Structural basis of interaction between proline-rich sequence of PNRC2 and EVH1 domain of hDcp1a and its implications in mRNA decapping

Lai, Ethan Tingfeng

2012

Lai, E. T. (2012). Structural basis of interaction between proline-rich sequence of PNRC2 and EVH1 domain of hDcp1a and its implications in mRNA decapping. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/53728>

<https://doi.org/10.32657/10356/53728>

Structural basis of interaction between proline-rich sequence of PNRC2 and EVH1 domain of hDcp1a and its implications in mRNA decapping

Ethan Lai Tingfeng

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University in fulfillment of the requirement for the degree of Doctor of philosophy

2012

Acknowledgement

I would like to thank my supervisor, Prof Haiwei Song for giving me the opportunity to work with him. Thank you for your invaluable guidance, your tolerance towards my ignorance and continued support throughout the course of this project. I shall always remember the three important factors that you told me which are “motivation, sense of urgency and time planning”.

I would also like to express my gratitude to both past and present members of SHW lab. To Dr Portia, thank you for being such a selfless mentor. To Dr Sharon ling, thank you for sacrificing your personal time to proofread the manuscript. To Mr Lim Meng Kiat, thank you for lending your assistance whenever possible. To Dr Piao Shunfu, thank you for your insightful advice and tips.

I am grateful to my collaborators, Prof Yoon Ki Kim and Dr Hana Cho from University of Korea, and Prof Roy Parker from University of Arizona for their contributions to this study. Thanks also to beamline scientists at the European Synchrotron Radiation Facility (ESRF, France) as well as the staff from shared facilities of A*STAR especially DNA sequencing facility and X-ray crystallography facility. I would like to acknowledge A*STAR for awarding me with the AGS scholarship and BMRC for financially supporting the work presented in this thesis.

I would like to thank my friends, Eric, Hendrick, Yu Jinn, Teck Siong, Richard, Nijun, Shuping, Grace, Janice and Dawn for all the joys and laughters we shared. Most importantly, I would like to thank my parents, Albert and Joanna, and my mzzd, Xueyi, for their precious words of advice and unconditional love. I would never be able to complete this journey without all of you.

Table of Contents

ACKNOWLEDGEMENT	2
LIST OF FIGURES	5
LIST OF TABLES	7
SUMMARY	8
ABBREVIATIONS.....	10
1 INTRODUCTION.....	15
1.1 NONSENSE MEDIATED DECAY.....	15
1.1.1 <i>Functions of Nonsense mediated decay.....</i>	15
1.1.2 <i>Premature termination codon definition</i>	18
1.1.3 <i>Surveillance complex.....</i>	26
1.1.4 <i>Decay of mRNA after premature termination codon recognition</i>	31
1.1.5 <i>Subcellular localisation of Nonsense mediated decay.....</i>	35
1.1.6 <i>Nonsense mediated decay factors</i>	38
1.1.7 <i>Evolution of Nonsense mediated decay pathway.....</i>	45
1.1.8 <i>Involvement of Nonsense mediated decay factors in additional cellular process</i>	47
1.1.9 <i>Alternative complexes that trigger Nonsense mediated decay in human.....</i>	48
1.1.10 <i>Nonsense mediated decay and diseases.....</i>	49
1.2 DECAPPING.....	51
1.2.1 <i>Deadenylation</i>	53
1.2.2 <i>Transition to translationally repressed messenger ribonucleoprotein.....</i>	53
1.2.3 <i>Sequestration of mRNA into P bodies.....</i>	54
1.2.4 <i>Assembly of decapping complex.....</i>	56
1.2.5 <i>Decapping enzyme</i>	58
1.2.6 <i>Regulators of decapping.....</i>	64
1.2.7 <i>Alternative decapping enzymes</i>	71
1.2.8 <i>Decapping in the nucleus</i>	71
1.3 ENABLED/VASP HOMOLGY 1 DOMAIN.....	73
1.3.1 <i>Overall fold.....</i>	73
1.3.2 <i>Proline rich sequence ligand binding</i>	74
1.3.3 <i>Classification.....</i>	76
1.3.4 <i>Promiscuity of EVH1 domain.....</i>	80
1.3.5 <i>Therapeutics.....</i>	80
1.4 PROLINE RICH NUCLEAR RECEPTOR CO-ACTIVATOR 2.....	82
1.4.1 <i>Role as nuclear receptor coactivator.....</i>	82
1.4.2 <i>Role in Stau1-mediated mRNA decay.....</i>	82
1.4.3 <i>Role in adipogenesis and energy expenditure</i>	83
1.4.4 <i>Role in aromatase activation</i>	84
1.5 RATIONALE OF MY STUDY	85
2 MATERIALS AND METHODS.....	87
2.1 MOLECULAR CLONING	87
2.1.1 <i>Polymerase chain reaction (PCR) and analysis</i>	87
2.1.2 <i>Construct sub-cloning.....</i>	88
2.1.3 <i>Ligation</i>	89
2.1.4 <i>DH5α transformation and positive clone selection</i>	89
2.1.5 <i>Site directed mutagenesis</i>	90
2.2 PROTEIN EXPRESSION.....	92
2.2.1 <i>Rosetta (DE3) transformation.....</i>	92
2.2.2 <i>Small scale protein expression.....</i>	92
2.2.3 <i>Large scale Rosetta (DE3) cell culture.....</i>	92

2.3	PROTEIN PURIFICATION	93
2.3.1	Cell lysis	93
2.3.2	Purification of PNRC2 Δ NR-hDcp1aEVH1 complex.....	93
2.3.3	Purification of PNRC2 (FL)-hDcp1aEVH1 complex.....	94
2.3.4	Purification of His-PNRC2 Δ NR (W114A).....	95
2.3.5	Purification of his-hDcp1aEVH1.....	95
2.3.6	Purification of hDcp2 (1-245).....	95
2.4	PROTEIN ANALYSIS.....	96
2.4.1	Protein concentration analysis.....	96
2.4.2	Mass spectrometry MALDI-TOF.....	96
2.4.3	Edman degradation N-terminal sequencing.....	96
2.5	PROTEIN CRYSTALLISATION	97
2.6	STRUCTURE DETERMINATION	97
2.6.1	Data collection	97
2.6.2	Structure determination of the PNRC2 Δ NR-hDcp1aEVH1 complex.....	98
2.7	FUNCTIONAL ANALYSIS	100
2.7.1	Isothermal titration calorimetry.....	100
2.7.2	His Tag affinity pull down assay.....	100
2.7.3	Decapping assay.....	101
2.7.4	His tag affinity pull down assay.....	102
2.7.5	Western blot.....	102
3	RESULTS.....	103
3.1	PURIFICATION OF PROTEINS.....	103
3.2	PROTEIN ANALYSIS	106
3.2.1	Mass spectrometry MALDI-TOF.....	106
3.2.2	Edman degradation N-terminal sequencing.....	107
3.3	CRYSTAL STRUCTURE OF PNRC2 Δ NR- Dcp1aEVH1.....	111
3.3.1	Structure determination.....	111
3.3.2	Overall structure description	116
3.3.3	Structural comparison with previous classes of EVH1 domain.....	117
3.3.4	The hDcp1aEVH1-PNRC2 interface.....	119
3.3.5	A novel recognition mechanism of PNRC2 by hDcp1aEVH1.....	124
3.4	MUTATIONAL EFFECTS ON PNRC2-HDCP1A INTERACTION	127
3.5	THE NR BOX OF PNRC2 IS REQUIRED FOR BINDING TO THE PHOSPHORYLATED UPF1	128
3.6	PNRC2 AND HDCP1A SYNERGISTICALLY STIMULATE DECAPPING <i>IN VITRO</i>	131
4	DISCUSSION	134
4.1	YEAST TWO-HYBRID ASSAYS.....	134
4.2	SiRNA PNRC2 KNOCK OUT ASSAY	137
4.3	ROLE OF PNRC2-HDCP1A COMPLEX IN NMD	138
4.4	PNRC2 Δ NR-hDcp1aEVH1 COMPLEX IS A NOVEL CLASS OF EVH1 DOMAIN.....	139
4.5	CRITICAL RESIDUES FOR PNRC2 AND hDcp1aEVH1 INTERACTION	140
4.6	CONSERVED BINDING MODE OF PRS PROTEINS WITH DCP1	141
4.7	ASSEMBLY OF DISTINCT DECAPPING COMPLEX.....	142
4.8	SYNTHETIC GRADASES.....	145
4.9	hDcp1A IS A POTENTIAL CONTROL POINT THAT DETERMINES CELL SURVIVAL	147
4.10	CIKS IS A POTENTIAL INTERACTION PARTNER OF hDcp1A.....	149
4.11	C-TERMINAL EXTENSION OF DCP1A	150
5	CONCLUSION AND FUTURE DIRECTION	152
5.1	CONCLUSION.....	152
5.2	FUTURE DIRECTION.....	153
	REFERENCES.....	154

List of Figures

Figure 1. Faux 3' UTR Model.	19
Figure 2. Model of the EJC.	23
Figure 3. Pioneer round of translation.	25
Figure 4. Assembly of the surveillance complex in mammals.	28
Figure 5. Recruitment of SMG7 by phosphorylated Upf1.	30
Figure 6. NMD degradative activities.	33
Figure 7. Branched NMD pathway models.	34
Figure 8. P body formation in yeast and human.	38
Figure 9. Phosphorylation of Upf1 is catalysed by SMG1.	41
Figure 10. Phosphorylated Upf1 mediated repression of translation initiation.	43
Figure 11. PNRC2's role in NMD.	46
Figure 12. Overall scheme of decapping.	52
Figure 13. The assembly of distinct decapping complexes	57
Figure 14. Ribbon diagram of spDcp2n.	59
Figure 15. Surface representation of <i>S. cerevisiae</i> Dcp1.	61
Figure 16. Open and closed forms of Dcp1-Dcp2 complex.	63
Figure 17. The promoter proximal "torpedo" model for premature termination of pol II transcription.	72
Figure 18. PRS ligands binding to respective classes of EVH1 domains.	74
Figure 19. Flow chart of structure determination of the protein complex	99
Figure 20. Purification of recombinant PNRC2 Δ NR-hDcp1aEVH1 protein complex	103
Figure 21. Purification of recombinant PNRC2 (FL)-hDcp1aEVH1 protein complex	104
Figure 22. Purification of recombinant His-PNRC2 Δ NR (W114A)	104
Figure 23. Purification of recombinant his-hDcp1aEVH1 protein	105
Figure 24. Purification of recombinant human Dcp2 (1-245) protein	105
Figure 25. Peptide mass fingerprints of the PNRC2 and hDcp1aEVH1 proteins	106
Figure 26. Edman degradation N terminal sequencing	110
Figure 27. Domain organisation of hDcp1a and PNRC2 proteins	111
Figure 28. Crystals of PNRC2 Δ NR- Dcp1aEVH1	112
Figure 29. Evaluation of 1000 trials by SnB software	115
Figure 30. Structure of PNRC2 Δ NR- Dcp1aEVH1.	116
Figure 31. Comparison of the hDcp1aEVH1 protein with EVH1 domains from the ScDcp1, Mena, Homer and N-Wasp proteins. Residues of PRS binding site that interacted with PRS ligand were highlighted in yellow.	118
Figure 32. Sequence alignment analysis	121
Figure 33. Electrostatic charge distribution hDcp1aEVH1 binding interface with PNRC2	124
Figure 34. Surface views of the EVH1 interfaces	126
Figure 35. <i>In vitro</i> His tag affinity pull down assay	128
Figure 36. Binding isotherms of ITC titrations.	130
Figure 37. The hDcp1a and PNRC2 proteins synergistically stimulate the decapping activity of the hDcp2 protein.	132
Figure 38. Interactions of PNRC2 and hDcp1a with Dcp2.	133
Figure 39. Yeast two hybrid assay	135
Figure 40. Mutational effects on P-body localisation and NMD.	136
Figure 41. Immunoprecipitates (IP) of FLAG-Dcp2 using the extracts of HEK293T cells depleted of endogenous PNRC2.	137

Figure 42. Superposition of PNRC2 Δ NR-hDcp1aEVH1 complex to SpDcp1-SpDcp2 interface.	144
Figure 43. Sequence alignment of PNRC with PNRC2	148
Figure 44. Sequence alignment of Ciks with PNRC2.	149

List of Tables

Table 1. Generation of point mutations of residues lining PNRC2 and Dcp1aEVH1 interface.....	91
Table 2. Selenium atom sites found by the software SnB	113
Table 3. Data Collection and refinement statistics of SeMet PNRC2 Δ NR-hDcp1aEVH1	114
Table 4. Binding network of PNRC2/hDcp1aEVH1 interface	122

Summary

Nonsense-mediated mRNA decay (NMD) is an important mRNA surveillance system that maintains the integrity of transcripts by removing aberrant mRNAs harbouring premature termination codons (PTCs). In the absence of NMD, truncated proteins with dominant negative or potentially deleterious gain of function activity may be expressed from these erroneous mRNAs. In addition to its function as a quality control for mRNAs, NMD is involved in the regulation of 3 – 10% of the *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Homo sapiens* transcriptomes.

PTC recognition in NMD requires the cross talk between terminating ribosome stalled at stop codon and downstream cis-acting signal that is not conserved across species. In *S. cerevisiae* and *D. melanogaster*, the faux 3' untranslated region (UTR) model has been proposed to be the main mechanism in recognising PTC. In mammals, PTC recognition is mediated through the interaction of the SMG1, Upf1, eukaryotic Release factor 1 and 3 (SURF) complex on the terminating ribosome at a PTC and downstream exon junction complex (EJC) during the pioneer round of translation.

Upon recognition of PTC, an evolutionarily conserved surveillance complex consisting of the Upf1, Upf2 and Upf3 proteins is assembled. In mammals, the assembly of the surveillance complex also activates a phosphoinositide-3-kinase related protein kinase, SMG1, which will phosphorylate the Upf1 protein. Till date, three mechanisms of how phosphorylated Upf1 protein leads to decay of mRNAs have been proposed. First, the phosphorylated Upf1 protein recruits the SMG7 protein and C terminal of SMG7 protein targets the associated mRNA for decay through an undetermined mechanism. Second, the phosphorylated Upf1 protein recruits

hDcp1a or Dcp2 proteins to decap mRNA for its subsequent 5'-3' decay. Third, the hyperphosphorylated Upf1 protein associates with the nuclear receptor coregulatory protein 2 (PNRC2), which targets the mRNA to the P body through the interaction between the PNRC2 and hDcp1a proteins. Subsequently, decapping and 5' - 3' decay of mRNA occurs in the P body. As the three mechanisms are inferred from the results of coimmunoprecipitation or two-hybrid interactions, no direct interaction between the hyperphosphorylated Upf1 protein and decapping enzymes has been demonstrated till date.

Here we present the crystal structure of hDcp1a protein in complex with the PNRC2 peptide. The proline-rich region of the PNRC2 peptide is bound to the EVH1 domain of the hDcp1a protein while isothermal titration calorimetry study demonstrates that NR-box of the PNRC2 protein mediates the direct interaction with hyperphosphorylated Upf1 protein. The mode of the PNRC2 protein interaction with the hDcp1a protein is distinct from those observed in other classes of EVH1/proline-rich ligands interactions. Additionally, PNRC2 mutagenesis study performed by our collaborator showed that disruption of the interaction of the PNRC2 protein with the hDcp1a protein abolishes its P-body localization and ability to promote mRNA degradation when tethered to mRNAs. Furthermore, the PNRC2 protein acts in synergy with the hDcp1a protein to stimulate the decapping activity of the Dcp2 protein by bridging the interaction between the hDcp1a and Dcp2 proteins. The formation of a novel PNRC2-hDcp1a-Dcp2 decapping complex suggests that the PNRC2 protein is a decapping coactivator in addition to its adaptor role in NMD.

Abbreviations

AGO	Argonaute
ARE	AU rich elements
ATP	Adenosine triphosphate
AU	Asymmetric unit
CBC	Cap binding complex
CCP4	Collaborative computational project no. 4
CNS	Crystallographic and NMR system
DcpS	Scavenger decapping enzyme
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
Edc	Enhancer of decapping
eIF4E	Eukaryotic translation initiation factor
eRF	Eukaryotic release factor
EJC	Exon junction complex
EVH1	Enabled/VASP homology 1
FPLC	Fast protein liquid chromatography
GST	Glutathione S-transferase

hDcp1a	Human decapping enzyme 1a
Hedls	Human enhancer decapping large subunit
HEK	Human embryonic kidney
Hsp	Heat shocked protein
IP3R	Inositol 1,4,5 triphosphate receptor
IPTG	Isopropyl β -D-thiogalactopyranoside
IRES	Internal ribosome entry site
ITC	Isothermal titration calorimetry
Kd	Dissociation constant
Klf2	Kruppel-like factor 2
LB media	Luria-Bertani media
M ⁷ G	5' ^N 7 methylguanosine
MALDI-TOF	Matrix-assisted laser desorption/ionisation-time of flight
Mena	Mammalian enabled
mGluR	Metabotropic glutamate receptors
mRNA	Messenger ribonucleic acid
mRNP	Messenger ribonucleoprotein
NDP	Nucleoside diphosphate
NMD	Nonsense mediated decay

NR	Nuclear receptor
NTP	Nucleoside triphosphate
OD	Optical density
P Body	Processing body
PABPC	PolyA binding protein cytoplasmic
PARN	PolyA ribonuclease
PCR	Polymerase chain reaction
PIP	Phosphatidyl inositol-4-phosphate
PIP2	Phosphatidyl inositol-4,5-phosphate
PMSF	Phenylmethanesulfonyl fluoride
PNRC2	Proline rich nuclear receptor co-activator 2
PPII	Poly proline II
PP2A	Protein phosphatase 2A
PVDF	Polyvinylidene fluoride
QRt-PCR	Quantitative real time PCR
RyR	Ryanodine receptor
PRS	Proline rich sequence
PTB	Phosphotyrosine binding domain
PTC	Pre-Termination codon

RAS	Rat sarcoma
R.M.S.D	Root mean squared deviation
RNA	Ribonucleic acid
RRM	RNA recognition motif
RUST	Regulated unproductive splicing translation
SAD	Single wavelength anomalous dispersion
SDS PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeMet	Seleno-L-methionine
SH3	SRC homology 3
siRNA	Small interfering RNA
SMD	Stau1 mediated decay
SMG1	Morphogenetic effect on Genitalia factor 1
Spred	Sprouty-related protein with a EVH1 domain
SQ motif	Serine/glutamine motif
SURF	SMG-1, UPF1, eukaryotic release factor 1 and eukaryotic release factor 3 complex
TCEP	Tris [2-carboxyethylphosphine] hydrochloride
TERRA	Telomeric repeat containing RNA
TGF	Transforming growth factor

TLC	Thin layer chromatography
TPR	Tetratricopeptide
TSS	Transcription start site
Upf1	Up-frameshift factor 1
UTR	Untranslated region
UV	Ultra violet
VDW	Van der Waals
WASP	Wiskott-Aldrich syndrome protein
WIP	WASP interacting protein
WT	Wild type

1 Introduction

1.1 Nonsense mediated decay

1.1.1 Functions of Nonsense mediated decay

1.1.1.1 Quality control mechanism

The central dogma framework of molecular biology begins from the transcription of the sequence information from DNA to RNA and its subsequent translation into the proteome. As such, a typical eukaryotic gene expression pathway comprises of coordinated and regulated steps from transcription, 5' cap formation, mRNA splicing, polyadenylation, mRNA export, translation and subsequent mRNA degradation. It is established that all the steps are subjected to regulation, which serves to control the efficacy and fidelity of the production of proteins. Given the complexity of the gene expression system, errors can occur at any of the intermediate steps. Thus, eukaryotic cells have evolved elaborate mRNA surveillance mechanisms to detect and degrade erroneous mRNA to ensure the fidelity of its mRNAs (Chang et al., 2007).

One of the most well studied and characterised mechanisms is known as the nonsense-mediated decay (NMD). This mechanism is evolutionarily conserved in a wide variety of organisms and it serves to identify premature termination codon (PTC) on erroneous mRNA and rapidly target them to the mRNA decay machinery (Amrani et al., 2006; Behm-Ansmant et al., 2007b; Chang et al., 2007). It is important for cells to eliminate mRNAs that prematurely terminate to prevent the expression of truncated proteins with dominant-negative or potential deleterious gain-of-function activities in the cell.

Erroneous generation of PTC in transcript can occur at either the DNA or RNA level. At the DNA level, generation of PTC in a transcript can be attributed to base substitution which changes an amino acid encoding codon into any of the three termination codons, frameshifting due to insertion or deletion of random nucleotide and mutation that affect splicing signals which lead to defective splice variants (Muhlemann et al., 2008). Moreover, in mammals, PTC containing

mRNAs transcribed from unproductive rearrangement of immunoglobulin and T cell receptor genes also forms an important physiological class of NMD substrates (Chang et al., 2007; Rehwinkel et al., 2006; Weischenfeldt et al., 2008). Accordingly, during lymphocyte maturation, it was observed that two of three rearrangements of the V, D and J segments lead to frameshift, which results in PTC generation.

At the RNA level, transcription errors of the RNA polymerase II may introduce PTC in transcript. Nevertheless, based on a misincorporation rate of the RNA polymerase II at 10^{-5} per nucleotide and an average of 10^4 coding nucleotides in a gene, it is estimated that transcription errors only occur in 0.05% - 0.5% of transcript. In contrast, the error rate in alternatively spliced mRNAs is much higher. Computational analysis of human EST databases showed that 60%-70% of human pre-mRNAs are alternatively spliced and 45% had a splice form that is a target of NMD (Lewis et al., 2003). Therefore, the result implies that one third of all human coding genes produce mRNAs with PTCs which are potential NMD substrates. In addition, in yeast, mRNAs with leaky scanning ribosome may result in the skipping of the original AUG sequence and subsequent initiation of internal downstream out of frame AUGs which are also observed to be a substrates for NMD.

1.1.1.2 Regulation of gene expression

Given the high level of conservation of NMD across species, it is unlikely that rare *de novo* nonsense mutations can provide sufficient selective pressure for evolutionary maintenance of NMD. Hence, it is postulated that the scope of NMD extends beyond quality control of mRNAs. Accordingly, NMD is shown to play a role in post-transcriptional regulation of physiological transcripts that mimic nonsense transcripts. Indeed, gene expression profiling of *S. cerevisiae*, *D. melanogaster*, and human cells depleted of NMD factors has provided evidence that NMD is involved (directly or indirectly) in the regulation of 3 – 10% of the transcriptome (He et al., 2003; Mendell et al., 2004; Rehwinkel et al., 2005; Rehwinkel et al., 2006; Weischenfeldt et al., 2008; Wittmann et al., 2006). The range of the transcriptome affected may be underestimated because the arrays used for analysis may monitor only a fraction of available cellular transcripts and the NMD factors may not be completely knocked down (5%-25%)

during analysis (Isken and Maquat, 2007; Mendell et al., 2004). Nevertheless, these studies showed that the physiological transcripts being regulated by NMD include mRNAs with programmed frameshifting, mRNAs regulated by stop codon read-throughs, mRNAs with selenocysteine codons, mRNAs with upstream open reading frames in the 5' UTRs, mRNAs with exceptionally long 3' UTRs, mRNAs with introns in 3' UTRs and mRNAs encoded by transposable elements. Further analysis showed that these mRNAs are involved in broad range of cellular processes like energy generation, cellular transport and organisation, transcription, cell proliferation, cell fate, DNA processing, cell defense and rescue, protein synthesis, cell cycle, telomere maintenance and metabolism (He et al., 2003; Rehwinkel et al., 2006). More importantly, the repertoire of genes regulated by NMD is not conserved across *S. cerevisiae*, *D. melanogaster*, and *H. sapiens* (Rehwinkel et al., 2005). For example, it is observed that inhibition of the NMD pathway is not lethal in *S. cerevisiae* (Hodgkin et al., 1989; Leeds et al., 1991). Similarly, disruption of NMD in *Caenorhabditis elegans* also results in viable nematodes with defects in male bursa and hermaphrodite vulva (Hodgkin et al., 1989). Conversely, Up-frameshift protein 1 (Upf1) null mice die in early embryonic development, which suggest that Upf1 is required for cell viability in mice (Medghalchi et al., 2001).

