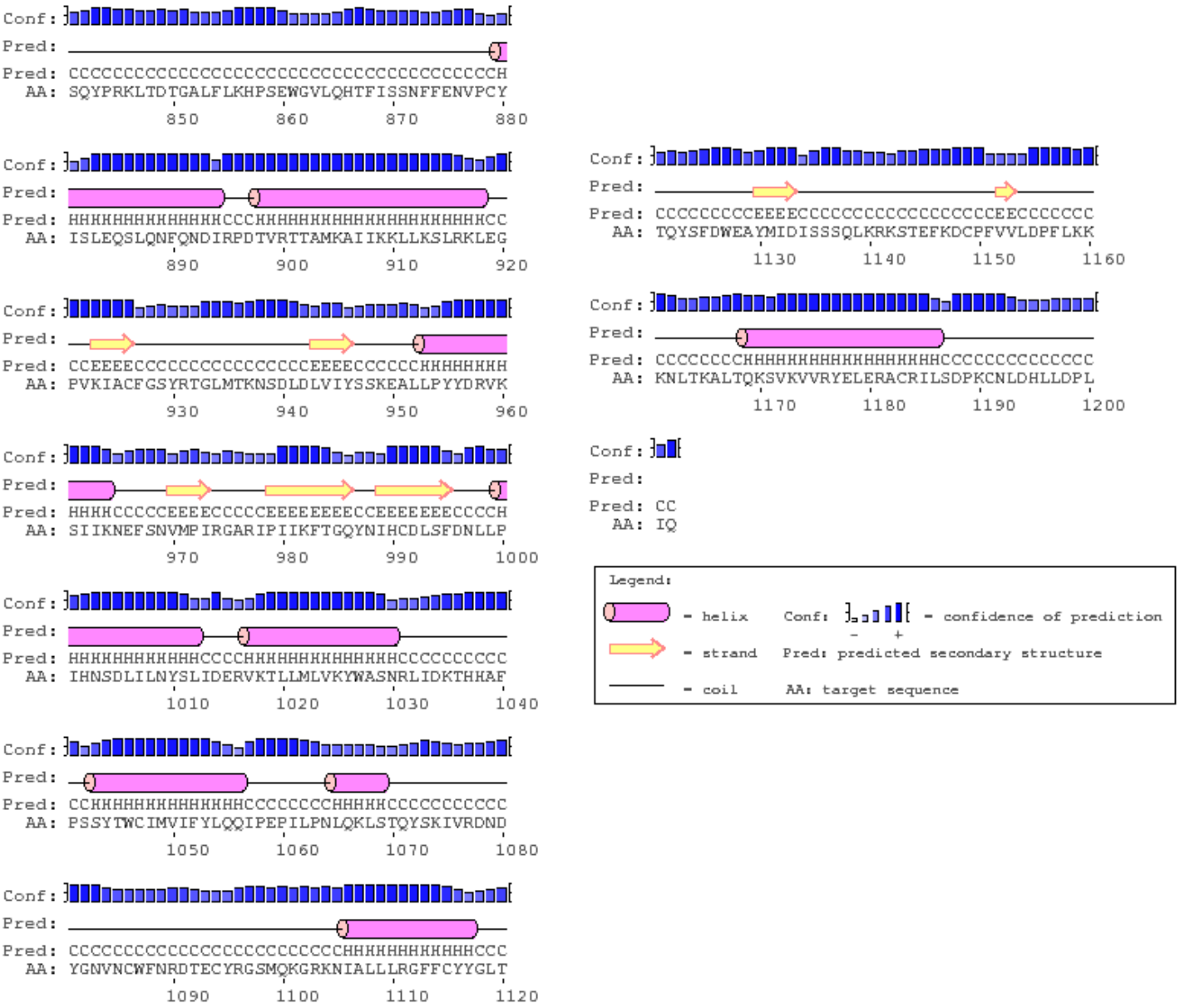
Italic Bold sequences possess 3' oligo(U) tails.

* retrieved in multiple clones.

† contains internal non-A residues in the poly(A) tail.