Interestingly, NMD plays a role in the auto-regulatory gene expression circuit known as “regulated unproductive splicing and translation” (RUST) (Lewis et al., 2003). This expression circuit proposes that pre-mRNA can undergo alternative splicing to generate a productive mRNA encoding the full-length protein and a non-productive mRNA with a PTC. At a threshold concentration, the protein binds to its own pre-mRNA and alter its splicing to favour the generation of a non-productive mRNA, which is rapidly degraded by NMD. In support of the model, *C. elegans* ribosomal proteins L3, L7a, L10a, L12 and SR proteins (SRp20 and SRp30), as well as the mammalian SR protein SC35, are regulated through RUST (Mitrovich and Anderson, 2000; Morrison et al., 1997; Sureau et al., 2001). This suggests that the coupling of NMD with alternative splicing to regulate gene expression post-transcriptionally may be conserved across species.

1.1.2 Premature termination codon definition

A key aspect of the study of NMD is to understand the mechanisms employed to discriminate PTCs from natural stop codons, where only the former will elicit mRNA decay. It is shown that the definition of PTC is a translation dependent process that relies on cross talk between the terminating ribosome stalled at the stop codon, and the downstream cis-acting signal which is not conserved across species (Amrani et al., 2006; Behm-Ansmant et al., 2007b; Chang et al., 2007). The interaction between the ribosome and the cis-acting signal conveys positional information to the ribosome (Conti and Izaurralde, 2005). Consistent with the model, it is observed that inhibiting translation leads to stabilisation of mRNAs, which are targeted for NMD. Furthermore, it is proposed that the diversity of mRNAs regulated by NMD in different species may be due to the differences in their respective PTC definition mechanism (Rehwinkel et al., 2005).

1.1.2.1 Faux 3' Untranslated region

In *S. cerevisiae*, *D. melanogaster* and *C. elegans*, the cis-acting signal is a faux 3' untranslated region (UTR) (**Figure 1**). The faux 3' UTR model proposed that the premature translation termination is intrinsically aberrant because the stop codon is not in appropriate context (Amrani et al., 2006; Behm-Ansmant et al., 2007a). For a translation termination event to be defined as normal termination, the translating ribosome needs to terminate in close proximity to the poly(A) tails such that it is flanked by a proper 3' UTR marked with specific proteins like the poly(A)-binding protein cytoplasmic (PABPC). Subsequently, the terminating ribosome is able to interact with these 3' UTR bound proteins to increase the efficiency of translation termination which involves the release of polypeptide chain and dissociation of the two ribosomal subunits (Behm-Ansmant et al., 2007a). Conversely, if the terminating ribosome stalls on an upstream PTC, the spatial distance between the ribosome and poly(A) tail is too large for potential interaction between the ribosome and the 3' UTR associated protein. This results in inefficient translation termination. In support of the proposed model, the use of toe print analysis for transcript undergoing translation termination reveals that ribosomes are released less efficiently on a

Figure 1. Faux 3' UTR Model.

In both *S. cerevisiae* and *D. melanogaster*, newly synthesised and steady state mRNAs are NMD substrates if they contain a faux 3' UTR. Ribosomes terminating at a PTC is inherently unstable due to the abnormally long distance between the termination site and the poly(A) bound PABPC. Thus, this results in inefficient translation termination, which leads to the recruitment of Upf1, Upf2 and Upf3 proteins. Subsequently, the Upf factors recruit and activate mRNA decay machinery. Figure is modified from Isken and Maquat, 2007

PTC, as compared to those that terminate at a normal stop codon (Amrani et al., 2006).

1.1.2.1.1 Poly (A) binding protein cytoplasmic

The importance of the PABPC in defining a normal stop codon is shown when artificial tethering of the PABPC at 3' end of a PTC suppresses NMD and promotes 5-10 fold increase in *S. cerevisiae* mRNA abundance (Amrani et al., 2006). A similar but independent study using *D. melanogaster* mRNA also shows that PABPC tethered at the 3' end of a PTC stabilised the mRNA. Moreover, the distance between the PTC and PABPC tethering site shares an inverse relationship with efficiency of the NMD suppression effect (Behm-Ansmant et al., 2007a). This suggests that the proximal PABPC binding defines natural stops in both *D. melanogaster* and *S. cerevisiae*.

Additionally, the PABPC consists of 4 non identical RNA recognition motifs (RRM) and it was observed that RRM1 and RRM2 are crucial for PABPC's high affinity binding to the poly(A) tail. A possible mechanism of NMD suppression is proposed whereby the association of the C-terminal domain of PABPC with eukaryotic release factor 3 (eRF3) facilitates ribosome release during translation termination (Behm-Ansmant et al., 2007a; Cosson et al., 2002; Hoshino et al., 1999; Kozlov et al., 2004). In support of the suggested model, tethering of the eRF3 downstream of a PTC also abolish NMD in *S. cerevisiae* and *D. melanogaster* (Behm-Ansmant et al., 2007b). In cells depleted of PABPC, NMD is suppressed. Interestingly, the translation efficiency is not affected (Behm-Ansmant et al., 2007a). Thus, this unexpected finding suggests that in addition to providing positional information for the PTC definition, PABPC may play a more direct role in NMD.

1.1.2.1.2 3' Untranslated region length

In *D. melanogaster*, increasing the 3' UTR length of natural stop triggers NMD (Behm-Ansmant et al., 2007a). Conversely, reducing the 3' UTR length causes a PTC to become an inefficient initiator of NMD (Behm-Ansmant et al., 2007a). Additionally, an independent study showed that *S. cerevisiae* mRNA with errors in 3' end processing became NMD substrates as a result of their long 3' UTRs (Muhlrad and Parker, 1999). Thus, the length of the 3' UTR may influence

the efficiency of NMD. However, it was observed that multicellular organisms have a large proportion of naturally occurring mRNAs with long 3' UTRs which are not NMD substrates (Behm-Ansmant et al., 2007a). In order to explain this, Isken and Maquat, (2007) suggested that in addition to considering the 3' UTR length, higher order 3' UTR structures must be taken into account when predicting the type of nonsense codon that will elicit NMD. Furthermore, it is plausible that these 3' UTRs have evolved features to suppress NMD. It was observed that the fusion of hsp70 3' UTR to a PTC containing mRNA confer NMD resistance to the transcript that elicit NMD under normal conditions (Behm-Ansmant et al., 2007b). Based on the findings, a novel mode of translation-dependent posttranscriptional gene regulation with NMD involvement was suggested. The model proposed that a mRNA is stabilised if the 3' UTR 3-dimensional conformation allows close proximity between its termination codon and its poly(A) tail. More importantly, this 3' UTR conformation can be altered by protein-protein interactions and protein-RNA interactions that can be spatially and temporally regulated in response to environmental cues (Eberle et al., 2008; Stalder and Muhlemann, 2008).

It is suggested that the NMD resistance conferred by hsp70 3' UTR to its transcript is due to the presence of several AU-rich elements that may play a role in rapid decay. Hence, it is postulated that the intrinsic instability of mRNA prevents further acceleration of its degradation caused by NMD. To validate the postulation, GW182, a protein which strongly reduces half lives of mRNA by promoting deadenylation and decapping, was co-expressed with a PTC containing mRNA that elicits NMD under normal conditions. Consistent with the postulation, it was found that the mRNA was no longer subjected to NMD upon co-expression with GW182 (Behm-Ansmant et al., 2007b). Additionally, an alternative scheme of conferring NMD resistance to transcript is proposed and it involves the sequence elements of a long 3' UTR forming secondary structures via RNA-RNA or RNA-protein interactions. As a result of the 3' UTR secondary structure, PABPC can be brought into close proximity with the stop codon and inhibit NMD. Alternatively, mRNA stabilising sequence elements may be present to antagonise NMD directly. For example, *S. cerevisiae* PGK1, GCN4 and YAP1 mRNAs have been reported to stop NMD when positioned downstream of a PTC (Ruiz-Echevarria and Peltz, 2000). This is due to the binding of protein, PUB1, onto the RNA. While the mechanism employed by PUB1 to antagonise NMD is

still unclear, it suggests that NMD can regulate gene expression in response to the levels of PUB1, which may be regulated by external stimuli.

The role of 3' UTR length and PABPC in PTC definition in mammals has not been demonstrated directly. Neu-Yilik et al., (2001) showed that PABPC is dispensable for NMD in human cells when a PTC is located upstream of an exon-exon boundary. In addition, it was reported that the exon junction complex (EJC) proteins components, which are essential for NMD in mammals, have no clear orthologs in *S. cerevisiae* (Gatfield et al., 2003). Thus, this suggests that *S. cerevisiae* PTCs can be defined independently of exon boundaries. Furthermore, in *D. melanogaster*, orthologs of EJC components like Barentsz, eukaryotic initiation factor (eIF) 4AIII, Y14 and MAGOH are dispensable for NMD (Gatfield et al., 2003; Le Hir et al., 2000; Palacios et al., 2004). Taken together, this suggests a divergence between the NMD pathway in mammals and invertebrates.

1.1.2.1.3 Loosely defined downstream elements (DSE)

In *S. cerevisiae*, an additional NMD mechanism, which involves DSE, has been reported. The DSEs reside in the translational reading frame and they bind to DSE-binding proteins, such as Hrp1, which cause inefficient translation termination that subsequently lead to the recruitment of NMD factors (Gonzalez et al., 2000; Wang et al., 2006). Consistent with the results, Hrp1 has been shown to coimmunoprecipitate with Upf1 and elicit NMD (Gonzalez et al., 2000).

1.1.2.2 Exon junction complex

In mammals, the cis-acting signal is the EJC (**Figure 2**) which is a multiprotein complex deposited by the spliceosome at the exon-exon junction during splicing at 20 - 24 nucleotides upstream of a splice junction (Behm-Ansmant et al., 2007a; Chang et al., 2007; Le Hir et al., 2000). The EJC remains bound to the mRNPs after its export from the nucleus to the cytoplasm, and its components essential for NMD include Y14, MAGOH, eIF4AIII, Barentsz and RNPS1 (Conti and Izaurralde, 2005; Lejeune and Maquat, 2005). In general, a stop codon is defined as a PTC when it occurs more than 50 - 55 nucleotide (nt) upstream of an exon-exon boundary (Nagy and Maquat, 1998). Nevertheless, there are exceptions to this 50 - 55 nt rule. For example, the 5' PTC in β -globin

first exon fails to elicit NMD despite its localisation at >50-55 nt upstream of Exon1-exon2 junction (Danckwardt et al., 2002; Romao et al., 2000; Silva et al., 2006). The reason for this discrepancy remains to be elucidated.

In support of the requirement of EJC to trigger NMD in mammals, it is observed that mammalian genes lacking introns are immune to NMD (Brocke et al., 2002; Maquat and Li, 2001). Furthermore, the tethering of either the EJC components RNPS1 or the Y14-MAGOH complex downstream of a normal stop codon elicit NMD (Fribourg et al., 2003; Gehring et al., 2003; Lykke-Andersen et al., 2001). Additionally, down regulation of EJC components like Y14, Barentsz, RNPS1 and eIF4AIII inhibits NMD (Ferraiuolo et al., 2004; Gehring et al., 2005; Gehring et al., 2003; Palacios et al., 2004). Taken together, the results suggest that EJC components are necessary for NMD in mammals.

Figure 2. Model of the EJC.

The EJC is deposited at exon exon junction after splicing and its core consists of eIF4AIII, MAGOH, Barentsz and Y14 where eIF4AIII and Barentsz bind directly to mRNA. Importantly, Y14 binds to eIF4AIII and Upf3, which further recruits Upf2. Upon PTC recognition, Upf1 binds to Upf2-Upf3 to form the surveillance complex.

1.1.2.2.1 Pioneer round of translation for mammalian cells

It has been proposed that mammalian PTC-containing transcripts undergo NMD during its pioneer round of translation (**Figure 3**) (Belgrader et al., 1994; Cheng and Maquat, 1993; Sun et al., 2000). Degradation of aberrant transcripts during the pioneer round of translation prevents multiple rounds of translation, which may lead to the production of high levels of truncated protein with potentially deleterious activities (Chang et al., 2007). The model suggests that the translating ribosomes displace downstream EJCs on the transcript to prevent NMD in the absence of PTCs (Chang et al., 2007). Conversely, in the presence of PTCs, ribosomes stall prematurely and downstream EJCs are not displaced from the erroneous transcript.

The discovery of the cap binding complex (CBC), consisting of CBP80 and CBP20, as a unique marker of newly synthesised mRNA and its subsequent substitution by eIF4E for bulk translation of mRNA allows discrimination between new and old mRNA. Through comparison of the properties in either CBC-bound or eIF4E-bound mRNA, the events occurring on newly synthesised mRNA and old mRNA undergoing bulk decay can be dissected (Fortes et al., 2000; Gorlich et al., 1996; Ishigaki et al., 2001; Izaurralde et al., 1995; Lewis and Izaurralde, 1997; Visa et al., 1996). These comparisons have yielded several evidences to support NMD occurrence at the pioneer round of translation. First, the presence of PTCs on mRNAs decreases only the level of CBC-bound mRNA. The levels of mRNAs, which are bound to eIF4E remain unaffected (Ishigaki et al., 2001). Second, EJC and NMD factors like SMG-1, Upf2, Upf3a, Upf3b and eIF4AIII are present in mRNPs containing CBC but not those containing eIF4E (Chiu et al., 2004; Ishigaki et al., 2001; Kashima et al., 2006; Lejeune et al., 2002). Third, the large subunit of CBC, CBP80, promotes NMD by recruiting Upf1 to the transcript and enhances the interaction between Upf1 and Upf2 (Hosoda et al., 2005). Taken together, the results propose that mammalian cells utilise a CBC-dependent pioneer round of translation for quality control of mRNAs. After the degradation of erroneous mRNAs, the remaining transcripts with bound CBC are substituted by eIF4E which mediates steady-state rounds of translation for bulk protein production.

Figure 3. Pioneer round of translation.

Mammalian NMD occurs after pre-termination codon (PTC) recognition during the pioneer round of translation, which is dependent on the cap binding complex (CBC). In order for NMD to initiate, Upf1 must bind to Upf2-Upf3 complex which associate with the EJC, to form the conserved surveillance complex. For an aberrant transcript (right panel), there is an EJC deposited downstream of the PTC which allow recruitment of Upf1 by CBC. Subsequently, Upf1 bind with eRF 1 and 3 and interact with Upf2 on the EJC. Conversely, the normal transcript (left panel) prevents the onset of NMD as all the EJCs are upstream of the normal termination codon. Thus, before Upf1 is recruited, the EJCs are displaced by a pioneer round of translation by the ribosome, which effectively eliminate the cis signal required to trigger NMD. After the proofreading step, eIF4E replaces CBC at the 5' end of the normal transcript and proceed to bulk translation.

While a pioneer round of translation is vital for NMD in aberrant mammalian transcripts, such a mechanism is absent in *S. cerevisiae*. This is shown in a study where a galactose inducible promoter was stimulated to rapidly induce expression of the Upf1, Upf2 or Upf3 proteins in an Upf null yeast strain. Following the expression of the Upf proteins, NMD of preexisting transcripts with PTCs is triggered (Maderazo et al., 2003). In addition, NMD of *S. cerevisiae* target newly synthesised (Cbc1-Cbc2 bound) and old mRNA (eIF4E bound) (Gao et al., 2005). A possible explanation for the absence of a similar mechanism is that NMD in *S. cerevisiae* is inefficient and thus, requiring multiple rounds of translation (200 or more) to degrade a significant amount of PTC bearing transcripts (Keeling et al., 2004).

1.1.3 Surveillance complex

A surveillance complex is assembled to bridge a premature termination event at the PTC of a transcript to mRNA degradation. This bridging role of the surveillance complex is accomplished by interaction with both eukaryotic translation termination factors and the general decapping machinery (Amrani et al., 2006; Conti and Izaurralde, 2005; Lejeune and Maquat, 2005). The surveillance complex is made up of a conserved trimeric complex consisting of trans-acting factors Upf1, Upf2, and Upf3. The trimeric Upf1-Upf2-Upf3 protein complex forms the conserved core of the NMD machinery and it is assembled in all organisms despite the differences in the PTC definition mechanisms (Conti and Izaurralde, 2005; Lejeune and Maquat, 2005). Gene knockdowns in *D. melanogaster* and genetic studies in *S. cerevisiae* and *C. elegans* have also shown that the Upf1, Upf2 and Upf3 proteins are essential for NMD (Gatfield et al., 2003; Lelivelt and Culbertson, 1999; Pulak and Anderson, 1993). Furthermore, three types of single gene deletion mutants involving Upf1, Upf2 and Upf3 lead to the stabilisation of PTC-containing mRNAs. Accordingly, gene expression profiles of entire transcriptomes had been analysed for *S. cerevisiae* and it was revealed that the three types of single gene deletion mutants of *S. cerevisiae* shows similar expression profile (He et al., 2003). Similarly, *D. melanogaster* cells with depleted Upf1 or Upf2 also displayed matching expression profiles (Rehwinkel et al., 2005) and thus, providing further evidence that Upf1, Upf2 and Upf3 are essential for NMD across species.

1.1.3.1 Stepwise assembly of surveillance complex in mammals

The assembly of surveillance complex in mammals is relatively well characterised (**Figure 4**). In Y14 or Upf2 depleted cells, it was observed that the association of the eRF1 and eRF3 proteins with unphosphorylated UPF1 and Morphogenetic effect on Genitalia-1 (SMG1) is enhanced (Kashima et al., 2006). Furthermore, immunoprecipitation assay using the N-terminal fragment of SMG1 (1-2223 amino acids (aa)) showed that it associates with Upf1 (Kashima et al., 2006). Taken together, it was proposed that upon stalling of the translating ribosome at a PTC, release factors eRF1 and eRF3 assemble on the ribosome and recruit the unphosphorylated Upf1 which then binds to the N - terminal domain of SMG1 kinase. The interaction of SMG1-Upf1-eRF1-eRF3 forms the SURF complex and its formation is independent from the EJC assembly (Chang et al., 2007; Czaplinski et al., 1998; Kashima et al., 2006). Additionally, this model also proposed that the definition of PTC and subsequent trigger of NMD is dependent on the competition between Upf1 and PABPC to bind eRF3 on the terminating ribosome (Singh et al., 2008).

The immunoprecipitation studies done on the cytoplasmic human cell extracts using anti-SMG1 antibodies in the presence of RNases, revealed SMG1 co-precipitates with the Upf1, Upf2, Upf3a/b, Y14, MAGOH, eIF4AIII and SMG7 proteins (Kashima et al., 2006). In order to identify the domain of SMG1 which are involved in the interaction of each individual protein components, subsequent immunoprecipitation studies of the C-terminal SMG1 (2068-3657 aa) have been performed. It was found that C-terminal SMG1 directly associates with only Upf2 (Kashima et al., 2006). Thus, SMG1 interacts with both Upf1 and Upf2 via its N-terminal and C-terminal domain respectively. These data suggest that SMG1 forms a second complex with the EJC components Upf3b, Y14, MAGOH and eIF4AIII in a Upf2 - dependent manner (Kashima et al., 2006). Studies also showed that Upf2 mutants which are defective in binding Upf3, do not coimmunoprecipitate with the Y14 protein. Furthermore, Kim et al., (2005) showed that the tethering of the Upf3 protein to a mRNA more than 50 - 55 nt downstream of a normal termination codon elicits NMD in a Upf2-dependent manner, and the tethering of the Upf2 protein at the same site elicits NMD in a Upf1-dependent but Upf3-independent manner. This lends support to the model where the temporal order of Upf association with EJC, is Upf3, Upf2

Figure 4. Assembly of the surveillance complex in mammals.

During the pioneer round of translation (step 1), eRF1 and eRF3 (labeled as eRF) binds to the stalled ribosome at the premature stop codon (step 2). Due to inefficient translation termination, Upf1 is recruited (step 2) which in turn binds to SMG1 to form the SURF complex (step 3). The SURF complex interacts with downstream EJC to form the conserved surveillance complex (Upf1, Upf2 and Upf3b) which triggers the phosphorylation of Upf1 (step 4) by SMG1. Phosphorylation of Upf1 triggers the decay of aberrant mRNA as well as the dissociation of the SURF complex and 40S and 60S ribosomal subunits (step 5).

and Upf1. Thus, the authors propose a stepwise surveillance complex assembly mechanism, whereby Upf3b associates with Y14 of EJC deposited on the mRNA after splicing in the nucleus (Gehring et al., 2003), the Upf3b protein then further recruits Upf2 (Kadlec et al., 2006; Kadlec et al., 2004; Serin et al., 2001) after mRNA export via the nuclear pore complex to the cytoplasm. Subsequently, using the mRNA as a loading platform, interaction of Upf1 and SMG1 in the SURF complex, with Upf2 in the EJC downstream of PTC, forms a molecular bridge between the SURF complex and the EJC.

The assembly of the SURF-EJC complex activates SMG1, which leads to phosphorylation of Upf1 protein (**Figure 4**) (Behm-Ansmant et al., 2007b; Chang et al., 2007; Kashima et al., 2006; Lejeune and Maquat, 2005). Consistent with the model, knocking down of EJC components like Upf2, Upf3b or Y14, strongly inhibits Upf1 phosphorylation (Kashima et al., 2006). Additionally, it was observed that phosphorylation of the Upf1 proteins leads to the loss of SURF complex (Kashima et al., 2006). Conversely, overproduction of the kinase inactive mutant form of SMG1 causes the accumulation of the SURF complex. As Upf2 and eRF1 are Upf1 binding competitors (Kashima et al., 2006), it is likely that eRF1 and eRF3 dissociate from Upf1 and SMG1 after Upf1 binds Upf2.

While the molecular mechanism is revealed for Upf1 phosphorylation, the function of Upf1 phosphorylation remains elusive. It has been proposed that the phosphorylated Upf1 protein recruits SMG5 - 7 and the C terminal domain of SMG7 will target the bound mRNA for decay via a mechanism that is still not determined (**Figure 5**) (Unterholzner and Izaurralde, 2004). The role of SMG7 proteins in NMD appears to be specific to higher eukaryotes. In *D. melanogaster*, there is no ortholog of SMG7. However, orthologs of SMG5 and SMG6 proteins are important in effecting NMD (Gatfield et al., 2003). In *S. cerevisiae*, there is no known orthologs for SMG1, SMG5 and SMG6. Thus, this may suggest that the regulation of Upf1 phosphorylation state is limited to metazoans.

Figure 5. Recruitment of SMG7 by phosphorylated Upf1.

Phosphorylated Upf1 recruits the SMG5-7 heterodimer and PP2A (step 1 and 2). Unknown events trigger dephosphorylation of Upf1 by PP2A and loss of mRNA 5' cap by decapping (step 3). Dephosphorylation of Upf1 leads to the dissociation of SMG5-7 and PP2A from the mRNP (step 4). SMG7 targets mRNA for decay via an unknown mechanism. After decapping, the mRNA is susceptible to rapid 5'-3' decay by the Xrn1 exonuclease.