Appendix E: Structure prediction for Cid16 by the PSIPRED algorithm




Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCH
AA: SQYPRKLTDTGALFLKHPSEWGVLQHTFISSNFFENVPCY
850 860 870 880
Conf:
Pred:
Pred: HHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHCC
AA: ISLEQSLQNFQNDIRPDTVRTTAMKAIIKKLLKSLRKLEG
890 900 910 920
Conf:
Pred:
Pred: CCEEEECCCCCCCCCCCCCCCCEEEECCCCCCCHHHHHHH
AA: PVKIACFGSYRTGLMTKNSDLDLVIYSSKEALLPYYDRVK
930 940 950 960
Conf:
Pred:
Pred: HHHHCCCCCEEEECCCCCEEEEEEECCEEEEEECCCCH
AA: SIIKNEFSNVMPIRGARIPIIKFTGQYNIHCDLSFDNLLP
970 980 990 1000
Conf:
Pred:
Pred: HHHHHHHHHHHHCCCCHHHHHHHHHHHHHCCCCCCCCCC
AA: IHNSDLILNYSLIDERVKTLLMLVKYWASNRLIDKTHHAF
1010 1020 1030 1040
Conf:
Pred:
Pred: CCHHHHHHHHHHHHHHCCCCCCCCHHHHHCCCCCCCCCC
AA: PSSYTWCIMVIFYLQQIPEPILPNLQKLSTQYSKIVRDND
1050 1060 1070 1080
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCCC
AA: YGNVNCWFNRDTECYRGSMQKGRKNIALLLRGFFCYYGLT
1090 1100 1110 1120
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: TQYSFDWEAYMIDISSSQLKRKSTEFKDCPFVVLDPFLKK
1130 1140 1150 1160
Conf:
Pred:
Pred: CCCCCCCCCHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCC
AA: KNLTKALTQKSVKVVRYELERACRILSDPKCNLDHLLDPL
1170 1180 1190 1200
Conf:
Pred:
Pred: CC
AA: IQ
Legend:
= helix
= strand
= coil
Conf: - +
= confidence of prediction
Pred: predicted secondary structure
AA: target sequence

Appendix F: List of genes >2-fold up- or down-regulated in one replicate by *cid1* deletion

Appendix F.1: List of genes >2-fold upregulated in one replicate by cid1 *deletion*

Systematic Name	Common Name	Expression Change		Mean Expression Change	Comments
		Run 1	Run 2		
SPNCRNA.389		5.51	0.99	3.25	intergenic ncRNA
SPAC186.06		5.18	0.98	3.08	subtelomeric
SPAC20H4.06c.RC		2.64			antisense to *SPAC20H4.06c*
SPBC1683.06c		2.59	1.72	2.16	subtelomeric
SPAC212.06c		2.58			subtelomeric, homology to *tlh1-4*
SPAC27E2.04c.RC	*mug155.RC*	2.53			antisense to *mug155*
SPAC23G7.13c		2.48	1.64	2.06	
SPNCRNA.139		2.43			intergenic ncRNA, adjacent to *str3*
SPAC1F8.03c	*str3*	2.31	1.35	1.83	
SPCC965.11c		2.24	1.68	1.96	
SPAC1002.17c	*urg2*	2.21	1.62	1.92	uracil regulatable gene
SPAC56F8.13		2.19	1.51	1.85	
SPAC4G9.22		2.18	1.78	1.98	
SPBC1683.12		2.15	1.85	2.00	
SPAC823.14.RC	*ptf1.RC*	2.15	1.68	1.92	antisense to *ptf1*
SPAC186.05c		2.11	0.44	1.28	subtelomeric
SPNCRNA.373		2.08	1.33	1.71	centromeric ncRNA
SPBC83.02c	*rpl37a-2; rpl43-2*	2.07	1.63	1.85	
SPBPB2B2.06c		2.06	1.18	1.62	subtelomeric
SPAC3F10.18c	*rpl41-2*	2.04	1.65	1.85	
SPRRNA.21.RC			16.84		antisense to rRNA.21
SPNCRNA.130.RC	*omt3.RC*		3.70		antisense to *omt3*
SPMIT.10	*atp9*	1.38	3.35	2.37	mitochondrially encoded
SPRRNA.21			3.05		rRNA.21
SPSNORNA.11.RC			3.04		antisense to snoRNA
SPAC4F10.17.RC		1.72	2.64	2.18	antisense to *SPAC4F10.17*
SPAPB8E5.05	*mfm1*		2.51		M-factor precursor
SPMIT.05	*cob1*	1.15	2.45	1.80	mitochondrially encoded
SPAC212.12	*SPNG7*	1.83	2.43	2.13	subtelomeric
SPAC20G4.07c.RC	*sts1.RC*	1.40	2.42	1.91	antisense to *sts1*
SPAC750.07c		1.98	2.36	2.17	subtelomeric
SPAP7G5.06	*per1*	1.84	2.31	2.08	
SPAPB2B4.01c.RC	*gpi12.RC*	1.18	2.25	1.72	antisense to *SPAPB2B4.01c*
SPMIT.06.RC		1.00	2.17	1.59	mitochrondrially encoded
SPAC25B8.13c	*isp7*	1.58	2.16	1.87	
SPSNORNA.08			2.15		snoRNA
SPMIT.08	*urfA*	1.08	2.12	1.60	mitochondrially encoded
SPAC26H5.09c.RC (*intron*)			2.10		antisense to intron

Appendix F.1 (continued): List of genes >2-fold upregulated in one replicate by cid1 *deletion*

Systematic Name	Common Name	Expression Change		Mean Expression Change	Comments
		Run 1	Run 2		
SPSNORNA.01			2.08		snoRNA
SPAC1399.01c		1.84	2.07	1.96	
SPAC1039.01		1.77	2.05	1.91	subtelomeric
SPBC13A2.04c	*ptr2*	1.40	2.04	1.72	
SPCC1223.09		1.65	2.04	1.85	
SPCC1020.09	*gnr1*	1.98	2.02	2.00	
SPMIT.07.RC	*atp6.RC*	0.93	2.01	1.47	antisense, mitochondrially encoded
SPMIT.01	*cox1*	1.05	2.01	1.53	mitochondrially encoded