1.1.4 Decay of mRNA after premature termination codon recognition

Finally, the decay of the aberrant mRNA is carried out by ribonucleases that are involved in general mRNA decay. In eukaryotic cells, the general mRNA decay is initiated through the shortening of the poly(A) tail of the mRNA transcript by deadenylases (Parker and Song, 2004). However, the NMD surveillance complex is able to bypass deadenylation and promotes the decay of PTC-containing mRNAs directly. Several reports have shown that mRNAs with PTCs undergo decapping-dependent 5' - 3' decay pathway and exosome mediated 3' - 5' decay pathway in yeast and mammals (**Figure 5 and 6**) (Cao and Parker, 2003; Chen and Shyu, 2003; Couttet and Grange, 2004; Lejeune et al., 2003; Mitchell and Tollervey, 2003; Muhlrads et al., 1994; Parker and Song, 2004). For decapping - dependent 5' - 3' decay pathway, decapping exposes the mRNA to digestion by the 5' - 3' exonuclease Xrn1 (Muhlrads et al., 1994). This process occurs in P bodies which are enriched in Xrn1, Dcp1, Dcp2 and decapping activators. Correspondingly, genome wide expression profiling of yeast showed that 70% of the transcripts upregulated in Upf1 deleted (Upf1 Δ), Upf2 Δ , Upf3 Δ strains are also upregulated in Dcp1 Δ and Xrn1 Δ strains. This suggests that the majority of the NMD targets are degraded via the decapping dependent 5'-3' pathway in yeast (He et al., 2003).

For the exosome 3' - 5' decay pathway, the yeast NMD surveillance complex is proposed to accelerate deadenylation, which is immediately followed by the 3' - 5' decay effected by the exosome and the Ski complex (Cao and Parker, 2003; Mitchell and Tollervey, 2003; Parker and Song, 2004). The exosome is a large protein complex consisting of multiple 3' - 5' nucleases. Nine core exosome subunits form ring-like structure and use phosphate as an attacking group in RNA metabolism, and produce nucleoside diphosphates (NDPs) (Symmons et al., 2002).

In *D. melanogaster*, mRNA decay is initiated by a distinct pathway that involves the endonucleolytic cleavage near the PTC (Gatfield and Izaurralde, 2004). This produces highly unstable decay intermediates that can only be detected after respective exonuclease activities are sufficiently inhibited.

Thus, bypassing the need for decapping or deadenylation, the resultant 5' decay intermediate is degraded by the exosome and the Ski complex, while the 3' decay intermediate is degraded by the Xrn1 protein (Gatfield and Izaurralde, 2004).

In mammals, it was found that RNA interference (RNAi)-mediated downregulation of Dcp2 (5' - 3' decay) or PM-Scl100 component of exosome (3' - 5' decay) lead to significant increases in the abundance of PTC - containing mRNAs (Lejeune et al., 2003). In support of the observations, it was also shown that the decay of the NMD substrates in mammals occurs via 3' – 5' exonucleolytic pathway (Chen and Shyu, 2003; Couttet and Grange, 2004; Lejeune et al., 2003). These results suggest that mammalian cells undergo similar exonucleolytic pathways.

Interestingly, reports have shown that the mammalian NMD also involves endonucleolytic cleavage near the PTC (Eberle et al., 2009; Huntzinger et al., 2008). It was observed that polyadenylated 3' fragments are produced in Xrn1 depleted human cells and its production is dependent on Upf1 and not the Dcp2 protein (Eberle et al., 2009). In addition, the size of the 3' fragments is proportional to the location of the PTC. These strongly suggest that the polyadenylated 3' fragments are generated by the endonuclease. Furthermore, the cloning of the 3' fragment revealed that cleavage occurs without any sequence preference and within 40 nucleotides upstream or downstream of the PTC.

The discovery of the Pilt N-terminus (PIN) domain of the NMD factor SMG6 degrading circular RNAs *in vitro* demonstrates that SMG6 is an endonuclease. This makes SMG6 a prime candidate responsible for the endonucleolytic decay of mammalian NMD substrates (Eberle et al., 2009). The knockdown and reconstitution experiments have provided additional evidence that SMG6 is indeed the factor that mediates the endonucleolytic cleavage of NMD substrates in mammals and *D. melanogaster* (Eberle et al., 2009; Huntzinger et al., 2008).

Given that mammalian NMD substrates also undergo exonucleolytic decay, it was proposed that after mammalian PTC recognition, specific

A. *S. cerevisiae*

B. *D. melanogaster*

Figure 6. NMD degradative activities.

(A) NMD in *S. cerevisiae* is effected mainly through Dcp1-Dcp2 mediated decapping and subsequent 5' - 3' decay by the Xrn1 exonuclease. A minority of the transcripts undergo Pan2 and Ccr4-mediated deadenylation and exosome mediated 3' - 5' decay. (B) NMD in *D. melanogaster* involves endonucleolytic cleavage in the vicinity of the PTC. Subsequent decay intermediates are subjected to 5' - 3' decay by Xrn1 and 3' - 5' decay by the exosome and the Ski7 protein.

Figure 7. Branched NMD pathway models.

Phosphorylation of Upf1 can lead to recruitment of either SMG6 (left panel) or SMG5-7 heterodimer (right panel). Binding of SMG6 endonuclease leads to cleavage in the vicinity of the PTC and decay intermediates are degraded from 5' - 3' and 3' - 5' directions by Xrn1 and Ski7-Exosome respectively. Alternatively, binding of SMG5-7 leads to either decapping mediated 5' - 3' decay, or deadenylase mediated 3' - 5' decay.

interaction with either SMG6 or SMG7 determines if these NMD substrates undergo exonucleolytic or endonucleolytic cleavage (**Figure 7**). Consistent with the model, tethering assays, where the SMG7 protein is tethered to a reporter mRNA, showed that SMG7 activates 5' - 3' and 3' - 5' decay of mRNA. Additionally, both decay pathways are independent from the SMG6 (Unterholzner and Izaurralde, 2004). Similarly, in *S. cerevisiae*, activation of exonucleolytic decay of NMD substrates is independent of SMG6 but dependent on Ebs1 which is an orthologue of SMG7. Conversely, in *D. melanogaster*, endonucleolytic cleavage of NMD substrates occurs in the presence of SMG6 and not SMG7. Taken together, these suggest that the recruitment of either SMG6 or SMG7 may lead to distinct decay pathways of NMD substrates in mammals.

1.1.5 Subcellular localisation of Nonsense mediated decay

As translation occurs in the cytoplasm, it is likely that mammalian NMD takes place in a cytoplasmic site since it is dependent on translation. An inhibition assay showed that upon the cytoplasmic expression of dominant negative polypeptides that disrupt interaction of NMD factors, NMD is inhibited. Conversely, NMD is unaffected when these polypeptides are confined in the nucleus (Singh et al., 2007).

In *S. cerevisiae*, there are evidences that NMD takes place in the P body (Bruno and Wilkinson, 2006; Sheth and Parker, 2006). First, the proteins Dcp2, Dcp1 and Xrn1, which catalyse the degradation of PTCs containing mRNAs, are localised in P bodies (Sheth and Parker, 2003). Second, Upf1 is shown to be sufficient in targeting mRNAs to P bodies (Sheth and Parker, 2006). Third, the Upf1, Upf2 and Upf3 proteins localised to P bodies when NMD is inhibited at the nucleolytic step (Sheth and Parker, 2006). In mammals, it is speculated that NMD also takes place in the P bodies. A study showed that overexpression of SMG7 recruits SMG5 and Upf1 into the cytoplasmic foci that has the characteristics of P bodies (Unterholzner and Izaurralde, 2004). Nevertheless, it is not established if endogenous SMG7 localised to P bodies because its low expression levels prevents the determination of its intracellular location (Ohnishi et al., 2003; Reichenbach et al., 2003). In contrast, it was observed that NMD was not affected in HeLa cells when P body formation was disrupted by

depletion of the Human enhancer of decapping large subunit (Hedls) protein (Stalder and Muhlemann, 2009). Thus, more evidences are required to determine the cellular location of NMD.

1.1.5.1 Formation of P bodies

P bodies are found in yeast, insect cells, nematode and mammalian cells and they are formed through aggregation of translationally repressed mRNPs that are associated with translation repression and mRNA decay machinery (Sheth and Parker, 2003). The rate of P body formation is affected by the concentration and rate of aggregation of translationally repressed mRNAs. In addition, deletion mutants of the individual components of P bodies demonstrated that the assembly of a P body is redundant and no single component of the P body is absolutely required (Teixeira and Parker, 2007). Nevertheless, the molecular basis for the aggregation of mRNPs awaits elucidation. While studies shown that P bodies are formed by mRNPs, the exact composition of P bodies remains to be solved. In general, there are three classes of P body components. The first class belongs to a set of proteins that are conserved from yeast to mammals. These proteins are part of the general repression/decay machinery and they include the decapping enzyme Dcp1-Dcp2, activators of decapping Dhh1, Pat1, Lsm1-7, exonuclease, Xrn1 and the deadenylase Ccr4-Pop2-Not complex (Behm-Ansmant and Izaurralde, 2006; Cougot et al., 2004; Sheth and Parker, 2003; van Dijk et al., 2002). It was shown that Dcp2 and Pat1 are required to recruit Dcp1 and the Lsm1 - 7 complex to the p bodies respectively (Teixeira and Parker, 2007).

The second class includes proteins whose functions are limited to specific organisms or subclasses of mRNAs. For metazoans, the P bodies contain proteins for mediating miRNA repression like Argonautes and TNRC6B (paralog of GW182) (Behm-Ansmant et al., 2006; Liu et al., 2005). Specifically in human P bodies, there are proteins that are involved in antagonising viral functions (Wichroski et al., 2006). The final component of P bodies is mRNAs, which are not actively in translation (Liu et al., 2005; Teixeira et al., 2005). It was observed that there was a direct correlation between the cellular concentration of translationally repressed mRNPs and the extent of P body formation in the cells (Bregues et al., 2005; Cougot et al., 2004; Eulalio et al., 2007). In support of the

proposed translation inactive status of mRNAs in P bodies, translation initiation factors and ribosomes have been observed to be absent or inactivated in P bodies (Andrei et al., 2005; Brengues et al., 2005; Ferraiuolo et al., 2005; Teixeira et al., 2005).

In *S. cerevisiae*, the formation of P bodies was coupled to Dhh1 and Pat1 as deletion of both proteins disrupt P body formation upon glucose starvation. (Coller and Parker, 2005). In addition, deletion of the C - terminal dimerisation domain, Yjef-N, of the enhancer of decapping protein (Edc3) and the C - terminal Glu/Asn rich domain of Lsm4 disrupt the formation of P bodies (Decker et al., 2007; Reijns et al., 2008). In support of the observation, Edc3 depletion also impairs P body formation in *D. melanogaster* showing that function of Yjef-N dimerisation domain of Edc3 in P body assembly is highly conserved across species (Tritschler et al., 2007). While the Glu/Asn rich domain of yeast Lsm4 was not conserved in metazoans, similar Glu/Asn rich domains can be found in the human Hedls and GW182 (Decker et al., 2007; Reijns et al., 2008). Accordingly, deletion of either Hedls or GW182 leads to loss of P bodies in humans (Eulalio et al., 2007; Jakymiw et al., 2005; Liu et al., 2005; Yu et al., 2005). Taken together, an evolutionary conserved model was proposed whereby the translation repressor proteins (Dhh1 and Pat1) catalyse the exit of mRNPs from their translationally active status and these mRNPs assemble together to form P bodies in the presence of Edc3 and Lsm4 (**Figure 8**) (Decker et al., 2007; Pilkington and Parker, 2008).

A

B

Figure 8. P body formation in yeast and human.

(A) Edc3 and Lsm4 contain YjeF-N and Q/N domains respectively. Assembly of associated mRNPs into P bodies monomers (PB monomers) requires the interaction of the Lsm-Pat1 complex and the decapping complex. Subsequently, P body is formed by the aggregation of the PB monomers through YjeF-N and Q/N homomeric interactions as indicated by the arrows. (B) Hedls and GW182 contains similar Q/N domains that may be involved in cross-linking of PB monomers. Therefore, PB monomers in humans may include interaction of Lsm-Pat1 complex, decapping complex and RNA induced silencing complex (RISC). Similarly, the P body is formed through the aggregation of PB monomers which involve YjeF-N, Q/N domain and RGG domain homomeric interactions. Figure is from Franks and Lykke-Andersen, 2008

1.1.6 Nonsense mediated decay factors

1.1.6.1 Up-frameshift factor 1 protein

The Upf1 protein is a nucleo-cytoplasmic shuttling protein and it is a key effector protein in NMD that links the translation machinery to the surveillance complex via its interactions with the translation release factors (eRF1 and eRF3 proteins) and the Upf2 protein. Upf1 is a 109 kDa protein and its nucleic acid dependent DEAD-box ATPase and RNA helicase have the highest sequence

conservation among a number of organisms (Bhattacharya et al., 2000; Culbertson and Leeds, 2003; Czapinski et al., 1995; Leeds et al., 1991). The Upf1 protein has seven conserved group I helicase motifs, which is a characteristic of members belonging to the RNA/DNA helicase superfamily 1 (Altamura et al., 1992; Koonin, 1992; Leeds et al., 1992). In addition, its ATPase activity resides in two of the helicase motifs, and it is linked to 5' - 3' helicase activity (Bhattacharya et al., 2000). Studies have shown that the ATPase activity is essential for NMD across species and the RNA binding ability of the Upf1 protein is dependent on the presence of ATP (Bhattacharya et al., 2000; Cheng et al., 2007). Moreover, the ATPase deficient Upf1 mutant in yeast showed the accumulation of nonsense mRNA and decay associated proteins (Dcp2, Dhh1, Pat1 and Lsm1) in P bodies (Sheth and Parker, 2006). More importantly, the Upf2 and Upf3 proteins were absent in P bodies. These suggest that the ATPase activity of *S. cerevisiae* Upf1 is important in recruiting the Upf2 and Upf3 proteins, which would trigger the actual decay of the aberrant transcript. Nevertheless, it is shown that mammalian Upf1 protein co-immunoprecipitates with the Upf2 and Upf3 proteins in an ATPase - independent manner (Kashima et al., 2006). Thus, the role of Upf1 ATPase in the recruitment of the Upf2 and Upf3 proteins may not be conserved in mammals. Similarly, it was suggested that the helicase activity of Upf1 is critical in dissociating the cap-binding complex from mRNA. This leads to translation inhibition of mRNA, and thus makes the mRNA accessible to mRNA decay machinery (Jankowsky et al., 2001; Lykke-Andersen, 2002; Sun et al., 1998; Weng et al., 1996).

1.1.6.2 Recruitment of decapping machinery by Up-frameshift factor 1 protein

The recruitment of the decapping enzyme and its subsequent activation upon PTC recognition remains poorly understood. It has been shown that both *S. cerevisiae* and *H. sapien* Upf1 proteins can co-immunoprecipitate and also display two-hybrid interaction with decapping enzymes. Nevertheless, no direct interaction between Upf1 and decapping enzymes has been demonstrated (Fenger-Gron et al., 2005; He and Jacobson, 1995, 2001; Isken et al., 2008; Lejeune et al., 2003; Lykke-Andersen, 2002). Accordingly, it was observed that the loss of either *S. cerevisiae* Upf1, Upf2 or Upf3 proteins inhibit decapping of nonsense containing mRNAs (He and Jacobson, 2001). Taken together, these

observations suggest that the Upf1 protein recruits the decapping enzyme to the PTC-containing mRNA through direct or indirect protein-protein interactions. Interestingly, inactivation of either *S. cerevisiae* Upf1 or Upf2 proteins in Xrn1 Δ cells, alters the levels of capped wild type Ura5, Tcm1 and Pgk1 transcripts, even though the effect is modest, as compared to nonsense containing mRNAs (He and Jacobson, 2001). This shows that the Upf1 and Upf2 proteins may also affect the decapping of wild type transcripts.

1.1.6.3 Up-frameshift factor 1 protein phosphorylation

In metazoans, the N and C termini of the Upf1 protein contain multiple serine residues, which are regulated by cycles of phosphorylation and dephosphorylation (Behm-Ansmant et al., 2007b; Chang et al., 2007). In the presence of Upf2 and Upf3 proteins, Upf1 phosphorylation at multiple serine residues (Ser1078 and Ser1096) in serine glutamine (SQ) motifs is catalysed by SMG1, which is a phosphoinositide-3-kinase related protein kinase (**Figure 9**) (Behm-Ansmant et al., 2007b). Consistent with these observations, the depletion of Upf2 or Upf3 proteins inhibits the accumulation of phosphorylated Upf1, which suggests that Upf2 or Upf3 stimulates the activity of SMG1 (Kashima et al., 2006; Wittmann et al., 2006). Although the phosphorylation of the Upf1 protein is a critical event in mammalian cells, this importance across species remains to be determined. In *C. elegans* and mammals, SMG1 is essential for NMD (Pulak and Anderson, 1993; Yamashita et al., 2001). However, knocking down of SMG1 does not inhibit the NMD of PTC-containing alcohol dehydrogenases in *D. melanogaster* (Chen et al., 2005). In addition, the ortholog of SMG1 is not encoded in the yeast genome (Yamashita et al., 2001).

After Upf1 is phosphorylated, it recruits SMG5, SMG6 and SMG7 (Chang et al., 2007; Denning et al., 2001; Page et al., 1999; Pal et al., 2001), which adopt a 14-3-3 like fold, comprising of several tetratricopeptide (TPR) repeats. The 14-3-3 protein's phosphoserine-binding pocket also shares similar residues with SMG5, SMG6 and SMG7 (Fukuhara et al., 2005). As 14-3-3 is a phosphoserine binding protein involved in signal transduction, this implies that SMG5, SMG6 and SMG7 may be a phosphoserine binding protein. In support of this model, SMG7 binds to phosphorylated SQ motifs of the Upf1 protein and

Figure 9. Phosphorylation of Upf1 is catalysed by SMG1

SMG5, SMG6 and SMG7 proteins bind to phosphorylated Upf1 protein via specific domains which adopt a 14-3-3-like fold. The SMG proteins also trigger Upf1 dephosphorylation via recruitment of PP2A.

mutation of the residues in the binding pocket of SMG7 impairs its binding ability to Upf1 *in vitro* (Fukuhara et al., 2005) which leads to the failure in recruitment of the Upf1 protein to the P bodies (Chang et al., 2007). Additionally, it was found that SMG6 forms a complex with the protein phosphatase 2A (PP2A) and phosphorylated Upf1 (Chiu et al., 2003; Ohnishi et

al., 2003). Subsequently, SMG5 and SMG7 also associate together and form a distinct complex with PP2A and phosphorylated Upf1 (Anders et al., 2003; Ohnishi et al., 2003). Thus, as SMG5, SMG6 and SMG7 form protein complexes containing the phosphorylated Upf1 and PP2A, it was proposed that these complexes bridge PP2A to phosphorylated Upf1 to trigger Upf1 dephosphorylation. (Anders et al., 2003; Behm-Ansmant et al., 2007b; Chiu et al., 2003; Ohnishi et al., 2003). In support of this model, loss of function mutations in any of the three genes in *C. elegans* lead to accumulation of phosphorylated Upf1 proteins (Page et al., 1999).

Interestingly, the phosphorylation of the Upf1 protein was discovered to trigger translational repression in mammalian cells. This was a previously unappreciated step in NMD (Isken et al., 2008). It was shown that the phosphorylated Upf1 has higher affinity for mRNA decay enzymes like the Dcp1a, Xrn1 and Rps4 proteins. More importantly, the phosphorylated Upf1 protein also has greater affinity for eIF3 and its association with eIF3 interferes with the recruitment of the 60S ribosomal subunit to the 40S pre-initiation complex (**Figure 10**). This leads to the abrogation of the eIF3-mediated conversion of 40S/Met-tRNA^{Met}/mRNA to translationally competent 80S/Met-tRNA^{Met}/mRNA (Isken et al., 2008). In addition, eIF3 mediated cap-dependent and Hepatitis C virus internal ribosome entry site (HCV-IRES) -dependent translation were also inhibited. The translation repression of the mRNA-protein (mRNP) in mammals was synonymous with the decapping activators, Dhh1 and Pat1, inhibiting the translation of the *S. cerevisiae* mRNPs (Coller and Parker, 2005). Accordingly, the authors suggest that the mRNA decay in eukaryotes require existing mRNAs to transit from their active translation status to a repressed translation status. This transition is necessary to sequester mRNAs from translational machinery and bring about the assembly of the decay machinery on the mRNAs (Isken et al., 2008).

1.1.6.4 Up-frameshift factor 2 and Up-frameshift factor 3 proteins

The generation of a conditional Upf2 knockout mouse showed that long-term hematopoietic stem cells in the bone marrow are rapidly removed upon Upf2 depletion. Nevertheless, differentiated cells remains relatively indifferent to

Figure 10. Phosphorylated Upf1 mediated repression of translation initiation.

Phosphorylated Upf1 binds to eIF3 and inhibit eIF3 mediated joining of 60S ribosomal subunits to 40S/Met-tRNA^{Met}/mRNA to form translational competent 80S/ Met-tRNA^{Met}/mRNA. Subsequently, mRNA decay factors are recruited.

the loss of Upf2. This suggests that Upf2 plays a critical role in the survival of proliferating cells (Weischenfeldt et al., 2008). However, it remains to be elucidated if the observation is due to the disruption of the NMD, or is evidence of a novel function of the Upf2 protein.

More importantly, the Upf2 protein is shown to interact with eIF4A and this association is proposed to facilitate the breakdown of the eIF4F complex, which results in the inhibition of the translation of mRNAs (Mendell et al., 2000). In addition, as observed in Upf2 Δ and Upf3 Δ *S. cerevisiae* mutant strains, P bodies increase in size and number (Sheth and Parker, 2006). Further analyses revealed that even though Upf1 and Dcp2 accumulate in these P bodies, the aberrant mRNAs do not undergo NMD. This suggests that after Upf1 dependent targeting of the NMD substrates to P bodies, Upf2 and Upf3 are required to facilitate decay of the aberrant mRNAs.

The Upf3 is the least conserved among the Upf proteins (Culbertson and Leeds, 2003). Furthermore, mammals possess two Upf3 genes while *S. cerevisiae*, *D. melanogaster* and *C. elegans* have only one Upf3 gene. Human Upf3a is located on chromosome 13, while Upf3b is located on the chromosome X. Upf3 is predominantly a nuclear protein which can shuttle to the cytoplasm, and in human cells, it associates with the spliced mRNA via interactions with the EJC components (Conti and Izaurralde, 2005). In addition, both Upf3a and Upf3b bind the Upf2 protein through their N termini (Chang et al., 2007). However, Upf3b strongly elicits NMD when it is tethered downstream of a stop codon, while Upf3a only exerts a moderate effect. Sequence comparison showed that Upf3b differs from Upf3a by an amino acid that facilitates its interaction with Y14, and hence Upf3b is more important in NMD, as compared to Upf3a (Gehring et al., 2003; Kunz et al., 2006). Loss of function assays also showed that NMD substrates are stabilised only when Upf3b is depleted (Chang et al., 2007).

The interaction of human Upf2 and Upf3 has been visualised with atomic details (Kadlec et al., 2004). Upf2 consists of three middle portions of eIF4G domains (MeIF4G) while Upf3b contains a canonical RNP type RNA binding domain (RBD). The interaction between Upf2 and Upf3 protein involves the binding of the third MeIF4G domain of Upf2 and the RBD domain of Upf3. In addition, the Upf2 protein also utilises its C-terminal region (1084-1272 aa) to bind to the N-terminal zinc finger domain of Upf1 (Mendell et al., 2000; Serin et al., 2001). The N terminal zinc finger of Upf1 resembles the ring-box and U-box of ubiquitin ligases (Kadlec et al., 2006), and it is formed by the cysteine-histidine rich domain (115-272 aa). Subsequently, it was found that N-terminal region of Upf2 (94-133 aa) also contributes to its binding to Upf1 (He et al., 1996;

Serin et al., 2001). Thus, in the Upf2-Upf3 complex, the N and C terminal regions of Upf2 should be accessible to UPF1 to form a trimeric complex.

1.1.6.5 Proline-rich nuclear receptor coregulatory protein 2

Cho et al., (2009) demonstrated that the human proline-rich nuclear receptor coregulatory protein 2 (PNRC2) is a Dcp1a- and Upf1-interacting protein, and it interacts preferentially with the hyperphosphorylated Upf1. Thus, they propose a model where the hyperphosphorylated Upf1 triggers dissociation of eRF1 and eRF3 proteins. Subsequently, the interaction between Upf1 and PNRC2 proteins targets the Upf1 bound transcripts to P bodies for NMD (**Figure 11**) (Cho et al., 2009). In support of the model, downregulation of the PNRC2 protein through small interfering RNA (siRNA) leads to redistribution of hyperphosphorylated Upf1 from P bodies to cytoplasm (Cho et al., 2009).

Subsequently, Dcp1a, which is in complex with the PNRC2 protein, may facilitate 5' decapping of the aberrant transcript through the formation of the Dcp2-Dcp1 complex in P bodies. After decapping, the mRNAs in P bodies undergo exonucleolytic or endonucleolytic decay. Accordingly, downregulation of the endogenous PNRC2 protein abrogates NMD and the artificial tethering of PNRC2 downstream of a normal stop codon promotes NMD (Cho et al., 2009). Additionally, microarray analyses showed that at least 30% of the endogenous NMD substrates require PNRC2 for decay (Cho et al., 2012). These results provide evidence that PNRC2 is essential for NMD and it mediates the cross talk between Upf1-containing NMD machinery and the decapping complex by serving as an adaptor protein to bridge the interaction between the Dcp1a and Upf1 protein. Till date, an atomistic detail of PNRC2 binding with Dcp1a remains to be elucidated. Given that Edc3 stimulate decapping activity upon binding to Dcp1, it is postulated that PNRC2 is able to modulate the decapping activity.

1.1.7 Evolution of Nonsense mediated decay pathway

Initially, it was suggested that the splicing dependent PTC recognition mechanism discovered in mammalian cells has evolved from the more ancestral PABPC dependent PTC recognition mechanism found in the invertebrates. Nevertheless, it was discovered that both extended 3' UTR (300, 500, 700 nt)

Figure 11. PNRC2's role in NMD.

Phosphorylated Upf1 and its associated transcript are targeted into the P body through Upf1 and PNRC2 interaction. Additionally, PNRC2 serves as a bridge between the decapping complex and NMD surveillance complex, which leads to decay of the aberrant transcript in the P body. Figure is from Cho et al, 2009.

and exon-exon boundary, placed 99 nt downstream of a stop codon, in *Nicotiana tabacum* PHA minigene elicit NMD in a Upf1-dependent manner. This shows that both mechanisms can co-exist (Kertesz et al., 2006).

1.1.8 Involvement of Nonsense mediated decay factors in additional cellular process

The splice-dependent PTC definition in mammalian cells has implications in the evolution of NMD. It has been suggested that the EJC protein components Y14, MAGOH, eIF4AII, Barentsz and RNPS1, which are essential for PTC recognition in mammalian cells, have functionally substituted Upf2 or Upf3 (Chan et al., 2007; Gehring et al., 2005). Hence, these functionally released NMD effectors may have further evolved to acquire novel functions. In support of the hypothesis, the mammalian Upf1, SMG1, SMG5, SMG6 and SMG7 have roles in other cellular processes besides NMD (Rehwinkel et al., 2006).

Interestingly, Upf1, Upf2, SMG1 and SMG6 are enriched at telomeres and they negatively regulate the association of the telomeric repeat-containing RNAs (TERRAs) with chromatin (Azzalin et al., 2007). More specifically, human SMG6 associates with telomerase activity in Hela cells and overexpression of SMG6 leads to accumulation of chromosome-end fusions (Reichenbach et al., 2003). Overall, the association of NMD factors with telomeres provide a link between TERRA regulation and telomere maintenance, which is distinct from NMD.

SMG1 is also found to phosphorylate tumour suppressor p53 protein and activate p53 protein upon genotoxic stress. Thus, depletion of SMG1 leads to spontaneous DNA damage and increased sensitivity to ionizing radiation (Brumbaugh et al., 2004). Moreover, it was discovered that genotoxic stress also leads to phosphorylation of Upf1 by Ataxia telangiectasia and Rad3 related kinase (ATR) and causes the association of phosphorylated Upf1 with chromatin (Azzalin and Lingner, 2006). Two observations suggest that Upf1 plays a role in DNA repair. First, Upf1 is identified as the DNA polymerase δ associated helicase (Carastro et al., 2002). Second, depletion of Upf1 causes cell cycle arrest in early S phase and ATR dependent DNA damage response (Azzalin and Lingner, 2006). Subsequently, attempts in generation of embryonic stem cell lines from Upf1-null mouse blastocyst were unsuccessful and Upf1 knock out mice suffer from early embryonic lethality (Medghalchi et al., 2001).

1.1.9 Alternative complexes that trigger Nonsense mediated decay in human

While genetic and structural studies have shown evidence that Upf1-Upf2-Upf3 protein complex is necessary to invoke NMD in *S. cerevisiae*, *D. melanogaster* and *C. elegans*, Gehring et al., (2005) showed that two alternative complexes can initiate NMD in human cells. Using tethering assays, a complex comprises of the Upf1 protein, RNA binding EJC protein RNPS1 and Upf2 triggers NMD. Furthermore, the complex activation of NMD is independent of Upf3 and EJC core proteins like Barentsz or eIF4AIII. The second complex found to activate NMD consists of Upf1, Upf3b, Y14, MAGOH, Barentsz and eIF4AIII (Gehring et al., 2005). Interestingly, this complex is independent of Upf2 and its activity is inhibited by the absence of Barentsz or eIF4AIII.

A third branch of NMD pathway was proposed when RNAi-mediated depletion of Upf3a, Upf3b has no effect on downregulation of TCR- β transcripts harbouring PTCs (Chan et al., 2007). As the Upf3 independent NMD pathway is dependent on both Upf2 (Mendell et al., 2002; Wang et al., 2002a) and the EJC core proteins (Palacios et al., 2004; Shibuya et al., 2004; Shibuya et al., 2006), it suggests that the transcripts are not degraded by the Upf3/EJC-independent pathway or Upf2-independent pathway. Furthermore, the proposed Upf3 independent pathway does not follow the 55 boundary rule as dictated by the classical EJC model. In addition, PTCs that reside closer to the 5' end of the penultimate exon elicit NMD more efficiently as compared to PTCs at the 3' ends (Wang et al., 2002a). This is the only known mammalian mRNA to exhibit polar NMD and hence strongly suggesting that the pathway is distinct.

The transcripts regulated by the three distinct NMD pathways in humans are downregulated by the Upf1 protein, which suggests that the different pathways converge at Upf1 to elicit mRNA decay (Chang et al., 2007). More importantly, there is evidence that NMD machinery has started incorporating the EJC components for its activity and some of the EJC components like Y14, MAGOH, Barentsz and eIF4AIII are able to substitute the NMD function of Upf2 (Gehring et al., 2005). Earlier studies by Gehring et al., (2003) have demonstrated that the Upf3b protein lacking its RBD, which is deficient in binding Upf2, still elicits NMD in a tethering assay. Furthermore, it has also been

reported that Upf1 can be recruited to the 3' UTR via interactions with Staufen-1 or histone-stem-loop-binding protein and trigger NMD using a distinct but related mechanism (Kaygun and Marzluff, 2005; Kim et al., 2005). Thus, these data lends support to the suggestion that the Upf2 independent NMD pathway exists in human cells.

1.1.10 Nonsense mediated decay and diseases

NMD plays a role in suppressing the clinical manifestation of specific human genetic disorders. In general, nonsense codons are generated by nonsense mutations or frameshifts, which account for 30% of inherited genetic disorders (Frischmeyer and Dietz, 1999; Holbrook et al., 2004). Thus, NMD serves as a protective surveillance mechanism to degrade mutant alleles, which lead to a recessive mode of inheritance. Alternatively, the failure to degrade aberrant nonsense transcripts will lead to the expression of truncated proteins which may have a dominant negative genetic effect. An example is the haemolytic anaemia β -thalassaemia, which arise from the introduction of PTCs into the β globin gene. This results in truncated β globin protein expression that exert a dominant negative effect affecting the integrity of haemoglobin (Hall and Thein, 1994). In most patients with β -thalassaemia, they display a recessive mode of inheritance. This is due to NMD degrading the aberrant transcript before the defective β globin can be expressed. However, if the PTC lies in the last exon of the β globin gene, it is able to bypass NMD and hence leads to a dominant mode of inheritance (Holbrook et al., 2004). Nevertheless, the existence of varying prognosis for the same mutation in different individuals suggests that the NMD efficiency varies among individuals.

Besides inherited genetic disorders, NMD is also shown to protect against acquired mutations. mRNAs of the mutant tumour suppressor proteins BRCA1 and p53 are eliminated by NMD. Nevertheless, intronless variants of these mutants, which are immune to NMD, express truncated proteins which lead to increased cell survival, tumorigenicity and resistance to chemotherapy (Holbrook et al., 2004).

Conversely, there are situations where the truncated proteins are partially functional. Thus, NMD aggravates the disease phenotype through the

destruction of mRNA that prevent sufficient accumulation of partially functional truncated protein (Kuzmiak and Maquat, 2006). An example is the Ulrich's disease (Usuki et al., 2004). In this disease, the exon 18 of the collagen VI α 2 mRNA undergoes a 26-nucleotides frameshift mutation, which results in a PTC in exon 22. The subsequent NMD of the mRNA leads to insufficient collagen VI α 2 protein production. Nevertheless, wortmannin and caffeine have been employed to suppress the phosphorylation of Upf1, which reduces NMD. This restores the expression of the collagen VI α 2 protein, which partially rescues the extracellular matrix function. Therefore, it is suggested that a higher order RNA structure or stably associated protein complex or small molecule drug situated between the PTC and the downstream cis signal EJC, may sever the functional PTC-EJC association and suppress NMD (Isken and Maquat, 2007). This strategy could be useful for therapy of NMD related diseases.

Additionally, treatment of diseases aggravated by NMD may be achieved by promoting readthrough of the PTC instead of inhibiting NMD. The aminoglycoside antibiotics are known to promote readthroughs of stop codons at high concentration and beneficial effects have been reported from clinical trials with cystic fibrosis patients (Wilschanski et al., 2003). Nevertheless, high concentration of aminoglycoside leads to undesirable toxic side effects that affect the kidneys and ears. Currently, a small organic molecule, PTC124, is able to discriminate between normal and premature termination codons, and it selectively promotes readthrough of PTCs (Welch et al., 2007). Nevertheless, the drug is still in phase 3 clinical trials for patients with cystic fibrosis.

1.2 Decapping

mRNA decay is an important step in the control of gene expression in response to various environmental and developmental changes (Mitchell and Tollervey, 2003; Parker and Song, 2004). Before mRNA decay is elicited, either the 5' cap or 3' poly(A) tail must be removed because both features stabilises mRNA. The decapping process is the removal of the 5' cap from mRNA and it is a key and irreversible step in several mRNA decay pathways such as the general 5' - 3' mRNA decay pathway, microRNA (miRNA)-mediated mRNA decay (Behm-Ansmant and Izaurralde, 2006; Behm-Ansmant et al., 2006; Eulalio et al., 2007), AU-rich elements (ARE) stimulated mRNA decay (Fenger-Gron et al., 2005; Gao et al., 2001; Lykke-Andersen and Wagner, 2005; Stoecklin et al., 2006; Wilusz and Wilusz, 2004), 3' uridylation mediated mRNA decay (Rissland and Norbury, 2009; Song and Kiledjian, 2007), non-polyadenylated histone mRNA decay (Mullen and Marzluff, 2008), initiation mediated mRNA decay (Heikkinen et al., 2003), nuclear decay of pre-mRNAs (Kufel et al., 2004) and nonsense mediated decay (Cho et al., 2009; He and Jacobson, 1995, 2001; Lykke-Andersen, 2002). As such, decapping is a site of numerous control inputs. The mRNA decapping enzyme is evolutionarily conserved among all eukaryotes and it consists of a catalytic core, the Dcp2 protein, and a coactivator, the Dcp1 protein (Chowdhury and Tharun, 2009; Deshmukh et al., 2008; Liu and Kiledjian, 2006; Parker and Song, 2004; Simon et al., 2006).

Biochemical analyses of decapping showed that the Dcp1-Dcp2 complex cleaves the 5'^{N7}-methylguanosine (m⁷G) cap at the 5' end of mRNA to form products m⁷GDP and a 5' monophosphate mRNA (LaGrandeur and Parker, 1998; Stevens, 1980). The mRNA m⁷G cap is critical for translation and it also protects the mRNA from 5' - 3' degradation (Hsu and Stevens, 1993). Therefore, removal of the cap simultaneously shuts down translation and allows rapid degradation of 5' monophosphate mRNA by 5' - 3' exonuclease, Xrn1 (Cao and Parker, 2003; Chen and Shyu, 2003; Couttet and Grange, 2004; Lejeune et al., 2003; Mitchell and Tollervey, 2003). In general, decapping consists of deadenylation, transition to translationally repressed mRNP complex, sequestration of mRNP in P bodies and assembly of decapping complexes on mRNP (**Figure 12**) (Coller and Parker, 2004).

-

Figure 12. Overall scheme of decapping.

The transcript slated for 5'-3' decay has to undergo deadenylation. Removal of poly(A) tail leads to the transition of translation active state to translation repressed state of the transcript which causes the disassembly of translation machinery. Translation repressed transcript is targeted to the P bodies where decapping complex is assembled on the transcript to carry out decapping. Figure is from Collier and Parker, 2004.

1.2.1 Deadenylation

The majority of mRNA degradation is initiated by the shortening of the poly(A) tail (Beelman et al., 1996) known as deadenylation, and it is often the rate limiting step in mRNA turnover (Muhlrad and Parker, 1994). There are different deadenylases such as the Ccr4, Pop2, Pan2-Pan3 complex and Poly (A) ribonuclease, which are proposed to target specific subset of mRNP substrates. This allows better control of deadenylation since altering the activity of one deadenylase will only affect a specific class of mRNPs. Furthermore, adaptor proteins that bind mRNA to control their rate of deadenylation can recruit specific deadenylase and activate the enzyme.

1.2.2 Transition to translationally repressed messenger ribonucleoprotein

It has been observed that loss of function mutations of translation initiation factors promote decapping (Schwartz and Parker, 1999). Additionally, inhibition of translation elongation by cyclohexamide leads to a decrease in decapping (Beelman and Parker, 1994; Peltz et al., 1992). These observations demonstrate that factors, which increase the association of mRNA with ribosomes, inhibit decapping while factors that repress translation may activate decapping. Thus, these suggest that translation and decapping are fundamentally coupled and they are in competition. The source of competition between translation and decapping remains to be elucidated. Nevertheless, it was observed that eIF4E, a component of translation initiation complex eIF4F, binds to the 5' cap of the mRNA (Cougot et al., 2004) and addition of purified eIF4E inhibits decapping. Conversely, *in vivo* loss of eIF4E function stimulates decapping (Schwartz and Parker, 2000). Similarly, the structure of the Dcp2 protein obtained by crystallographic studies showed that Dcp2 Nudix domain interacts with both the 5' cap and mRNA body (She et al., 2008). These data demonstrate that the translation and decapping machinery may compete for association with the 5' cap (Schwartz and Parker, 2000; Wilusz et al., 2001).

The proposed model has a clear implication that the translating mRNP must transit to a translationally inactive mRNP before decapping can commence. This suggests that cap-binding translation factors must be displaced before the

Dcp2 protein can bind the 5' mRNA cap and catalyse decapping. While the mechanism of the transition is still unsolved, co-immunoprecipitation studies have shown that the components of decapping machinery like Pat1, Lsm complex and the decapping complex (Dcp1-Dcp2) associate with the mRNP after deadenylation (Tharun and Parker, 2001). Thus, this suggests that the critical transition in mRNP organisation occurs after deadenylation.

An alternative model to explain the requirement of mRNPs to exit their translationally active state for decapping was also suggested when it was observed that decapping of transcript occurs in P bodies, whereas translating polysome-bound mRNAs are excluded from it (Beelman and Parker, 1994; Sheth and Parker, 2003). The model proposed that elongating ribosomes may be responsible for the inhibition of decapping by preventing the entry of translational competent mRNPs into P bodies. In support of the theory, increase ribosome loading was shown to decrease the decapping rate (Beelman and Parker, 1994; Cao and Parker, 2003; LaGrande and Parker, 1998; Peltz et al., 1992).

Subsequently, it was observed that not all mRNAs which are translationally repressed are targets of decapping (Brenques et al., 2005; Teixeira et al., 2005). Hence, it was proposed that in addition to translation repressed status of mRNAs, the eIF4F cap complex at 5'mRNA has to be destabilised (Franks and Lykke-Andersen, 2008). This can be achieved through the binding of pro-decapping factors to the mRNA 5' cap.

1.2.3 Sequestration of mRNA into P bodies

Several assays provide evidence that P bodies are highly dynamic sites where mRNAs are decapped and degraded in the 5' - 3' direction (Sheth and Parker, 2003). In *S. cerevisiae*, proteins involved in decapping like Dcp1, Dcp2, Lsm1, Pat1, Dhh1, Xrn1 are localised in P bodies in cytoplasm (Sheth and Parker, 2003). In addition, it was observed that the formation and size of P bodies correlate with the flux of mRNA molecules undergoing decapping. For example, addition of the translation elongation inhibitor, cyclohexamide, inhibits decapping and leads to reduction or loss of P bodies (Beelman and Parker, 1994). In contrast, inhibiting 5' - 3' exonuclease digestion leads to an increase in size and number of P bodies (Sheth and Parker, 2003). Finally, mRNA decay

intermediates are found to specifically localised in P bodies upon inactivation of Xrn1 (Sheth and Parker, 2003).

There are three advantages to the compartmentalisation of mRNA decapping in P bodies. First, the sequestering of decapping factors in P bodies separate the degradation machinery from actively translating mRNPs. This can prevent the premature decapping of mRNAs. Second, the partitioning of mRNA decay added an extra layer of control in the process as the delivery of mRNAs slated for decay to the P bodies can be regulated. Finally, sequestering of non-translational mRNAs prevent its competition for limited translation factors and ensure that the ratio of translation capacity to available pools of translating mRNAs are at optimum level in the cytoplasm.

1.2.3.1 Re-entry of mRNAs into translational pool

Translationally repressed mRNPs in P bodies can either be degraded or stored (Franks and Lykke-Andersen, 2008). Accordingly, *S. cerevisiae* mRNAs that assembled into P bodies upon glucose starvation are released back into the translational pool of mRNAs in cytoplasm after addition of glucose (Brenques et al., 2005). Hence, it was proposed that the protein complexes they associate within the P bodies decide the fate of translationally repressed mRNAs. Thus, if the mRNAs associate with protein complex that does not have mRNA decay factors, the resulting mRNA will be stored and subsequently released into the cytoplasm for translation (Franks and Lykke-Andersen, 2008). Consistent with the model, in mammalian cells or *D. melanogaster* neuronal cells, translation initiator eIF4E is bound to its suppressor protein CUP and 4E-T respectively. Together, the protein complex associate with the 5' cap of mRNA and translationally repress the resultant mRNA (Richter and Sonenberg, 2005). It has been observed that such mRNPs accumulate in P bodies and it is speculated that due to an absence of NMD factors in the protein complex, the mRNP is preferentially stored than degraded in the P bodies (Andrei et al., 2005; Barbee et al., 2006; Ferraiuolo et al., 2005).

1.2.4 Assembly of decapping complex

Distinct decapping complexes are found to be assembled on mRNA, which enhance the interaction of the decapping enzyme with mRNA (**Figure 13**). This leads to increased rate of decapping, and the type of decapping complexes assembled is dependent on the context of the mRNA. In *S. cerevisiae*, the majority of mRNAs assemble a decapping complex that consists of the Lsm1-Lsm7 complex, Pat1, Dhh1 and Dcp1-Dcp2 proteins (Beelman et al., 1996; Boeck et al., 1998; Bouveret et al., 2000; Coller et al., 2001; Fischer and Weis, 2002; Hatfield et al., 1996; Tharun et al., 2000). Consistent with the observations, Cho et al., (2009) demonstrated that the human Upf1 associates with the Dcp1a protein via PNRC2 interaction to form a distinct decapping complex specific to the PTC containing transcript. Furthermore, it was shown that Upf1 and SMG7 may recruit the Dcp1 and Dcp2 proteins to form the decapping complex for aberrant mRNA with PTCs (Unterholzner and Izaurralde, 2004). For mRNAs with AREs in its 3' UTR, TTP associate with the Dcp2 and Dcp1 proteins to form a distinct decapping complex specific for ARE containing mRNA (Fenger-Gron et al., 2005). Additionally, human 3' uridylated mRNA binds specifically to a decapping complex that consists of the Dcp2, Dcp1a, Hedls, Lsm1 and Lsm4. The presence of Lsm1 and Lsm 4 suggests that Lsm1-7 complex may also be involved in stimulation of decapping (Song and Kiledjian, 2007). Finally, Stau1 is also involved in stimulating the decapping complex of Stau1 mediated decay (SMD) (Cho et al., 2012). Taken together, these show that a common pool of decay components binds to specialised adaptor proteins to form unique decapping complexes targeted for specific subset of mRNAs.

Yeast decapping complex

ARE mediated decay

3' Uridylation mediated decay

Nonsense mediated decay

Stau1 mediated decay

Figure 13. The assembly of distinct decapping complexes

Different decaying complexes are assembled which are dependent on the context of mRNA. Common decay components like Dcp2 and Dcp1 bind to a range of adaptor proteins to target distinct subset of mRNAs.

1.2.5 Decapping enzyme

1.2.5.1 Dcp2

The Dcp2 protein is found to be a high copy suppressor of temperature sensitive *S. cerevisiae* strain with dysfunctional Dcp1 protein (Dunckley and Parker, 1999). Subsequently, orthologs of the Dcp2 protein in yeast, humans, *D. melanogaster*, *C. elegans* and *A. thaliana*, were found to have decapping activity which is independent of Dcp1 (Cohen et al., 2005; Iwasaki et al., 2007; Lin et al., 2008; Lykke-Andersen, 2002; Steiger et al., 2003; van Dijk et al., 2002; Wang et al., 2002a). Thus, this establishes Dcp2 protein as the main catalytic subunit in a decapping holoenzyme with Dcp1. The decapping activity of this holoenzyme is greatly stimulated as compared to Dcp2 alone (Beelman et al., 1996; Dunckley and Parker, 1999; Steiger et al., 2003), and it was observed that the Dcp2-Dcp1 decapping complex has a binding preference for longer mRNA substrate (LaGrande and Parker, 1998; Piccirillo et al., 2003; Steiger et al., 2003; Wang et al., 2002a). The loss of either the Dcp1 or Dcp2 protein in *S. cerevisiae* leads to a complete disruption of decapping in vivo (Beelman et al., 1996; Dunckley and Parker, 1999). This suggests that Dcp1-Dcp2 is the only decapping enzyme in *S. cerevisiae*.

The crystal structure of *S. pombe* Dcp2 protein revealed that its N terminal consists of an all-helical domain while the C terminal is made up of a classic Nudix fold (**Figure 14**) (She et al., 2006). The Nudix fold is common to the Nudix hydrolase family, which consists of enzyme that catalyse the hydrolysis of nucleoside triphosphates, nucleotide sugars, dinucleoside polyphosphates and capped mRNA. More importantly, the Nudix family of proteins are evolutionary conserved in eukaryotes, archaea, bacteria and viruses (McLennan, 2006). Structure analysis showed that the Nudix fold composed of an $\alpha/\beta/\alpha$ sandwich which contains a loop-helix-loop structure, that is formed by a signature Nudix motif (**Figure 14**) (Bessman et al., 1996; Mildvan et al., 2005; She et al., 2006). The consensus sequence of the Nudix motif is $GX_5EX_7REUXEEXGU$ whereby X is any residue and U is Ile, Leu or Val (Bessman et al., 1996; Mildvan et al., 2005). Accordingly, the active site for the decapping activity of Dcp2 has been mapped to the Nudix / MutT domain as this domain alone is sufficient for decapping RNAs *in vitro* (Dunckley and Parker, 1999; Lykke-Andersen, 2002;

Steiger et al., 2003; Wang et al., 2002b). Furthermore, mutations of glutamate residues found in close proximity or part of the Nudix motif (E143A, E147Q and E192A) are defective in an *in vitro* decapping assay (she et al, 2006). The results suggest that the three glutamate residues may act as a base to coordinate a divalent cation, Mg^{2+} or Mn^{2+} , which is responsible for the activation of a water molecule for nucleophilic attack of the pyrophosphate bond present in the mRNA cap (Mildvan et al., 2005; She et al., 2006).