Appendix F.2: List of genes >2-fold downregulated in one replicate by cid1 *deletion*

Systematic Name	Common Name	Expression Change		Mean Expression Change	Comments
		Run 1	Run 2		
SPAC27D7.13c.RC	*ssm4.RC*	0.49			antisense to *ssm4*
SPAC11H11.03c.RC		0.49			antisense to *SPAC11H11.03c.*
SPCC663.07c		0.49	0.67	0.58	pseudogene
SPAC26H5.12.RC	*rpo41.RC*	0.49	0.51	0.50	antisense to *rpo41*
SPNCRNA.200		0.49			ncRNA
SPCC1259.13	*chk1; rad27*	0.49	0.79	0.64	key checkpoint regulator
SPAP27G11.16		0.49	0.82	0.66	
SPAC26F1.11		0.48	0.62	0.55	
SPAC13G6.01c.RC	*rad8.RC*	0.48	0.93	0.71	antisense to *rad8*
SPAPB8E5.08		0.48	0.86	0.67	
SPCC364.06	*nap1*	0.48	0.68	0.58	
SPCC24B10.12	*cgi121*	0.48	0.89	0.69	
SPAC2F3.10.RC	*vps54.RC*	0.48			antisense to *vps54*
SPAC1002.06c	*bqt2*	0.47	1.74	1.11	telomere structure protein
SPAC16E8.06c.RC	*nop12.RC*	0.47	1.52	1.00	antisense to *nop12*
SPAC926.08c	*rpf2*	0.46	0.74	0.60	
SPAPB8E5.07c.RC	*rrp12*	0.46			adjacent to *mfm1*
SPNCRNA.240		0.46	0.92	0.69	intergenic ncRNA
SPCC63.03		0.46	0.88	0.67	
SPNCRNA.268		0.45	0.68	0.57	intergenic ncRNA
SPCC1393.11		0.45	0.92	0.69	
SPAC2C4.17c.RC		0.45	0.83	0.64	antisense to *SPAC2C4.17c*
SPAC12B10.10		0.45	1.00	0.73	sequence orphan
SPAC23C4.02	*crn1*	0.45	0.94	0.70	
SPCC1919.12c		0.44	0.79	0.62	
SPCC61.02	*spt3*	0.43	0.73	0.58	SAGA complex subunit
SPAC23H4.17c	*srb10*	0.43	0.74	0.59	
SPAC4F8.13c	*rng2*	0.43			
SPACUNK4.12c.RC	*mug138.RC*	0.42			antisense to *mug138*
SPBC19F8.05		0.42	0.79	0.61	sequence orphan
SPCC188.07	*ccq1*	0.42	0.61	0.52	telomere maintenance protein
SPBC3E7.02c	*hsp16*	0.42	0.86	0.64	
SPNCRNA.39	*prl39*	0.40			poly(A)$^+$ ncRNA
SPBC1718.05	*trs31*	0.40	0.98	0.69	
SPAC3A12.04c		0.38	0.87	0.63	RNAse P/MRP subunit
SPBC29A3.03c		0.37	0.81	0.59	
SPBC1685.13	*fhn1*	0.35	1.02	0.69	
SPNCRNA.537		0.35	0.86	0.61	
SPBC1604.20c	*tea2; klp4*	0.34	0.91	0.63	
SPAC1002.17c.RC	*urg2.RC*	0.28			antisense to *urg2*

Appendix F.2 (continued): List of genes >2-fold downregulated in one replicate by cid1 *deletion*

Systematic Name	Common Name	Expression Change		Mean Expression Change	Comments
		Run 1	Run 2		
SPAC23C4.11.RC	*atp18.RC*	0.21			antisense to *atp18*
SPBC8E4.10c		0.73	0.48	0.61	
SPBPB2B2.12c	*gal10*	0.63	0.45	0.54	subtelomeric
SPAC186.05c		2.11	0.44	1.28	subtelomeric
SPAC1B4.02c.RC		0.97	0.43	0.70	
SPNCRNA.289.RC			0.43		antisense to ncRNA
SPAC977.15		0.76	0.38	0.57	subtelomeric
SPAPYUG7.03c.RC	*mid2.RC*		0.38		
SPNCRNA.31.RC	*prl31.RC*		0.34		antisense to *prl31*
SPAC23H3.13c.RC (*intron*)	*gpa2.RC (intron)*		0.29		antisense to *gpa2* intron
SPNCRNA.313.RC			0.12		antisense ncRNA
SPNCRNA.61	*prl61*		0.05		subtelomeric ncRNA
SPNCRNA.61.RC	*prl61.RC*		0.04		antisense to *prl61*

. . . pation

- Dr. Frank-N-Furter