In addition, Dcp2 has two additional conserved regions, Box A and Box B. While the function of Box A remains to be elucidated, it has been demonstrated that Box B is required for RNA binding and efficient decapping *in vitro* (Piccirillo et al., 2003). Furthermore, electrostatic potential mapping on the molecular surface of Dcp2 showed that this region of Dcp2 consists of many basic residues which makes it a potential RNA binding site (She et al., 2008). Direct evidence for RNA binding to Box B was also shown in NMR chemical shift perturbation experiments whereby the chemical shift observed are mapped to residues on Box B (Deshmukh et al., 2008).

Figure 14. Ribbon diagram of spDcp2n.

Salmon, the N terminal domain; green, The C terminal Nudix domain; red, Nudix motif. The Nudix motif is the catalytic centre of Dcp2p. Glutamate residues on the Nudix motif are responsible for catalysing the decapping reaction.

Besides decapping, *S. cerevisiae* Dcp2 is found to have an additional role of promoting P body assembly (Teixeira and Parker, 2007). It was observed that Dcp2 Δ yeast strain has consistently smaller P bodies as compared to the Dcp1 Δ or Xrn1 Δ strains. Thus, it was proposed that the interactions between Dcp2 and its partners (Dcp1, Dhh1, Edc3, Pat1) may stabilise P bodies mRNP or provide cross linking interactions between individual mRNP to stimulate the aggregation of mRNPs to form bigger P bodies.

1.2.5.2 Dcp1 protein

An additional key activator protein involved in decapping is identified as the Dcp1 protein (Beelman et al., 1996; Sakuno et al., 2004; Steiger et al., 2003) and its interaction with individual mRNA influences the rate of decay of that transcript. Beelman et al., (1996) showed that Dcp1 is essential for decapping in *S. cerevisiae*. Dcp1 mutant maintain normal rate of deadenylation but has impaired rate of decapping which results in the accumulation of full length capped mRNA transcripts (Beelman et al., 1996).

The comparison of the proline rich sequence (PRS) binding site sequence and subsequent crystal structure of *S. cerevisiae* Dcp1 showed that it belongs to the Drosophila enabled (Ena)/ vasodilator-stimulated phosphoprotein (VASP) homology 1 (EVH1) family of protein domains (She et al., 2004). Subsequent analysis of the Dcp1 structure revealed two structurally conserved patches that are critical for its function in decapping (**Figure 15**).

The Patch 1 of the Dcp1 protein is located in a concave V-shaped groove formed by the beta strands (β -strands) 1, 2, 5, 6 and 7. Patch 1 corresponds to the PRS binding site of other classes of EVH1 domain and it is suggested that the Dcp1 protein may bind to additional proline rich proteins that may possibly stimulate decapping. In support of its role in decapping, mutation of three residues (Trp56, Tyr47, Trp204) residing in Patch 1 cause a partial loss of decapping activity *in vivo* (Tharun and Parker, 1999). Nevertheless, it was observed that the mutation of residues forming Patch 1 does not affect the binding of Dcp1 to Dcp2. Thus, the loss in decapping activity is not due to reduced interaction between Dcp1 and Dcp2.

Figure 15. Surface representation of *S. cerevisiae* Dcp1.

Invariant residues in Patch 1 and Patch 2 are labelled. Patch 1 of *S. cerevisiae* Dcp1 is proposed to be the PRS-binding site while Patch 2 is analogous to the hotspot of disease mutations in N-Wasp. Additionally, Patch 2 is also suggested to be critical for Dcp1 function.

Patch 2 is located in a concave surface opposite of Patch 1 and it is formed by the N-terminal 3_{10} -helix and β -strands 1 - 4. This region corresponds to the hotspot of disease mutations in N-Wasp EVH1 domain (Volkman et al., 2002). A highly conserved cluster of charged residues, Arg70 and Asp31, reside in Patch 2 and subsequent mutation of either residues to alanine inhibit decapping *in vivo* (Tharun and Parker, 1999). Similarly, the mutation of conserved residues in Patch 2 does not affect Dcp1 binding with Dcp2. Thus, one aspect of Dcp1 enhancement of decapping activity is independent of Dcp2p binding. Furthermore, a large hydrophobic patch resides between Patch 1 and Patch 2. Their contribution to decapping activity is demonstrated when a quadruple mutant (L37A, L38A, L217A, L221A) targeting this hydrophobic patch has severe defect in decapping. Nevertheless, the hydrophobic patch did not affect Dcp1 association with Dcp2.

1.2.5.2.1 Transforming growth factor β /bone morphogenetic protein 4 (BMP4) signaling pathway

While the human Dcp1a protein is commonly known to enhance the decay of mRNA, Bai et al., (2002) observed that it also serves as a transcription co-activator role in the TGF β /bone morphogenetic protein 4 (BMP4) signaling pathway. Upon TGF β /BMP4 stimulation, the Dcp1a protein interacts with an

important mediator of the pathway, Smad4, to form a protein complex, which can translocate to the nucleus to activate transcription (Bai et al., 2002). It was revealed that the EVH1 domain of Dcp1a protein is responsible for binding to a well conserved proline rich sequence at the linker region at the C terminus of the Smad4 protein (275–308 aa). Coincidentally, this region is part of the Smad4 activation domain (SAD), which conveys the entire intrinsic transcriptional activity of the Smad4 protein. Detailed mutational analysis showed that Tyr301 and Trp302 of Smad4 are crucial for its interaction with EVH1 domain of hDcp1a.

1.2.5.3 Dcp1-Dcp2 decapping complex

Mutations in *S. cerevisiae* Dcp2 N-terminal region that disrupt interaction with Dcp1 led to impaired decapping activity *in vitro* and *in vivo* (Sakuno et al., 2004; She et al., 2006; She et al., 2004; She et al., 2008; Tharun and Parker, 1999). Indeed, the crystal structure of *S. pombe* Dcp2-Dcp1 decapping complex shows that N - terminal of Dcp1 associate with N - terminal α helical domain of the Dcp2 protein through hydrogen bonds and hydrophobic interactions with a buried solvent-accessible surface of 1016Å².

While *S. pombe* Dcp1 (SpDcp1) protein is shown to interact with the SpDcp2 protein, the residues of Dcp1 and Dcp2 involved in the interaction are not conserved in metazoans (Deshmukh et al., 2008; She et al., 2006; She et al., 2008). Consistent with the observations, when human Dcp1 protein is overexpressed in human embryonic kidney (HEK) 293 cells, it does not stably associates with Dcp2. Furthermore, it was shown that in humans, *C. elegans* and *D. melanogaster* Dcp1 does not stimulate Dcp2 activity *in vitro* (Cohen et al., 2005; Fenger-Gron et al., 2005; Iwasaki et al., 2007; Lin et al., 2008; Lykke-Andersen, 2002; van Dijk et al., 2002). These seemingly conflicted results across species suggest that an additional protein may be required for the association of the metazoan Dcp1 and Dcp2. In support of the proposed model, the Hedls protein is required to bridge the interactions between Dcp1 and Dcp2 in humans (Fenger-Gron et al., 2005).

Figure 16. Open and closed forms of Dcp1-Dcp2 complex.

(A) Ribbon diagram of Dcp1-Dcp2 complex in the open and inactive form. Dcp1 is shown in green, the N-terminal domain (NTD) of Dcp2 is shown in pink and C-terminal domain (CTD) of Dcp2 is shown in blue with the Nudix motif shown in red. (B) Ribbon diagram of Dcp1-Dcp2 complex in the closed and active form with ATP as an RNA mimic. NTD of Dcp2 is brought into close proximity to the CTD of Dcp2 and it was proposed that residues from both domains form a binding pocket to bind mRNA substrate. Coloring scheme as in (A).

Structure analysis of *S. pombe* Dcp1-Dcp2 holoenzyme showed that Dcp2 exists in open and closed form (**Figure 16**) (She et al., 2008). Consistent with the structural data, studies using small angle X-ray scattering (SAXS) demonstrated that upon the addition of mimics, ATP, m⁷GpppA or pAA, the Dcp2-Dcp1 complex transits from an open conformation to a closed form in solution. Moreover, it was observed that addition of mRNA mimics to the Dcp2 protein alone did not produce any measurable changes in the scattering pattern in the SAXS experiment. Taken together, the results suggest that binding of substrates in the active site promote the closed conformation of the Dcp2 protein in a Dcp1 dependent manner (She et al., 2008).

Several evidences showed that the closed conformation of the Dcp2 protein is the catalytically active form. First, upon forming the closed conformation, the N-terminal domain of Dcp2 protein is brought into close proximity to the Nudix domain and residues from both domains form a binding

pocket for the mRNA substrate. Consistent with the results, Glu39, Trp43, Asp47 and Phe48 residing in N terminal of Dcp2 are shown to affect the decapping activity of Dcp2 upon mutation (She et al., 2006). Second, ATP promotes the formation of the closed form of Dcp2 protein and it is bound to the active site in this closed form. Third, the closed form of the Dcp2 protein leads to the ordering of loop 185-193 which brings the catalytic residue, Glu192, to an appropriate position and orientation towards the active site to catalyse decapping. Fourth, mutation of hinge region residue, Arg95, to proline (R95P) mutant increases rigidity of the hinge and prevents the formation of the closed complex which strongly reduced decapping activity to a level that is comparable to a Dcp2 mutant with deletion of the N terminal domain. In addition, the R95P mutant decapping activity is not stimulated by addition of the Dcp1 protein, even though it is able to bind to Dcp1 in 1:1 ratio (She et al., 2008). The results suggest that the Dcp1 protein is able to bind to open form of the Dcp2 protein and enhance the decapping activity of Dcp2 by promoting or stabilising the closed form of Dcp2.

1.2.5.4 The scavenger decapping enzyme (DcpS)

The DcpS protein is a second type of decapping enzymes found in eukaryotic cells and it is part of the HIT family of pyrophosphatases which uses the histidine triad for decapping. It is shown that the DcpS protein decaps oligonucleotides produced by the exosome mediated 3'-5' mRNA decay (Wang and Kiledjian, 2001). In addition, the DcpS protein is required to hydrolyse m^7GDP , which is the product from mRNA decapping, to form m^7GMP and phosphate (van Dijk et al., 2003). Nevertheless, the DcpS protein is unable to decap long mRNA substrate and thus this may prevent premature decapping of mRNAs by DcpS (Liu et al., 2002).

1.2.6 Regulators of decapping

1.2.6.1 Decapping activators

Decapping is a highly regulated process that involves many protein-protein interactions acting upon the Dcp2 protein. The proteins that facilitate decapping are classified into general and pathway specific activators (Coller and

Parker, 2004; Eulalio et al., 2007; Garneau et al., 2007; Isken and Maquat, 2007; Krogan et al., 2006; Parker and Song, 2004). Their proposed mechanism is to provide a wider interface to bind RNA and also promote a tighter enzyme-substrate decapping complex for increased efficiency. Besides affecting the decapping complex, some activators of mRNA decapping like the Pat1 and Dhh1 proteins may indirectly promote decapping by repressing translation, as decapping requires mRNPs to exit from their translation competent state (Coller and Parker, 2005).

1.2.6.1.1 *Enhancer of decapping 1 and Enhancer of decapping 2*

In *S. cerevisiae*, the Edc1 and Edc2 proteins are shown to be high copy suppressor of decapping defects present in strains harbouring temperature sensitive alleles of the Dcp1 and Dcp2 proteins (Dunckley et al., 2001). In addition, Edc1 and Edc2 proteins stimulate decapping activity of Dcp2 in cell free extracts or with purified decapping enzyme. The results suggest that the Edc1 and Edc2 proteins may bind directly to the decapping holoenzyme Dcp1-Dcp2 (Schwartz et al., 2003; Steiger et al., 2003). Through sequence alignment analysis, it was observed that the C-terminal 30 residues of the Edc1 protein share high sequence identity with that of the Edc2 protein. The deletion of this proline rich region abolished the ability of the Edc1 protein to enhance decapping both *in vitro* and *in vivo* (Schwartz et al., 2003). Taken together, it was proposed that the Edc1 and Edc2 protein may use their proline rich sequence to bind to decapping holoenzyme and stimulate decapping activity. Besides direct binding of decapping holoenzyme, northwestern analysis showed that the Edc1 and Edc2 proteins are also RNA binding proteins (Schwartz et al., 2003). Thus, it was proposed that the RNA binding ability of the Edc1 and Edc2 proteins may also contribute to the stimulation of decapping activity of the Dcp1-Dcp2 complex (Schwartz et al., 2003) by increasing the interface for mRNA binding. Additionally, it was observed that Edc1 Δ and Dcp1 Δ yeast strain are unable to grow on glycerol at 37°C. Nevertheless, both strains are able to survive at 37°C on dextrose (Schwartz et al., 2003). The results suggest that Edc1 mediated decapping is necessary following a carbon source shift. More significantly, it suggests that specific Edc1 bound decapping complex only decaps subset of mRNAs that regulate carbon source shift.

1.2.6.1.1.1 Binding of Dcp1 with Enhancer of decapping 1 and Enhancer of decapping 2

The observations that the purified Edc1 and Edc2 proteins can stimulate decapping suggest that the Edc1 and Edc2 proteins are able to bind directly to the Dcp1-Dcp2 holoenzyme. Recently, Borja et al., (2011) showed that the Edc1 and Edc2 proteins can bind separately to the EVH1 domain of Dcp1 protein using proline rich consensus sequence of [E/D] XXX[L/F/W]PXP[S/T][F/W]. In agreement with previous results, independent phage display assays to identify peptides that interact with the Dcp1 protein also demonstrated that the Dcp1 protein has a preference for consensus sequence FPRP[S/T][F/W]. This gives evidence that Pro170 and Pro172 of Edc1 are essential for binding and a basic residue between the prolines might be a common motif for coactivators of the Dcp1 protein (Borja et al., 2011).

To quantify the binding between the Dcp1 and Edc1 proteins, fluorescence anisotropy assay was used. The PRS peptide of Edc1 was found to bind wild type Dcp1-Dcp2 with a K_d of 12 μ m (Borja et al., 2011). However, upon double mutation of Pro170 and Pro172 of the Edc1 protein, binding between the Edc1 protein and the Dcp1-Dcp2 protein complex was abolished (Borja et al., 2011).

1.2.6.1.1.2 Correlation between binding and stimulation of decapping activity

Single-turnover kinetic assay to study the RNA decapping by the Dcp1-Dcp2 protein complex (Jones et al., 2008) was used to determine the existence of a correlation between binding of either the Edc1 or Edc2 proteins to the Dcp1 - Dcp2 protein complex and subsequent stimulation of decapping activity. Subsequently, the wild type Edc1 and Edc2 proteins were found to enhance the catalytic efficiency (K_{max}/K_m) of the Dcp1-Dcp2 protein complex by 3000-fold and 1000-fold respectively (Borja et al., 2011). Additionally, P170A and P172A double proline mutants of Edc1 and P142A of Edc2 displayed a 35-fold and 33-fold reduction in the ability to stimulate decapping respectively (Borja et al., 2011). Moreover, the respective improvement of substrate binding by 140-fold and 40-fold upon addition of the Edc1 and Edc2 proteins to the Dcp1-Dcp2 protein complex agrees with previous data that both proteins bind RNA

(Schwartz et al., 2003). In addition to reducing K_m , the Edc1 and Edc2 proteins were found to enhance the rate limiting catalytic step of decapping which increases the K_{max} . Taken together, these results demonstrate that the Edc1 and Edc2 proteins may provide additional interface for the Dcp1-Dcp2 complex to bind RNA (reduction in K_m) and also promote a tighter enzyme-substrate decapping complex for increased efficiency (enhancement of K_{max}).

1.2.6.1.1.3 Enhancer of decapping 1 and Enhancer of decapping 2 binding residues

Despite direct biochemical evidence showing that both the Edc1 and Edc2 proteins regulate the decapping activity of the Dcp2 protein through interaction with the PRS binding site on the Dcp1 protein, the lack of structural data prevents the understanding of the atomic details of the binding mechanism. Due to the poor solubility of the Dcp1 protein, Borja et al., (2011) failed to identify a definitive mapping of the binding site of the Dcp1 protein to PRS of the Edc1 and Edc2 proteins. Nevertheless, upon binding of the Edc1 protein to the Dcp1 protein, it was noted that the most prominent chemical shift changes of the Dcp1 protein occur in the indole nitrogen of tryptophan, leucine and valine residues which localise in the vicinity of the putative PRS binding site. Taken together, this suggests that in addition to being a general coactivator, the Dcp1 protein may have a novel role of coupling pathway specific coactivators of decapping to the Dcp2 protein to further enhance its decapping activity and substrate recognition (Borja et al., 2011). Thus, the Dcp1 protein serves as an additional control point for decapping and 5' - 3' decay.

1.2.6.1.2 *Enhancer of decapping 3*

While the Edc1 and Edc2 proteins' stimulation of decapping activity are specific to *S. cerevisiae*, the role of the Edc3 protein in stimulating decapping is conserved throughout eukaryotes (Cougot et al., 2004; Dunckley et al., 2001; Fenger-Gron et al., 2005; Kshirsagar and Parker, 2004; Schwartz et al., 2003). The depletion of the Edc3 protein in cells impairs decapping while translation remains unaffected (Badis et al., 2004; Coller and Parker, 2005; Kshirsagar and Parker, 2004). Furthermore, as the Edc3 protein was shown to interact directly with the Dcp2 and Dcp1 proteins, it is proposed that the Edc3 protein might recruit or activate the decapping complex on target mRNAs (Decker et al., 2007;

Tritschler et al., 2007). Finally, the C terminal dimerisation domain, Yjef-N, of the Edc3 protein promotes interaction of mRNPs, which is critical for formation of the P bodies (Decker et al., 2007; Tritschler et al., 2007).

1.2.6.1.3 Pat1

The Pat1 protein stimulates decapping directly by its association with the Lsm1-7 protein complex (detailed below). Additionally, the Pat1 protein also interacts with the Dhh1 protein (detailed below) and stimulates decapping indirectly by promoting translation repression of mRNPs. Additionally, the Pat1 protein is found to co-immunoprecipitate with the Pabp1 and eIF-4E proteins in a RNA-dependent manner (Tharun and Parker, 2001). Subsequently, the overexpression of the Pat1 protein causes translation repression and P body formation (Coller and Parker, 2005). These results suggest that Pat1 may associate with translationally competent mRNPs and facilitate the removal of translational machinery as well as subsequent assembly of decapping machinery.

1.2.6.1.4 Lsm1-7 complex

The Lsm1-7 protein complex is a decapping activator and it is made up of seven Sm-like proteins. The complex is conserved in all eukaryotes and it associates with the Pat1 protein to form a hetero-octameric ring complex, which binds to mRNA after deadenylation (Coller and Parker, 2004). The Lsm1-7-Pat1 complex increases the efficiency of decapping *in vivo* and it is proposed to facilitate the assembly of the decapping complex (Bouveret et al., 2000; Salgado-Garrido et al., 1999; Tharun, 2009; Tharun et al., 2000). In addition, it is suggested that part of the decapping stimulating effect of the Lsm1-7-Pat1 protein complex is due to its ability to recognise and bind preferentially to oligoadenylated mRNA at its 3' end (Chowdhury et al., 2007). Correspondingly, mutations in the Lsm1 protein, that abolished the complex ability to recognise the oligoadenylation status of the mRNA, leads to a reduction in mRNA decay (Chowdhury and Tharun, 2008). In addition, the Lsm1-7 protein complex is also implicated in the mammalian 3' uridylation mediated decay (Song and Kiledjian, 2007). It was shown that it binds specifically to the 3' end oligo U-tract and activate decapping at the 5' end of mRNA, while inhibiting 3' - 5' decay. Thus,

this demonstrates that the Lsm1-7 protein complex association with the oligo U-tract leads to decapping dependent 5' - 3' decay of the targeted mRNA.

1.2.6.1.5 Dhh1

The Dhh1 protein is part of a family of DexD/H box ATPases. Through pull down assay and subsequent mass spectrometry analyse, the Dhh1 human homolog, Rck/p54 protein is found to be an interaction partner of the human Dcp1 protein in the presence of the Edc3 protein (Fenger-Gron et al., 2005). Thus, it is proposed that the Edc3 protein bridges the Rck/p54 protein to the Dcp1 protein. Together, the formation of the Rck/p54-Dcp1-Edc3 protein complex stimulates the decapping activity of the human Dcp2 protein.

Alternatively, it was observed that deletion of the Dhh1 and Pat1 proteins prevents translational repression upon glucose starvation in *S. cerevisiae* (Coller and Parker, 2005). In support of the observation, the overexpression of the Dhh1 and Pat1 proteins lead to inhibition of translation and a subsequent increase in the size of P body (Coller and Parker, 2005). Thus, this suggests that the Dhh1 and Pat1 proteins are critical in repressing the translation of mRNAs in *S. cerevisiae*, and hence stimulate decapping indirectly by increasing the pool of translationally inactive mRNP which may be substrates for decapping complex (Coller and Parker, 2005).

Coller and Parker, (2005) have proposed that the Dhh1 protein employs two mechanisms to inhibit translation. First, the Dhh1 protein is a member of the DEAD box family of RNA helicases. After binding to RNA, the Dhh1 protein may harness the energy from ATP hydrolysis to rearrange the structure of the mRNP to facilitate its association with the translation repression complex. Subsequently, the recruitment of translation repression complex on mRNP may trigger a cascade of events that cause the mRNP to become translationally repressed. Second, the Dhh1 protein may bind to a translation factor or the 40S ribosomal subunit, and inhibit their function in translation.

Several observations suggest that the function of the Dhh1 protein in *S. cerevisiae* is also conserved in other eukaryotes. First, it was observed that the Dhh1 protein is highly conserved having 66% identical and 81% similar residues with its human homolog, the RCK/p54 protein. Second, the RCK/p54 protein is

shown to repress translation *in vitro* and it was implicated in silencing by miRNAs (Chu and Rana, 2006; Collier and Parker, 2005). Finally, the RCK/p54 protein accumulates in P bodies (Cougot et al., 2004) and knocking down of the RCK/p54 protein using siRNA reduce P bodies formation (Andrei et al., 2005).

1.2.6.1.6 Hedls

The Hedls protein has no known homologue in yeast. Nevertheless, it is shown that the *D. melanogaster* Hedls ortholog, Ge-1, is a decapping activator that is required for silencing ~15% of AGO1 targets in miRNA mediated decay (Eulalio et al., 2007). Furthermore, co-depletion of the Ge-1 protein with other decapping activators like the Dcp1 protein is sufficient to relieve the silencing effect of miRNA, which suggests that one of the mechanisms employed by miRNA to regulate protein expression is through decapping of transcripts (Eulalio et al., 2007). It is subsequently revealed that Hedls and its orthologs in human, *D. melanogaster* and *Arabidopsis thaliana* localise to P bodies and they are required for P body integrity (Eulalio et al., 2007; Fenger-Gron et al., 2005; Xu et al., 2006; Yu et al., 2005). Accordingly, overexpression of the wild type Hedls protein, but not a mutant Hedls with deletion of the Asp/Asn rich domain, leads to the assembly of P bodies-like structures in human cells (Fenger-Gron et al., 2005; Yu et al., 2005). In addition, co-expressed Hedls protein was required for association of the hDcp1a protein with the Dcp2 protein. This suggests that Hedls is a scaffold protein that bridges the two proteins, enhancing the decapping efficiency of the Dcp2 protein (Fenger-Gron et al., 2005). Alternatively, it was observed that the human Dcp2 accumulates at 10 - to 20 - fold in the presence of the Hedls protein, suggesting that it may mediate the folding or modification to the Dcp2 protein so that it can form a stable complex with the hDcp1a protein. Accordingly, it was observed that the Hedls protein forms a complex with the Dcp2 protein and the Hedls protein serves as a decapping activator that enhances the Dcp2 protein decapping activity *in vitro* (Fenger-Gron et al., 2005).

1.2.7 *Alternative decapping enzymes*

While the Dcp1-Dcp2 protein complex is the only decapping enzyme in *S. cerevisiae* (Beelman et al., 1996; Dunckley and Parker, 1999), three studies show that mammals may possess other novel decapping enzymes. First, the Dcp2 protein is differentially expressed in different organs in mouse and it was absent in mouse's heart, liver, kidney and muscle. Similarly, Dcp2 protein is also absent in the liver in humans (Song et al., 2010). Second, the global analysis of the Dcp2 dependent changes in mRNA stability revealed that the half-lives of mRNA in mouse embryonic fibroblast cells were only moderately altered upon the Dcp2 protein knockdown. Third, a novel decapping enzyme, Nudt16, is expressed in mammalian cells and it regulates a specific subset of mRNAs which is involved in cell motility and migration (Song et al., 2010; Taylor and Peculis, 2008; Troyanovsky et al., 2001). The discovery of the Nudt16 protein suggests that there may be additional decapping enzymes present in metazoans.

1.2.8 *Decapping in the nucleus*

While mRNA decapping enzymes are mainly involved in the cytoplasmic mRNA turnover, Brannan et al., (2012) reported that the Dcp2, Dcp1a, Edc3 proteins are also found in the nucleus (**Figure 17**). These decapping proteins interact with termination factor TTF2 (Leonard et al., 2003) and the nuclear 5'-3' exonuclease "torpedo" Xrn2 proteins and they localised at the transcription start sites (TSSs). A proposed "torpedo" model of premature transcription termination was initiated by the decapping of nascent transcripts near sites of promoter-proximal RNA Polymerase II (pol II) pausing. Co-transcriptional decapping exposes a 5' monophosphate end on nascent RNA and provides an entry point for the Xrn2 exonuclease. Subsequently, the Xrn2 protein degrades the transcript and displaces pol II from the DNA template which result in premature termination of transcription (Brannan et al., 2012). In addition, the "torpedo" model was facilitated by the TTF2 protein, which physically interacts with the Dcp1a protein (Leonard et al., 2003) and it is found to displace pol II from template DNA *in vitro* (Liu and Doetsch, 1998). Moreover, the "torpedo" model affects thousands of genes and thus it is proposed to be a widespread gene expression regulation mechanism to limit productive pol II elongation of human genes.

Figure 17. The promoter proximal “torpedo” model for premature termination of pol II transcription.

The “torpedo” model is a widespread gene expression regulation mechanism that occurs in the nucleus during transcription. Decapping occurs at the 5' end of newly transcribed mRNA to allow TTF2 facilitated Xrn2 protein to degrade the transcript and displaces pol II (yellow) from the template DNA (left panel). Polymerases that escapes decapping are bound by CBC where productive elongation continues (right panel). Figure is from Brannan et al, 2012.

1.3 Enabled/Vasp Homology 1 domain

The EVH1 domain is an approximately 115 residues protein-protein interaction module commonly found in proteins that are involved in the reorganisation of the actin cytoskeleton and the modulation of actin dynamics for actin based cell motility (Ball et al., 2002). The EVH1 domain is also found to be involved in the signal transduction in postsynaptic compartments of the chemical synapse, suggesting that the EVH1 domain plays a part in synaptic plasticity, which affects memory formation and learning (Ball et al., 2002; Beneken et al., 2000).

1.3.1 Overall fold

The overall fold of the EVH1 domain consists of a parallel β sandwich, which is closed by an α helix. In addition, two concave surfaces are formed on the solvent exposed side of β sandwich and this structure is highly similar to that of the pleckstrin homology (PH) and the phosphotyrosine-binding domains (PTB). In all the classes of the EVH1 domain, a highly conserved aromatic triad of residues (Tyr16, Trp23 and Phe77 in Mena, Phe14, Trp24 and Phe74 in Homer and Tyr46, Trp54 and Phe104 in N-Wasp) is critical in forming the recognition site of their PRS ligand (**Figure 18**) (Beneken et al., 2000; Prehoda et al., 1999; Volkman et al., 2002). Trp23 in the Mena protein remains invariant in all the classes of the EVH1 domain and its sidechain is positioned between Phe77 and Tyr16 to form an exposed hydrophobic surface in precise stereochemistry to dock proline rich sequence (PRS) ligands. Nevertheless, the PRS recognition surface is varied across all classes of EVH1 domain and this suggests a possible mechanism to for specifying of PRS ligands specific to the different classes of EVH1 domain.

Figure 18. PRS ligands binding to respective classes of EVH1 domains.

The proline rich motif of PRS ligands adopt a PPII helix conformation and binds to a triad of aromatic residues (Tyr16, Trp23 and Phe77 in Mena, Phe14, Trp24 and Phe74 in Homer and Tyr46, Trp54 and Phe104 in N-Wasp) which form part of the PRS binding site in all classes of EVH1 domains. Figure is modified from Volkman et al, 2002.

1.3.2 Proline rich sequence ligand binding

1.3.2.1 Polyproline II structure

The target PRS ligand spontaneously form a unique secondary structure known as the polyproline helix II (PPII) structure ($\phi = -78^\circ$, $\psi = +146^\circ$) which is a left handed helix with three residues per turn (Ball et al., 2002). The PPII structure is favoured in sequences with multiple prolines due to the formation of a covalent bond between the side chain and main chain of proline, which greatly reduces the main chain torsional conformations available. The binding of a pre-formed structure to a protein surface results in a significant smaller loss of configurational entropy, which require fewer interactions for compensation (Petrella et al., 1996). Thus, the PPII structure is favoured as binding targets. This structure is triangular in cross section and it contains a proline residue in at least every third amino acid position. This ensures that the pyrrolidine rings at one edge of PPII prism forms a continuous hydrophobic patch. In Class I EVH1 domain, the ligand maintains its PPII structure while the ligand in class II EVH1 domain has its PPII structure distorted. The differences in structure can be attributed to the unique interaction formed by Phe6 (Homer numbering) of the PRS ligand with the PRS binding site of Homer (Beneken et al., 2000).

1.3.2.2 Binding mode

Structures of different classes of EVH1 domains with their respective ligands showed that the mechanism of the proline motif recognition is dependent on the PPII conformation of the peptide and not the specific identity of residues. In support of the proposed binding mode, the binding interface on EVH1 domains complements the 3-fold helical screw axis adopted by the PRS ligand. Furthermore, the presence of non-proline amino acids in these PRS does not affect formation of PPII secondary structure (Ball et al., 2002) and hence they are still capable of binding EVH1 domain. This suggests that even though EVH1 domains have a preference for high-density proline ligands with PPII conformation, it does not require prolines to occupy specific position. In addition, the PPII structure has a 2-fold pseudosymmetry, which allows the protein to bind in either orientation.

In general, the core of PRS ligand forms a proline rich motif that recognises and binds to a homologous hydrophobic pocket on EVH1 domain with low affinity. More importantly, residues flanking the core region of PRS can form additional affinity increasing interactions with residues of the EVH1 domain (Ball et al., 2002). Hence, these flanking residues play a crucial role in determining the binding orientation of the protein and also in modulating the specificity of PRS ligand binding to different classes of EVH1. Furthermore, different classes of EVH1 domain also employ distinct but overlapping binding sites, which are conserved among the same subset of EVH1 domains. Studies have shown that residues localised in these subset specific sites are important in forming critical interactions with their respective PRS ligand (Beneken et al., 2000; Prehoda et al., 1999; Volkman et al., 2002). An example will be Gly89 in Homer EVH1 domain. Due to the absence of side chain in Gly89, phenylalanine of Homer ligand is able to pack directly on Gly89. Mutation of Gly89 to either alanine or asparagine disrupts the binding of Homer ligand to its EVH1 domain. Furthermore, structural comparison of EVH1 domains shows that all non Homer EVH1 domains possess a different amino acid (typically asparagine) at this position, which suggests that other classes of EVH1 are unlikely to bind Homer ligand and hence prevent cross-reaction. Taken together, the variations of flanking residues of the core PRS ligand and minor structural differences on the EVH1 PRS binding site contribute to unique ligand binding mode for respective

classes of EVH1 domains. This causes different target specificity among various EVH1 domains, which will minimise the probability of cross-reaction of diverse proline rich ligands to different classes of EVH1 domain.

1.3.3 Classification

The classification of the EVH1 domains is based on the consensus sequences of the target PRS ligands. Currently, there are three classes of the EVH1 domains (Ball et al., 2002).

1.3.3.1 Class I EVH1 domain

The class I EVH1 domain comprises of the Ena/Vasp family of adaptor proteins and they couple signalling pathways to actin assembly. This is achieved through their function as scaffold proteins to recruit or activate downstream actin assembly machinery (Beckerle, 1998; Machesky et al., 1997; Pollard, 1995). In support of its adaptor function, the overall structural organisation of Ena/Vasp family of proteins revealed that the proteins have a specific region to bind with upstream signalling proteins and another distinct region to associate with components of the actin assembly machinery (Gertler et al., 1996; Symons et al., 1996). Given its importance in actin assembly, the Ena/Vasp family proteins are required for diverse actin-based processes from budding, cytokinesis in yeast (Li, 1997), neutrophils chemotaxis (Symons et al., 1996) and axon guidance in mammals (Lanier et al., 1999).

The class I EVH1 domain protein recognises the binding motif FPX ϕ P (where X is any residue and ϕ is a hydrophobic residue) and such sequences can be found in adhesion proteins, such as zyxin, vinculins and axon guidance protein, SAX-E/Robo. (Prehoda et al, 1999; Niebhur et al, 1997; Kidd et al, 1998; Zallen et al, 1998). Interestingly, the motile intracellular pathogen, *Listeria monocytogenes*, uses its surface protein, ActA, which comprises of multiple EVH1 binding motifs to recruit Ena/Vasp family of proteins, the Mena and Vasp proteins, to the base of its *Listeria* comet tail. Subsequently, the Mena and Vasp proteins recruit the actin assembly machinery to generate a propulsive tail to facilitate the movement and infection of neighbouring host cells (Chakraborty et

al., 1995; Laurent et al., 1999; Niebuhr et al., 1997; Theriot, 1994; Tilney and Portnoy, 1989).

The structure of the Mena protein in complex with ActA peptide showed that the FPPP core of the ActA peptide packed against the Mena EVH1 domain with one of its pointed end directed into a V shaped wedge (Prehoda et al., 1999). The V-shaped wedge of the Mena protein consist of a conserved aromatic triad of residues Tyr16, Trp23 and Phe77 and they engage in hydrophobic interaction with the proline rich core of the ActA peptide. While the C-terminal acidic residues are not visible in the structure, it is speculated that they are involved in the interaction with the basic residues of the Mena EVH1 domain that is immediately adjacent to the C-terminal of the Acta core peptide. The interaction of the flanking acidic residues with the EVH1 domain increases the strength of association between the Mena protein and the ActA peptide. In support of the proposed model, quantitative fluorescence perturbation assay demonstrated that minimal “core” peptide of ActA binds the Mena protein with low affinity ($K_d > 400\mu\text{M}$). Nevertheless, when C terminal acidic residues of ActA flank these core peptides, interaction between the Mena protein and the Acta peptide increased significantly ($K_d > 5\mu\text{M}$) (Niebuhr et al., 1997). Furthermore, the single Phe residue flanking the core motif is critical in setting the register of binding and hence, maintains an unambiguous directionality of the PRS ligand. Mutation of this Phe to any other amino acids with the exception of Trp will result in the loss of binding to EVH1 domain (Ball et al., 2002).

1.3.3.2 Class II EVH1 domain

The class II EVH1 domains consist of the Homer-Vesl protein family of postsynaptic receptor associated proteins whose expression is up-regulated in neurons after synaptic activity. This suggests that the Homer protein is a potential mediator of synaptic plasticity and thus it may play a role in learning and memory (Brakeman et al., 1997; Kato et al., 1997). This class of EVH1 domains localise to postsynaptic density and functions as adaptor proteins that cross link and bind to core motif, PPXXF (Tu et al., 1998), which are present in group I metabotropic glutamate receptors (mGluRs), inositol-1, 4, 5-triphosphate receptors (IP3Rs), ryanodine receptors (RyRs) and Shank family proteins (Beneken et al., 2000).

High-resolution crystal structure of the rat Homer EVH1 domain in complex with a mGluR-derived peptide, TPPSPF, has been solved (Beneken et al., 2000). Two major differences exist between the Mena and the Homer PRS binding mode. First, unlike the Mena PRS ligand, the PRS ligand of the Homer EVH1 domain does not uniformly adopt the PPII conformation. Instead, it incorporates a tight beta turn at its C terminus to appropriately position the prolines and C terminus phenylalanine for its interaction with Homer EVH1 domain. Second, the Homer protein uses an overlapping but distinct PRS binding site from the Mena protein, to bind its PRS ligand. His12 and Phe14 (Homer numbering) are critical residues that bind PRS ligand and they are localised on a distinct PRS binding site that is exclusively present on the Homer EVH1 domain. Taken together, the structure of Homer protein in complex with its PRS ligand has shown that Homer utilises a different binding mode for its PRS ligand, and thus it is classified as class II EVH1 domain.

1.3.3.3 Class III EVH1 domain

The homolog of the Wiskott-Aldrich Syndrome protein (Wasp), N-Wasp, is a signal transduction protein that link actin polymerisation to upstream intracellular signals (Carlier et al., 1999; Higgs and Pollard, 1999; Millard and Machesky, 2001; Pollard et al., 2000). The B domain of the N-Wasp protein has been shown to mediate phosphatidyl inositol-4-phosphate (PIP) and phosphatidyl inositol-4, 5-phosphate (PIP2) regulation in the N-Wasp protein (Prehoda et al., 1999; Rohatgi et al., 2001). Through the interaction with both peptide motifs of actin assembly machinery and membrane phospholipids, the N-Wasp protein may be critical in targeting the actin assembly machinery to the leading edge sites of the cell. Moreover, the Wasp protein is mainly expressed in hematopoietic cells and mutation of this protein leads to the Wiskott-Aldrich syndrome (WAS), which is a X-linked recessive disorder. Symptoms of WAS includes immunodeficiency, eczema and thrombocytopenia and they are caused by cytoskeletal defects in hematopoietic cells (Derry et al., 1994; Greer et al., 1996; Kolluri et al., 1995; Villa et al., 1995; Zhu et al., 1997; Zicha et al., 1998). Similarly, knockout mutation of the N-Wasp protein also leads to embryonic lethality and defects in many actin-based motility processes (Snapper et al., 2001).

It was also observed that 28 out of 35 identified missense mutations of the EVH1 domain at the N terminus of N-Wasp is responsible for the onset of WAS (Derry et al., 1994; El-Hakeh et al., 2002; Greer et al., 1996; Kolluri et al., 1995; Zhu et al., 1997). In addition, the hotspot of the disease mutation associated with severe WAS is found at a surface opposite the PRS binding site of the N-Wasp protein (Beneken et al., 2000; Prehoda et al., 1999). This provides evidence that N-Wasp EVH1 domain is a critical domain that affect the onset of WAS.

The nuclear magnetic resonance (NMR) structure of the N-Wasp EVH1 domain in complex with a 25 residues peptide from the Wasp interacting protein (WIP) shows a novel-binding mode whereby the WIP peptide wraps around the EVH1 domain (Volkman et al., 2002). As the WIP peptide is twice as long as the ligands of other classes EVH1 domain, it contacts regions of the N-Wasp protein that extends beyond the canonical PRS binding site of the EVH1 domain. Nevertheless, N-terminal terminus of WIP, LPPP residues, still adopts the PPII helical conformation and binds to the canonical PRS binding site. However, the WIP peptide binds in an opposite orientation as compared to the PRS ligand of the Mena and Homer proteins. Taken together, the N-Wasp EVH1 domain shows a different binding mode to its PRS ligand and it is regarded as the third class of EVH1 domain.

In addition, the structure of the N-Wasp EVH1 domain gives insights on how WAS arise due to mutations of residues residing in the hotspot of disease mutation. First, mutation of the buried residues leads to destabilisation of the EVH1 domain structure and ultimately disrupt the precise stereochemistry of the residues lining the PRS binding site. Second, due to the long length of WIP, surface residues at the hotspot of disease mutation directly contact WIP residues. Thus, mutation of these surface residues impairs the N-Wasp protein ability to bind the WIP protein, leading to onset of WAS.

1.3.3.4 A novel class of EVH1 domain

The high resolution crystal structure of the Dcp1 protein (She et al., 2004) was determined. Subsequent comparison of PRS binding sites from other members of EVH1 domains suggests that the Dcp1 protein belongs to a novel class of EVH1 domain. However, a structure of the Dcp1 protein in complex with

its elusive PRS substrate is required to validate this hypothesis. In addition, the presence of EVH1 domain in the Dcp1 protein may also suggest that it may have an additional function of linking the Dcp2 protein to novel protein that activate the Dcp2 protein decapping activity.

Additionally, it was suggested that PRS binding site of EVH1 domain use a wider interface to recognise extended epitope. Borja et al., (2011) identified a conserved Trp204 of the Dcp1 protein, which lies outside the conventional PRS binding site. Additionally, mutation of Trp204 impairs the binding of PRS of the Edc1 protein to the Dcp1 protein. Furthermore, it is noted that the Dcp1 protein binds to an extended consensus sequence, which is unique from those described for other classes of EVH1 domain.

1.3.4 Promiscuity of EVH1 domain

The EVH1 domains have been shown to interact with a variety of ligands. Class II EVH1 domain, the Homer protein, binds to mGluRs, IP3Rs and Shank proteins (Beneken et al., 2000). Furthermore, *S. cerevisiae* Dcp1 protein binds to the Edc1 and Edc2 proteins while its human homologue, the Dcp1a protein, binds to the Smad4 and PNRC2 proteins (Bai et al., 2002; Borja et al., 2011; Cho et al., 2009). The promiscuity of the EVH1 domains allow proteins to form unique complexes that are specially configured for different metabolic pathways, and also to be regulated by an extensive network of proteins.

1.3.5 Therapeutics

The high resolution structures of EVH1 domains in complex with their respective PRS ligands have increased our understanding of EVH1-mediated interaction and identified residues that govern specificity and binding affinities. Using this knowledge, inhibitors that interfere with the binding between EVH1 domains and their PRS ligands can be designed. These inhibitors can be useful in dissecting biochemical pathways and serve as molecular tags to monitor the association and disassociation of EVH1 domain binding within the cells. More importantly, given the diverse role that EVH1 domains play in an organism, the inhibitors may help in the treatment of diseases related to binding of EVH1

domain, which includes the spreading of intracellular pathogen, pathologically altered adhesion and motility in inflammatory and metastatic cancers.

1.4 Proline rich nuclear receptor co-activator 2

1.4.1 *Role as nuclear receptor coactivator*

The PNRC2 protein is first identified as a nuclear receptor co-activator that possess a proline-rich Src homology domain 3 (SH3) binding motif, SEPPSPS, and a short hydrophobic motif, LKTLL, known as the NR-box (Zhou and Chen, 2001; Zhou et al., 2000; Zhou et al., 2006). It is shown that the PNRC2 protein uses its SH3 domain binding motif to interact with nuclear receptors to bind specific hormone-responsive elements in promoters to activate transcription of target genes and regulate their expression (Zhou and Chen, 2001; Zhou et al., 2000; Zhou et al., 2006). The PNRC2 protein was reported to bind orphan receptors SF1 and estrogen receptor-related receptor α 1 in a ligand independent manner. In addition, the PNRC2 protein can also bind to the ligand binding domain of the estrogen receptor, glucocorticoid receptor, progesterone receptor, thyroid receptor, retinoic acid receptor and retinoid X receptor in a ligand-dependent manner.

1.4.2 *Role in Stau1-mediated mRNA decay*

The Stau1 protein is a RNA binding protein which functions in mRNA transport, translational control, DNA replication and mRNA decay (Kim et al., 2005). Cho et al., (2012) demonstrated that SMD involves the Stau1 and PNRC2 proteins, which leads to subsequent decapping and 5' - 3' exonucleolytic activities. SMD is distinct from the mammalian NMD as it occurs independently of splicing and it is unaffected by downregulation of the Upf2 or Upf3 proteins (Kim et al., 2005). Similarly, NMD is not affected by downregulation of the Stau1 protein. Subsequently, a model was proposed where SMD is initiated by the binding of the Stau1 protein to approximately 230 nt region, known as Stau1 binding site (SBS) which is located at the 3' UTR of the target mRNA (Gong et al., 2009; Gong and Maquat, 2011; Kim et al., 2005; Kim et al., 2007). Upon the binding of the Stau1 protein, the interaction between the N-terminal of the hyperphosphorylated Upf1 and Stau1 proteins caused Upf1 to be recruited to SBS (Kim et al., 2005). In addition, as C-terminal of hyperphosphorylated Upf1 protein binds to the PNRC2 protein, the Upf1 protein serves to act as an adaptor

protein to bridge the Stau1 and PNRC2 protein together at the SBS (Cho et al., 2012). In turn, the PNRC2 protein associates with the human Dcp1a protein and hence recruits the decapping machinery to trigger the degradation of the target mRNA (Cho et al., 2009; Cho et al., 2012). Finally, two independently conducted microarray analyses showed that at least 22 293 mRNAs bind to the Stau1 protein and hence they are potential targets of SMD (Kim et al., 2005).

1.4.3 Role in adipogenesis and energy expenditure

Adipogenesis is a cell differentiation process where preadipocytes differentiate into adipocytes. The adipose tissue plays a role in hemostasis, energy balance and glucose homeostasis (Rosen and MacDougald, 2006). The PNRC2 protein was found to affect adipogenesis as downregulation of the mouse PNRC2 protein reduced expression levels of the mouse peroxisome proliferator-activated receptor γ (mPPAR γ) and caveolin-1 proteins. The reduction in expression of the two proteins attenuates adipogenesis which leads to a reduction in the rate of accumulation of cytoplasmic fat (Cho et al., 2012). One of the possible ways that the PNRC2 protein can accomplish this role is through forming a stable complex with the hyperphosphorylated Upf1 and Stau1 proteins to increase efficiency of SMD. The Kruppel-like factor 2 (KLF2) mRNA, which encodes an antiadipogenic factor that targets mPPAR γ mRNA for degradation, is shown to be a target of SMD (Banerjee et al., 2003; Cho et al., 2012; Rosen and MacDougald, 2006; Wu et al., 2005). Thus, SMD of KLF2 mRNA increases expression of the PPAR γ protein, which eventually accelerates adipogenesis.

The generation of PNRC2 null mice showed that the mice have higher metabolic rate and they are resistant to high fat diet-induced obesity (Zhou et al., 2008). Further analysis of the PNRC2 null mice suggests that the absence of the PNRC2 protein leads to increased energy expenditure. Hence, the PNRC2 protein is proposed to play an important role in controlling the energy balance between energy storage and expenditure.

1.4.4 Role in aromatase activation

The aromatase is a key enzyme involved in the synthesis of the estrogen hormone, which influence breast tumor growth and maintenance in mammary tumorigenesis (Lu et al., 1996; Miller, 1996; Rajhans et al., 2008). Accordingly, deregulation of PELP1, an estrogen receptor coregulator, leads to an upregulation of aromatase activity which causes breast tumours from postmenopausal women to have a higher amount of estrogen than the basal levels of estrogen circulating in plasma (James et al., 1987). As such, aromatase inhibitors are used in the therapy of mammary tumours with some success (Jordan and Brodie, 2007). Using yeast two hybrid screen, the PNRC2 protein was identified as the interacting partner of PELP1. Immunoprecipitation and luciferase assays further showed that the PELP1 protein activate aromatase expression by forming a functional complex with the ERR α and PNRC2 proteins *in vivo* (Rajhans et al., 2008).

1.5 Rationale of my study

Numerous studies in the field of mRNA decapping and NMD reveal many important insights with regards to the recognition of PTC on aberrant transcripts and their subsequent decay. Nevertheless, the accumulation of knowledge in these fields have also highlighted several knowledge gaps that remain to be solved:

- 1) After the PTC recognition, recruitment of decapping enzyme to the aberrant transcripts remains poorly understood. Till date, the binding of the human Upf1 protein to the Dcp1a, Dcp2 or PNRC2 proteins has been demonstrated through co-immunoprecipitation and yeast two hybrid assays. However, direct interactions between the proteins have not been shown.
- 2) While the structure of the *S. cerevisiae* Dcp1 protein has been visualised at atomic details, the binding mode between the Dcp1 protein and its PRS ligands (PNRC2, Edc1, Edc2 and Smad4) has not been elucidated. Thus, the Dcp1 EVH1 domain remains to be classified.
- 3) While the PNRC2, Edc1, Edc2 and Smad4 proteins are shown to be interaction partners of the hDcp1a protein, the proline rich consensus sequence has not been determined yet.
- 4) The interaction between the human Dcp1a and Dcp2 proteins is weak and it requires additional factors like the Hedls, Edc3 and TTP proteins to form distinct decapping complexes that specifically target their own unique subset of mRNAs. Similarly, as the PNRC2 protein is an interaction partner of the Dcp1a protein, it is postulated to be an adaptor protein that can stimulate decapping activity. Nevertheless, the decapping stimulatory effect of the PNRC2 protein has not been demonstrated.
- 5) The Dcp1a and PNRC2 proteins are bifunctional proteins that are involved in both the transcription and decay of mRNA. Given the contradictory functions, the activity of the Dcp1a and PNRC2 proteins must be precisely controlled. Nevertheless, the mechanism utilised by the Dcp1a and PNRC2 proteins to

switch their function from mRNA transcription to mRNA decay remains to be solved.

In order to provide answers to the prevailing knowledge gaps, a structure of the Dcp1a protein in complex with the PRS of PNRC2 has been determined by X-ray crystallography. The results were analysed and they formed the basis of further biochemical assays performed *in vitro* and *in vivo* to support our hypothesis.

2 Materials and Methods

2.1 Molecular cloning

2.1.1 Polymerase chain reaction (PCR) and analysis

2.1.1.1 PCR amplification

Constructs with the following gene fragments were made:

Human Dcp1a EVH1 (hDcp1aEVH1) (1-130 aa)

Human PNRC2 Δ NR (PNRC2 Δ NR) (1-121 aa)

Human PNRC2 full length (PNRC2 (FL)) (1-139 aa)

Human Dcp2 (hDcp2) (1-245 aa)

The constructs were amplified through PCR with the use of the human universal cDNA library (Clontech) as the template. PCR were performed on an iCycler thermal cycler (Bio-Rad) using the following protocol:

Reaction mixture components	Volume (μ l)
Template DNA (50ng)	1
10x Pfu buffer(Stratagene)	5
Turbo Pfu DNA polymerase (2.5 units/ μ l)	1
10 μ M 5' primer (Proligo, Sigma-Aldrich)	1
10 μ M 3' primer (Proligo, Sigma-Aldrich)	1
100 mM MgSO ₄	0.5
25 mM dNTPs	1
Distilled water	39.5
Total	50

Cycling parameters		
95°C	5 min	1 cycle
95°C	30 sec	35 cycles
55°C	45 sec	
72°C	3 min	
72°C	10 min	1 cycle
4°C	∞	

The molecular weight of each PCR product was checked by subjecting 5 μ l of each sample to agarose gel electrophoresis on a 1.2% (w/v) agarose gel in 1x TAE running buffer (40 mM Tris pH 8.0, 1.14% (v/v) acetic acid, 1 mM EDTA). Subsequent purification of PCR products was performed using QIAquick PCR purification kit according to manufacturer's protocol.

2.1.2 Construct sub-cloning

Purified PCR products and vector were subjected to restriction enzyme digestion to generate complementary restriction sites to clone gene constructs into vector. The restriction enzyme digestion was performed at 37°C for 3 hours using the following protocol:

Reagent	Volume (μ l)	
	Vector	PCR fragment
DNA	13	30
10x Buffer (New England Biolabs)	2	5
Enzyme 1 (New England Biolabs)	1	2
Enzyme 2 (New England Biolabs)	1	2
100x BSA	0.2	0.5
Distilled Water	2.8	10.5
Total	20	50

After the restriction enzyme digestion, the digested products were separated using agarose gel electrophoresis as described above. The respective digested bands of vector or insert DNA fragments were subsequently excised and purified using QIAquick Gel Extraction kit according to manufacturer's protocol.

2.1.3 Ligation

The two vectors used in this thesis are pETDuet-1 (Novagen) and pGEX-6P-1 (GE Amersham Biosciences). PNRC2 Δ NR (1-121 aa) and hDcp1aEVH1 (1-130 aa) were cloned into the multiple cloning site (MCS) 1 and MCS2 of pETDuet-1 vector (Novagen) respectively while the ligation of insert hDcp2 DNA fragments into MCS of pGEX-6P-1 vector were performed at 15°C overnight according to the following protocol:

Reaction mixture components	Volume (μ l)
Double digested vector	4
Double digested DNA insert	12
10xT4 buffer (New England Biolabs)	2
T4 DNA ligase (New England Biolabs)	2
Total	20

2.1.4 DH5 α transformation and positive clone selection

Using the heat-shock method, 10 μ l of ligated product was transformed into 100 μ l of E.coli DH5 α (Invitrogen) high efficiency competent cells. Accordingly, the heat-shock method was performed as followed. Frozen DH5 α competent cells were incubated on ice for 20 minutes before addition of ligated product. Subsequently, they were subjected to heat shock treatment in a 42°C water bath for 90 seconds followed by 2 minutes of incubation on ice. After incubation, 1 ml of Luria-Bertani (LB) medium was added to the cells and the cells were further incubated at 37°C for an hour. Following that, the cells were plated on 1.5% LB agar plate containing 100 μ g/ml ampicillin and incubated at 37°C overnight. Positive clones were picked and inoculated into 3 ml of LB medium. DNA plasmids were subsequently extracted with QIAprep Spin Miniprep kit and verified by DNA sequencing. The PCR sequencing reaction was performed as follows:

Reaction mixture components	Volume (μ l)
Big Dye V31 sequencing mixture	8
Plasmid	4
10 μ M Forward or Reverse primer	1
Distilled Water	7
Total	20

Cycling parameters		
95°C	1 min	1 cycle
95°C	10 sec	25 cycles
50°C	5 sec	
60°C	4 min	
15°C	∞	

Subsequent analysis of DNA sequences were carried out on ABI Prism 377 DNA sequencer (Perkin Elmer) by staff from the DNA sequencing facility at the Institute of Molecular and Cell Biology.

2.1.5 Site directed mutagenesis

Eight mutants were created on the co-expression plasmid (His-PNRC2 Δ NR-Dcp1aEVH1) described (Section 2.1.3) using the QuikChange Mutagenesis method (Stratagene), and verified by DNA sequencing (**Table 1**). Subsequently, the cDNA inserts encoding Dcp1aEVH1 and PNRC2 Δ NR were swapped, resulting in a second co-expression plasmid harbouring His-Dcp1aEVH1 and PNRC2 Δ NR in MCS1 and MCS2 of the pETDuet-1 vector (His-Dcp1aEVH1-PNRC2 Δ NR), respectively. Three mutants were created on the second co-expression plasmid using the same method as described above.

Table 1. Generation of point mutations of residues lining PNRC2 and Dcp1aEVH1 interface

Plasmid 1	His PNRC2ΔNR (MCS1)	Dcp1aEVH1 (MCS2)
1	W114A	WT
2	WT	Y36A
3	WT	F38A
4	WT	W45A
5	WT	F94A
6	WT	L96A
7	WT	R98A
8	WT	I104A
Plasmid 2	His Dcp1aEVH1 (MCS1)	PNRC2ΔNR (MCS1)
1	WT	K109A
2	WT	P110A
3	WT	W114A

2.2 Protein expression

2.2.1 Rosetta (DE3) transformation

Sequence verified plasmids were transformed into *E.coli* Rosetta (DE3) competent cells (Invitrogen) using the heat-shock method as described (Section 2.1.4). Subsequently, 1 μ l of plasmid was inoculated with 50 μ l of *E.coli* Rosetta (DE3) instead. Single colonies were inoculated into 3 ml LB medium containing 100 μ g/ml of ampicillin and grown overnight at 37°C. Cells were subsequently stored at -80°C in 20% glycerol.

2.2.2 Small scale protein expression

100 μ l of overnight culture was inoculated to 10 ml of fresh LB medium supplemented with 100 μ g/ml of ampicillin and cultured at 37°C. When the optical density at 595 nm (OD_{595}) approaches 0.5 - 0.6, 0.5 mM of IPTG was added to the culture to induce the expression of the protein of interest for 3 hours at 37°C. For analysis of protein expression, 200 μ l of uninduced cells and 100 μ l of induced cells were pelleted and boiled in Laemmli loading buffer. Subsequently, the samples were subjected to denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) on a 12 % SDS-PAGE gel (Laemmli, 1970). Following SDS-PAGE, the 12 % SDS-PAGE gel was stained in Coomassie staining buffer (30 % (v/v) methanol, 10 % (v/v) acetic acid and 0.025 % (w/v) Coomassie Brilliant Blue R-250) for 15 minutes. The gel was then destained with 30 % (v/v) methanol and 10 % (v/v) acetic acid.

2.2.3 Large scale Rosetta (DE3) cell culture

E.coli Rosetta (DE3) harbouring the protein expression plasmid was inoculated into 50 ml of fresh LB media supplemented with 100 μ g/ml of ampicillin for overnight culture at 37°C. The small scale culture was diluted 100-fold in 5 litres (L) of fresh LB media supplemented with 100 μ g/ml of ampicillin and cultured at 37°C till the OD_{595} reached 0.5 - 0.6. IPTG was added to the culture to a final concentration of 0.5 mM at 27°C for 5 hours before harvesting. Harvesting of cells was done through centrifugation and subsequent

resuspension of the cells in buffer (150 mM NaCl and 20mM Tris-HCl pH 7.5). Finally, the cells pellets were stored at -80°C before protein purification.

In Selenomethionine (SeMet)-substituted protein, *E.coli* Rosetta (DE3) was inoculated into 50ml of fresh LB media supplemented with 100 µg/ml of ampicillin for overnight culture at 37°C. The overnight culture was diluted 100-fold in 5 L MOPS minimal media (Neidhardt et al., 1974) and the subsequent steps of large scale cell culture followed as described above.

2.3 Protein purification

2.3.1 Cell lysis

The frozen bacterial cell pellet was resuspended in 80 ml lysis buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.1 mM PMSF, 5 mM imidazole, 2 mM benzamidine) supplemented with 0.5 mg/ml hen egg lysozyme and incubated on ice for 30 minutes. The cells were homogenised by sonication (15 seconds on, 90 seconds off, for 8 cycles at 15 amplitude microns). The crude extract was separated by centrifugation at 18,000 rpm for 1 hour at 4°C using J6-HC Centrifuge (Beckman Coulter).

2.3.2 Purification of *PNRC2ΔNR-hDcp1aEVH1* complex

The purification steps of SeMet substituted *PNRC2ΔNR-hDcp1aEVH1* complex was the same as described in the sections below. However, 10 mM DTT was added to the MonoQ Buffer A, the MonoQ Buffer B and the gel filtration buffer to reduce selenium oxidation.

2.3.2.1 His tagged fusion protein affinity chromatography

The following steps were done at 4°C. After obtaining the clarified supernatant from previous procedure, the supernatant was loaded onto a pre-equilibrated TALON metal affinity resin column (Clontech) and washed extensively with the lysis buffer. The His tagged fusion protein was eluted in 20

ml elution buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.1 mM PMSF, 2 mM benzamidine and 250 mM imidazole).

2.3.2.2 Ion exchange chromatography

The remaining protein purification steps were done using the ÄKTA fast protein liquid chromatography (FPLC) system (GE Amersham Biosciences). The protein sample was diluted by addition of 30 ml ion exchange MonoQ Buffer A (100 mM NaCl, 20 mM Tris-HCl pH 7.5 and 2 mM DTT) and loaded onto a FPLC MonoQ HR 10/10 column ion exchange column (GE Amersham Biosciences). The column was first washed with the MonoQ Buffer B (1 M NaCl, 20 mM Tris-HCl pH 7.5 and 2 mM DTT) and subsequently pre-equilibrated by the MonoQ Buffer A. Protein was eluted by a 100ml linear gradient of NaCl and peak fractions were analysed by SDS-PAGE before pooling the purified fractions for a further purification step by size exclusion gel filtration chromatography.

2.3.2.3 Gel filtration chromatography

The pooled peak fractions were concentrated to 5 ml and loaded onto FPLC Superdex-200 size exclusion gel filtration chromatography column (GE Amersham Biosciences), which was previously pre-equilibrated with the gel filtration buffer (100 mM NaCl, 20 mM Tris-HCl pH 7.5, 2 mM DTT). The purity of the peak fractions was verified by SDS-PAGE before they were pooled and concentrated to a final concentration of 20 – 25 mg/ml for storage at -80°C.

2.3.3 Purification of PNRC2 (FL)-hDcp1aEVH1 complex

The purification steps of PNRC2 (FL)-hDcp1aEVH1 complex were identical to the purification steps of PNRC2 Δ NR-hDcp1aEVH1 complex. However, the pH of lysis buffer, elution buffer, MonoQ Buffer A and MonoQ Buffer B was increased from 7.5 to 8.5.

2.3.4 Purification of His-PNRC2 Δ NR (W114A)

The purification steps of His-PNRC2 Δ NR (W114A) protein were identical to that of the PNRC2 Δ NR-hDcp1aEVH1 protein complex with an exception that MonoS HR 10/10 ion exchange column (GE Amersham Biosciences) was used instead. Additionally, the buffer and pH of lysis buffer, elution buffer, MonoS Buffer A, MonoS Buffer B and the gel filtration buffer was changed to MES pH 6.0.

2.3.5 Purification of his-hDcp1aEVH1

The purification steps of his-hDcp1aEVH1 were identical to the purification steps of the PNRC2 Δ NR-hDcp1aEVH1 protein complex.

2.3.6 Purification of hDcp2 (1-245)

The cell lysis step of *E.coli* Rosetta expressing hDcp2 (1-245) is the same as that described for the PNRC2 Δ NR-hDcp1aEVH1 protein complex. However, the components of lysis buffer were different (500 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.1 mM PMSF, 2 mM benzamidine and 2 mM DTT).

2.3.6.1 GST-fusion protein affinity chromatography

Clarified supernatant was loaded onto a pre-equilibrated Glutathione Sepharose 4B column (GE Amersham Biosciences) and washed extensively with lysis buffer. GST-fusion protein was eluted in 20 ml elution buffer (500 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.1 mM PMSF, 2 mM benzamidine and 2 mM DTT, 20 mM reduced glutathione). Subsequently, a 100-fold (w/w) dilution of PreScission protease (GE Amersham Biosciences) was added to the eluted fraction for GST tag cleavage at 4°C overnight.

2.3.6.2 Desalting and second GST affinity chromatography

The GST tag cleaved protein sample was loaded onto a HiPrep Desalting column (GE Amersham Biosciences) that was pre-equilibrated with MonoQ

Buffer A (100 mM NaCl, 20 mM Tris-HCl pH 7.5 and 2 mM DTT). After desalting, the sample was reloaded into a Glutathione Sepharose 4B column (GE Amersham Biosciences) to remove the GST tag. The flowthrough was subsequently subjected to ion exchange chromatography.

2.3.6.3 Ion exchange chromatography and gel filtration chromatography

The remaining protein purification steps of hDcp2 (1-245) were identical to the purification steps of PNRC2 Δ NR-hDcp1aEVH1 protein complex.

2.4 Protein analysis

2.4.1 Protein concentration analysis

Protein concentrations were determined using both the Bradford assay (Coomassie Protein Assay kit, Pierce) using bovine serum albumin (BSA) as standard, and the UV absorption method at 280 nm (ND-1000 Spectrophotometer, Nanodrop).

2.4.2 Mass spectrometry MALDI-TOF

The purified proteins were subjected to SDS-PAGE and the gel was stained with Coomassie blue. The desired bands were excised from SDS gel and submitted for protein identification by MALDI-TOF (Mass Spectrometry Facility, Nanyang Technological University).

2.4.3 Edman degradation N-terminal sequencing

Protein samples were run on SDS-PAGE gel and transferred to PVDF membrane (Biorad) (25 V, 1 hour). Subsequently, the membrane was stained with Coomassie blue and the desired protein bands were excised and submitted for N-terminal sequencing by Edman degradation (10 cycles) (Protein and Proteomics Centre, Department of Biological Sciences, National University of Singapore).

2.5 Protein crystallisation

Crystallisation screens were performed using the sitting drop method on 96-well plates (Greiner) at 15°C, using the Screenmaker robot (Innovadyne). The crystallisation buffer kits used were Nextal (QIAGEN) and Hampton Research crystallisation screen kits. Usually, each sitting drop has 230 nl protein solution mixed with 230 nl of reservoir solution. The plates were visually inspected for initial crystals on the third day and one week after the crystallisation screens. The crystals of PNRC2 Δ NR-hDcp1aEVH1 complex were grown using initial crystallisation condition at 15°C by hanging drop vapour diffusion method. The drop contained 1 μ l protein with a concentration of 25 mg/ml and 1 μ l of crystallisation reagent, and the well contains 300 μ l of reservoir solution. The crystallisation condition was further optimised and the final crystallisation reagents consisted of 27 % (w/v) PEG 3350, 0.15 M ammonium acetate, 0.1 M bis tris pH 5.5 and 0.01 M hexamine cobalt (III) chloride. The optimised buffer conditions yield diffraction quality crystals. The crystals of SeMet PNRC2 Δ NR-hDcp1aEVH1 protein complex were grown under the same conditions as the native ones. However, the dimension and quality of single crystals were further improved by microseeding technique. The crystals were harvested and transferred to the mother liquor containing cryoprotectant 15 % glycerol and flash-frozen in liquid nitrogen.

2.6 Structure determination

2.6.1 Data collection

Diffraction data were collected using the flash-frozen Se-Met crystals of the PNRC2 Δ NR-hDcp1aEVH1 complex. Empirical fluorescence scan was used to determine the appropriate x-ray energy for peak absorption by the selenium heavy atom present in the crystals. Subsequently, the x-ray beam of peak wavelength will be used to give maximal f'' value which leads to large Bijvoet differences. After wavelength determination, images separated by 90 degrees were processed to determine the strategy for data collection. All data were collected with an oscillation angle of 0.65 degrees. The diffraction data of SeMet PNRC2 Δ NR-hDcp1aEVH1 protein crystal were collected at the peak

wavelength of selenium absorption edge ($\lambda=0.9798$ Å) at beamline ID23-1, European Synchrotron Radiation Facility (ESRF), Grenoble, France. The best diffraction volumes of the crystals were defined using diffraction cartography (Bowler et al., 2010). The single wavelength anomalous dispersion (SAD) data were processed using the CCP4 suite (CCP4, 1994).

2.6.2 Structure determination of the *PNRC2 Δ NR-hDcp1aEVH1* complex

The structure of the PNRC2 Δ NR-hDcp1aEVH1 protein complex is solved by SAD phasing and the flow chart of structure determination is shown (**Figure 19**). Selenium sites were located by the software, Shake and Bake (SnB) (Miller, 1994) from a random trial structure with 1000 trials. Further refinement of the Selenium sites and phase estimation was accomplished through SHARP (De la Fortelle E, 1997). After solvent flattening, a partial model built from RESOLVE was used for manual model building with program COOT (Emsley and Cowtan, 2004). The model was then refined with Crystallography & NMR system (CNS) (Brunger et al., 1998), REFMAC (Murshudov et al., 1997) and Phenix (Adams et al., 2010) with the use of TLS (Painter and Merritt, 2006a, b). Water molecules were added using Arp/Warp solvent (Perrakis et al., 2001) and the model was further refined by Phenix.

Figure 19. Flow chart of structure determination of the protein complex

2.7 Functional analysis

2.7.1 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) measurements were performed at 20°C in either a VP-ITC microcalorimeter or an iTC₂₀₀ microcalorimeter (MicroCal Inc.). Protein and peptide samples were dialysed into a buffer containing 100 mM NaCl, and 20 mM Tris-HCl pH 7.5. For measuring the binding of PNRC2 (FL)-hDcp1aEVH1 to the Upf1 phosphopeptide (L(pS)QPEL(pS)QDSYLGD) (Peptide2.0), 150 µM of the phosphopeptide was injected into the calorimetric cell containing 15 µM protein complex. The titration was initiated with one 1 µl injection, followed by nineteen 2 µl injections with 60 seconds of equilibration time in between the injections in the iTC₂₀₀ microcalorimeter. The binding of the PNRC2ΔNR-hDcp1aEVH1 complex to the same phosphopeptide was measured by injecting 400 µM phosphopeptide into the calorimetric cell containing 20 µM of the protein complex. The titrations began with one 2µl injection, followed by twenty-eight 10 µl injections, with 240 seconds of equilibration in between the injections in the VP-ITC microcalorimeter. The data were collected and analysed by the Origin 7.0 software program (OriginLab), and fitted using a single-site binding model to obtain the dissociation constant (K_d) and stoichiometry (N) of the binding.

2.7.2 His Tag affinity pull down assay

His-PNRC2ΔNR-Dcp1aEVH1, His-Dcp1aEVH1-PNRC2ΔNR and its eleven mutants (**Table 1**) were expressed as described above. Clarified cell lysate of each protein samples was incubated at 4°C for 2h. After incubation for 2 hours with 10µl of TALON metal affinity resin (Clontech) at 4°C, protein was eluted with 60ul of 250mM imidazole dissolved in lysis buffer. 5ul of the reaction mixtures for F94A, L96A, I104A, L107A, K109A, P110A and WT while 50ul of the reaction mixtures for Y36A, F38A, W45A and W114A were used for running SDS-PAGE. This was because the concentration of soluble protein in Y36A, F38A, W45A, and W114A was much lower as compared to other samples.

2.7.3 Decapping assay

2.7.3.1 *In vitro* transcription

The Ampliscribe T7 transcription kit (Epicentre Technologies) was used to transcribe uncapped RNAs from the control linear λ DNA present in the kit. The steps were performed according to the manufacturer's instructions. The resulting RNAs were precipitated with 2.5 M ammonium citrate and centrifuged at 4°C for 15 minutes. The RNA pellet was washed with 70 % chilled ethanol and air-dried. Subsequently, RNA pellet was solubilised in 50 μ l of nuclease-free water and the concentration of RNA was determined through UV260/280 nm absorption (ND-1000 Spectrophotometer, Nanodrop). The RNA solution were then aliquot and stored at -20°C.

2.7.3.2 RNA capping

The incorporation of [α -³²P] GTP to the mRNA cap was done using ScriptCap m7 capping system kit (Epicentre Technologies) according to manufacturer's instructions. The resulting RNA was eluted in 100 μ l nuclease free water with >3,600 counts per minute (cpm) per μ l.

2.7.3.3 Decapping assay

Decapping assays were carried out in a buffer containing 50 mM Tris-HCl pH 7.9, 30 mM ammonium nitrate and 0.2 mM MnCl_2 . With exception of the negative control, 10 pmol of hDcp2 was used for each reaction. In assessing the stimulation of decapping, 5, 10 and 20 pmol of Dcp1aEVH1 or PNRC2 Δ NR (W114A) were added to the reaction buffer while the Dcp1aEVH1-PNRC2 Δ NR complex was added in 2, 4 and 6 pmol amounts. After the proteins were mixed with the reaction buffer, the reaction mixes were incubated on ice for 1 hour before 80 fmol of the capped RNA was added. The decapping reaction was carried out at room temperature before being terminated by the addition of 2 μ l of 0.5 M EDTA.

2.7.3.4 Thin layer chromatography

Aliquots (5 μ l) of the reaction products from decapping assays was spotted on 20 cm x 20 cm polyethyleneimine-cellulose TLC plates (Sigma) and developed in 0.75 M LiCl. The quantification of cap-labeled RNA was performed using a Phosphorimager (FLA 7000, Fuji).

2.7.4 His tag affinity pull down assay

Recombinant Dcp1aEVH1, PNRC2 Δ NR (W114A) and PNRC2 Δ NR-Dcp1aEVH1 protein complexes were each incubated with 3 μ l of TALON metal affinity resin (Clontech), respectively, at 4°C for 30 min. Subsequently, purified GST hDcp2 proteins were incubated with the beads above which were immobilised with the His tagged proteins at 4°C for 1 h. After washing with RIPA buffer, the bound protein complexes were eluted by boiling in Laemmli sample buffer containing 5 mM EDTA.

2.7.5 Western blot

For western blotting analysis, the eluted proteins were separated by 4 -12 % SDS-PAGE (Novex) and electrophoretically transferred to polyvinylidene fluoride (PVDF) membrane (Millipore). Non-specific sites were blocked with 5 % nonfat milk in TBST (20 mM Tris pH 7.6, 137 mM NaCl, 0.05 % Tween 20) at room temperature for 40 min. An anti-GST antibody (GWB-F35816; GenWay Biotech) was incubated with membrane in 2 % non-fat milk for 1 hour at room temperature at the dilution indicated in the manufacturer's protocol. After incubation with the peroxidase-coupled secondary antibody, proteins were detected using enhanced chemiluminescence (ECL) system (Thermo Fisher Scientific).

3 RESULTS

3.1 Purification of proteins

After purifying all protein samples using protein affinity and ion exchange chromatography, these samples are subjected to gel filtration chromatography for a final round of purification. Pure protein samples from peak fractions was verified by SDS-PAGE before they were pooled and concentrated to a final concentration of 20 - 25 mg/ml for storage at -80°C (Figure 20 - 24).

Figure 20. Purification of recombinant PNRC2 Δ NR-hDcp1aEVH1 protein complex

(a) The 26kDa PNRC2 Δ NR-hDcp1aEVH1 complex is eluted by size exclusion gel filtration chromatography column S200. The overall size of the complex suggests that there is only 1 copy of the PNRC2 and hDcp1aEVH1 protein. (b) Protein purity was visualised by SDS-PAGE

Figure 21. Purification of recombinant PNRC2 (FL)-hDcp1aEVH1 protein complex
 (a) The 28kDa PNRC2 (FL)-hDcp1aEVH1 complex is eluted by size exclusion gel filtration chromatography column S200. (b) Protein purity was visualised by SDS-PAGE.

Figure 22. Purification of recombinant His-PNRC2 Δ NR (W114A)
 (a) As PNRC2 Δ NR (W114A) is defective in binding Dcp1aEVH1, and hence it eluted as a single protein band of 13kDa by size exclusion gel filtration chromatography column S200. (b) Protein purity was visualised by SDS-PAGE

Figure 23. Purification of recombinant his-hDcp1aEVH1 protein

(a) hDcp1aEVH1 is eluted as a single protein band of 13 kDa by size exclusion gel filtration chromatography column S200. (b) Protein purity was visualised by SDS-PAGE.

Figure 24. Purification of recombinant human Dcp2 (1-245) protein

(a) hDcp2(1-245) is eluted as a monomer of 29kDa by size exclusion gel filtration chromatography column S200. (b) Protein purity was visualised by SDS-PAGE.

3.2 Protein Analysis

3.2.1 Mass spectrometry MALDI-TOF

Protein identity of the PNRC2 and the hDcp1aEVH1 proteins are verified by MALDI-TOF mass spectrometer (Figure 25). The peptide mass fingerprints derived from the methodology match the calculated absolute peptide masses from the PNRC2 and the hDcp1aEVH1 proteins.

A

Gel Idx/Pos	104/E3	Instr./Gel Origin	AK072/091009_ID_MW	Process Status	Analysis Succeeded					
Plate [#] Name	[1] 001300018198	Instrument Sample Name		Spectra	11					
Protein Name		Rank	Species	Accession No.	Protein MW	Protein Score	Protein Score C. I. %	Total Ion Score	Total Ion C. I. %	Confirmed
Proline-rich nuclear receptor coactivator 2 - Homo sapiens (Human)	1			PNRC2_HUMAN	15580.9	300	100	224	100	
Peptide Information										
Calc. Mass	Obsrv. Mass	± da	± ppm	Start Seq.	End Sequence Seq.	Ion Score	C. I. %	Modification		Rank Result Type
864.3879	864.4271	0.0392	45	31	37 EQNSQMK					Mascot
900.4646	900.5115	0.0469	52	20	26 NQQQLNR	49	98.052			Mascot
900.4646	900.5115	0.0469	52	20	26 NQQQLNR					Mascot
1045.5425	1045.5992	0.0567	54	7	15 YNIPAPQSR					Mascot
1093.5306	1093.5708	0.0402	37	29	37 TKEQNSQMK					Mascot
1151.5439	1151.5897	0.0458	40	87	97 SQANQNYAGAK	51	98.71			Mascot
1151.5439	1151.5897	0.0458	40	87	97 SQANQNYAGAK					Mascot
1328.703	1328.7461	0.0431	32	16	26 NVSKNQQLNR					Mascot
1385.7607	1385.7723	0.0116	8	20	30 NQQQLNRQKTK					Mascot
1804.8362	1804.9022	0.066	37	67	82 NFPNNQSWNSLSGPR	124	100			Mascot
1804.8362	1804.9022	0.066	37	67	82 NFPNNQSWNSLSGPR					Mascot
1820.8311	1820.8956	0.0645	35	67	82 NFPNNQSWNSLSGPR	25		0 Oxidation (HW)[8]		Mascot
1820.8311	1820.8956	0.0645	35	67	82 NFPNNQSWNSLSGPR			Oxidation (HW)[8]		Mascot
1934.8562	1934.9198	0.0636	33	46	64 GHGYNSSAAAWQAMQN GSK					Mascot

B

Gel Idx/Pos	103/E2	Instr./Gel Origin	AK072/091009_ID_MW	Process Status	Analysis Succeeded					
Plate [#] Name	[1] 001300018198	Instrument Sample Name		Spectra	11					
Protein Name		Rank	Species	Accession No.	Protein MW	Protein Score	Protein Score C. I. %	Total Ion Score	Total Ion C. I. %	Confirmed
mRNA decapping enzyme 1A (EC 3.-.-.) (Transcription factor SMIF) (SmaD4-interacting transcriptiona	1			DCP1A_HUMAN	63410.3	435	100	389	100	
Peptide Information										
Calc. Mass	Obsrv. Mass	± da	± ppm	Start Seq.	End Sequence Seq.	Ion Score	C. I. %	Modification		Rank Result Type
1118.5874	1118.6254	0.038	34	7	17 AGQEMSLAALK	46	95.984			Mascot
1118.5874	1118.6254	0.038	34	7	17 AGQEMSLAALK					Mascot
1313.6736	1313.7358	0.0622	47	48	58 TDIEGTLFVYR	99	100			Mascot
1313.6736	1313.7358	0.0622	47	48	58 TDIEGTLFVYR					Mascot
1407.7413	1407.7881	0.0468	33	73	84 LNMHNLVEPVNK	63	99.919			Mascot
1407.7413	1407.7881	0.0468	33	73	84 LNMHNLVEPVNK					Mascot
1423.7362	1423.7806	0.0444	31	73	84 LNMHNLVEPVNK			Oxidation (HW)[4]		Mascot
1448.728	1448.7948	0.0668	46	60	72 SASPYHGFTIVNR	95	100			Mascot
1448.728	1448.7948	0.0668	46	60	72 SASPYHGFTIVNR					Mascot
1464.7297	1464.7732	0.0435	30	338	350 NSTMMQAVKTTPR					Mascot
1469.7747	1469.8137	0.039	27	48	59 TDIEGTLFVYRR					Mascot
1604.8292	1604.8799	0.0507	32	59	72 RSASPYHGFTIVNR					Mascot
1706.8424	1706.8911	0.0487	29	99	112 NASLSIYINIFYDK					Mascot
1819.9377	1819.9982	0.0605	33	85	98 DLEQLHEPFLLYR					Mascot

Figure 25. Peptide mass fingerprints of the PNRC2 and hDcp1aEVH1 proteins
The proteins are cleaved into smaller peptides and the masses of these peptides are measured by MALDI-TOF mass spectrometer. The measured peptide mass fingerprints match the calculated peptide mass fingerprints of the (A) PNRC2 and (B) hDcp1aEVH1 proteins.

3.2.2 Edman degradation N-terminal sequencing

Edman degradation N-terminal sequencing was performed on an earlier construct of hDcp1aEVH1 protein (1- 140 aa) as this construct is unstable during purification. Results of this sequencing reveals that the N terminal of the hDcp1aEVH1 protein is stable and hence suggesting that degradation occur from the C terminal (**Figure 26**).

B

Cycle number	Residue No	Main Residue/s
1	BLANK STD	
2		
3	1	A
4	2	D
5	3	L
6	4	M
7	5	E
8	6	A
9	7	L
10	8	S
11	9	R
12	10	A

Figure 26. Edman degradation N terminal sequencing

(A) Results from 12 cycles of Edman degradation N terminal sequencing. Cycle 1 and cycle 2 serve as negative and positive control respectively. From cycle 3 to cycle 12, the identity of the amino acid is derived from the highest peak that corresponds to an equivalent amino acid peak in cycle 2. (B) The first three amino acids are not part of the hDcp1aEVH1 protein. They are added as a result of the ligation of the hDcp1aEVH1 gene into the open reading frame of the 2nd MCS of pETDuet-1 vector. The results show that the N terminal of hDcp1aEVH1 remains intact after expression.

3.3 Crystal structure of PNRC2 Δ NR- Dcp1aEVH1

3.3.1 Structure determination

Due to poor solubility of the full length hDcp1a protein, sequence alignments, secondary structure prediction and homology modelling using the *S. cerevisiae* Dcp1 protein (ScDcp1) protein structure as a template were carried out to determine a stable construct of the hDcp1a protein. Thus, a compact fragment of the hDcp1a protein containing the EVH1 domain which is of high solubility and purity was identified (Residue 1 to 130). Similarly, the full length PNRC2 protein is unstable which results in degradation of the protein after three days at 4°C. Subsequently, secondary structure prediction was employed to predict a relatively stable fragment of the PNRC2 protein (PNRC2 Δ NR) (Residue 1 to 121). The domain organisation of the hDcp1a and PNRC2 proteins was shown. (**Figure 27**)

Purified preparations of the hDcp1aEVH1 protein in complex with PNRC2 Δ NR were screened for crystallisation conditions. After optimisation of the initial crystallisation condition, the final crystallisation condition is fixed at 27 % (w/v) PEG 3350, 0.15 M ammonium acetate, 0.1 M bis tris pH 5.5 and 0.01 M hexamine cobalt (III) chloride. The crystals of SeMet PNRC2 Δ NR-hDcp1aEVH1 protein complex were grown under the same conditions as the native ones and it yielded diffraction quality crystals with diameter of 50 microns (**Figure 28**). The

Figure 27. Domain organisation of hDcp1a and PNRC2 proteins

Schematic diagrams of domain organisation of the PNRC2 and hDcp1a proteins. SH3-binding motif, PRS region and the NR box of PNRC2 are colored in red, pink and cyan respectively while the EVH1 and the trimerisation domains of the hDcp1a protein are shown in pale green and yellow.

Figure 28. Crystals of PNRC2ΔNR- Dcp1aEVH1

protein complex was crystallised in space group $I222$. Single-wavelength anomalous dispersion method was used to solve the structure at a resolution of 2.6 Å, using data obtained from SeMet-substituted crystal.

Additionally, six out of seven Selenium sites were found by the software, SnB (Miller, 1994). The coordinates of the Selenium sites are shown in **Table 2**. SnB uses the direct methods phasing algorithm which is based on the minimal principle of structure invariants which are linear combinations of phase angles. The minimal principle states that the minimal function $R(\phi)$, which determines the mean-square difference between estimated and calculated structure invariants, has a constrained global minimum when all the phases are equal to their true values for some choice of origin. The software uses a trial structure and uses the inverse fourier transform to calculate the corresponding phases in the reciprocal space. Subsequently, the associated value of $R(\phi)$ is calculated and phase refinement is carried out via parameter shift to minimise the $R(\phi)$. The refined phases are then subjected to fourier transform to produce an electron density map in the real space. Density modification via peak selection is carried out whereby n largest peaks are selected as atom position for a n -atom structure. These peaks are subjected to inverse fourier transformation to obtain the phases. Phase refinement and density modification is repeated for a predetermined number of cycles to derive the global minimum of $R(\phi)$ before a new trial is initiated. Evaluation of 1000 trials is shown (**Figure 29**). After further refinement as detailed in Section 2.6.2, the final model has good stereochemistry with a free R factor of 28.5 % and an R factor of 23.4 %. Data collection and refinement statistics are shown in **Table 3**.

Residues 1 - 130 of the hDcp1aEVH1 protein and a stretch of PNRC2 polypeptide chain comprising fourteen residues (residues 102 to 115) are visible in the electron density map with the rest of the residues (1-101 and 116-121 aa) assumed to be disordered as there is no interpretable electron density for these regions. The final model contains two complexes per asymmetric unit (AU) with two copies of the hDcp1aEVH1 (chains A and B) complexed with PNRC2 peptide (residue 102-115; chains C and D) with no major differences (root mean square deviation (rmsd) of 0.30 Å for all the equivalent C α atoms) (**Figure 30A**). As chains A and C are more complete than chains B and D, the subsequent analysis reported here is based on chains A and C only. A stereo view of PNRC2 Δ NR- hDcp1aEVH1 is shown (**Figure 30B**).

Table 2. Selenium atom sites found by the software SnB

	X	Y	Z	Height
1	0.652383	0.800469	0.857997	16.42
2	0.542104	1.003068	0.956079	14.02
3	0.750879	0.603071	0.764763	12.06
4	0.543925	0.915580	0.955449	11.23
5	0.951965	0.530608	0.667888	9.25
6	0.916071	0.553073	0.698968	8.06

Table 3. Data Collection and refinement statistics of SeMet PNRC2ΔNR-hDcp1aEVH1

Data collection	SeMet PNRC2ΔNR-hDcp1aEVH1
Wavelength (Å)	0.9793
Resolution limit (Å)	2.60
Space Group	I222
Unit cell dimensions	
a, b, c (Å)	77.195, 98.824, 107.011
α, β, γ (°)	90, 90, 90
Reflections	
Measured	42138
Unique	12401
Completeness (%)	98.4 (91.0)
Redundancy	3.4
I/σ	19.1 (4.9)
R _{merge} ^a	0.037(0.166)
Phasing statistics	
Number of Se sites	6
R _{cullis} ^b	0.69
Figure of Merit	
Before density modification	0.30
After density modification	0.82
Refinement	
Resolution Range (Å)	20-2.6
Number of atom	
Protein	2082
Water	67
R _{work} ^c (%)	23.35
R _{free} ^d (%)	28.47
R.M.S deviation	
Bond lengths (Å)	0.008
Bond angles (°)	1.24
Ramachandran plot	
Allowed (% residues)	99.1
Generously allowed (% residues)	0.9
Disallowed (% residues)	0

Values in parentheses indicate the values in the highest resolution shell.

^a $R_{\text{merge}} = \sum |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^b $R_{\text{cullis}} = \sum_h ||F_{\text{PH}} - F_{\text{H}}| - |F_{\text{H}}|| / \sum_h |F_{\text{PH}} - F_{\text{H}}|$.

^c $R_{\text{work}} = \sum ||F_{\text{o}} - F_{\text{c}}| / \sum |F_{\text{c}}|$, where F_{o} denotes the observed structure factor amplitude, and F_{c} denotes the structure factor amplitude calculated from the model

^d R_{free} is similar to R_{work} but calculated with 10.0 % of randomly chosen reflections which are omitted from the refinement.

A

B

Figure 29. Evaluation of 1000 trials by SnB software

(A) Histogram of Rmin function. The 156th trial has the lowest Rmin value of 2.70. R(true) and R(random) shows the expected values of the minimal function for sets of true and random phases respectively. Furthermore, the 40 trials structures that form a distinct cluster around Rmin value of 0.30 show that the correct phases for the SAD data has been found. (B) The Rmin trace. It shows the changes of Rmin value after each cycle of phase refinement for the best trial.

3.3.2 Overall structure description

The structure of the EVH1 domain of hDcp1a consists of seven anti-parallel β strands and two α helices. β sheets 1-4 form a signature β -sandwich which is capped by a N-terminal α -helix (**Figure 31A**). Consequently, there are two concave grooves formed on the solvent-exposed side of the β sheets. Following the naming convention of the ScDcp1 structure (**Figure 31B**), the first groove (designated as Patch 1) of the hDcp1a protein is formed by β 1, β 2, β 5, β 6, β 7 strands while the second groove (Patch 2), which is found opposite to the PRS site, is formed by β 1, β 4, β 5, β 6 and the α 1 helix. The PNRC2 peptide adopts a largely extended conformation and contacts the hDcp1a protein, which includes the canonical PRS binding site observed in other EVH1 domains in a 1:1 binding stoichiometry.

Figure 30. Structure of PNRC2 Δ NR- Dcp1aEVH1

(A) Crystals of PNRC2 Δ NR- Dcp1aEVH1 complex. (B) Two PNRC2 Δ NR- Dcp1aEVH1 exist in the asymmetric unit: Molecule A (hDcp1a; pale green), molecule B (hDcp1a; pale blue), molecule C (PNRC2; pale pink) and molecule D (PNRC2; yellow) (C) Stereo view of the Dcp1aEVH1-PNRC2 Δ NR interface. Residues 102 to 115 of the PNRC2 peptide and the residues of the Dcp1aEVH protein1 involved in the interface are labelled and shown as stick models. The colouring scheme is as Figure 27.

3.3.3 Structural comparison with previous classes of EVH1 domain

Following a search in the Protein Data Bank with the Dali server (Holm and Sander, 1993), the closest structural homolog ScDcp1 (She et al., 2004) (PDB entry 1Q67; Z score 16.0) was found. The N terminus of ScDcp1 protein shares 17% sequence identity with the EVH1 domain of hDcp1a protein. Additionally, superpositioning of equivalent C α atoms of the hDcp1a and ScDcp1 proteins gives rise to rmsd of 1.28Å. Overall, the core regions of the structure of hDcp1a and Dcp1 proteins are very similar (**Figure 31A and 31B**). Nevertheless, two prominent differences exist between these two structures. First, two short helices located at one of the two *S. cerevisiae*-specific insertions (She et al., 2004), are present in the ScDcp1 protein but absent in the hDcp1aEVH1. Second, the N-terminal region of ScDcp1 forms a 3_{10} -helix with an extended polypeptide whereas in Dcp1aEVH1 this region is a long α -helix, which bears similarity to the long N-terminal α -helix observed in the structure of *S. pombe* Dcp1 bound to Dcp2 (She et al., 2008).

Subsequently, three more structural homologs, Mena (Prehoda et al., 1999) (PDB entry 1EVH ; Z-score 14.5), Homer (Beneken et al., 2000) (PDB entry 1DDV; Z-score 13.2) and N-Wasp proteins (Volkman et al., 2002) (PDB entry 1MKE; Z-score 11.0) were also identified (**Figure 31**). The superposition of the equivalent C α atoms of the hDcp1aEVH1 protein with the three homologs proteins revealed that there are no crucial differences in their topologies with rmsd of 1.58Å, 1.61Å and 1.73Å for the Mena, Homer and N-Wasp proteins respectively. More importantly, the alignment also revealed that the position of PRS sites of the EVH1 domains is conserved at equivalent location.

Figure 31 Comparison of the hDcp1aEVH1 protein with EVH1 domains from the ScDcp1, Mena, Homer and N-Wasp proteins. Residues of PRS binding site that interacted with PRS ligand were highlighted in yellow.

(A) Structure of the hDcp1aEVH1 protein (pale green) in complex with the PNRC2 peptide (pink) with secondary elements labelled and the N-terminal helix colored in blue. (B) Structure of the ScDcp1 protein (pale green) with the N-terminal region and the insertion helices shown in blue and orange, respectively. (C) Structure of the Mena EVH1 domain (olive) in complex with the FPPPP peptide from ActA (magenta). (D) Structure of the Homer EVH1 domain (olive) in complex with TPPSPF peptide from mGluR (magenta). (E) NMR structure of the N-Wasp EVH1 (olive) in complex with WIP peptide (magenta).

3.3.4 The hDcp1aEVH1-PNRC2 interface

Most of the residues involved in the hDcp1aEVH1-PNRC2 interface are highly conserved (**Figure 32**), therefore highlighting their important roles in the interaction between the hDcp1a and PNRC2 proteins (**Table 4**). The interactions of the PNRC2 peptide with the Dcp1aEVH1 protein are predominantly hydrophobic in nature as shown by the electrostatic map of the binding interface (**Figure 33**). In addition, it is also supplemented by additional VDW contacts and hydrogen bonding (**Table 4**). The binding interface between the hDcp1aEVH1 protein and the PNRC2 peptide consists of two hydrophobic pockets with a central ridge made up by Trp45 and Leu96 (**Figure 34A**). The first pocket (designated as hydrophobic pocket 1), which corresponds to patch 1 of the ScDcp1 protein (She et al., 2004), comprises the side-chains of conserved residues Tyr36, Trp45, Phe94, Leu96 and Trp108 while the second pocket (designated as hydrophobic pocket 2) is made up of the side-chains of Phe38, Trp45, Leu96, Ile104 and the methylene groups of Glu87, Gln89 and Arg98. The N- and C-terminal portions of the PNRC2 peptide associate with hydrophobic pockets 1 and 2 respectively while its middle portion containing the sequence LPKP docks at the ridge formed by Trp45 and Leu96 through multiple Van der Waals (VDW) contacts. The LPKP motif adopts a polyproline II (PPII) conformation (a left-handed helix with three residues per turn) which is conserved in the proline rich motif of PRS ligands in other EVH1-peptide complexes ((Ball et al., 2002); see below). Importantly, the hydrophobic pocket 1 of hDcp1a protein that interacts with the PNRC2 peptide is analogous to the region of *S. cerevisiae* Dcp1 shown to be important for decapping and for binding to the decapping activator, Edc1 (Borja et al., 2011; She et al., 2004). This suggests that hydrophobic pocket 1 harbour conserved interaction sites for a variety of decapping activators in numerous species (see below).

Starting from the N-terminal portion of the PNRC2 peptide, Pro102 is involved in stacking interaction with the side chain of Trp108 of the hDcp1aEVH1 protein while Ser103 makes multiple VDW contacts with Tyr36, Phe94 and Trp108 of the hDcp1aEVH1 protein. Additionally, hydrogen bond formed between Ser103 carbonyl oxygen and the hydroxyl group of Tyr36 in the hDcp1aEVH1 protein. In the proline rich motif, LPKP, residing in the middle portion of the PNRC2 peptide, Leu107 and Pro108 protrudes into the

hydrophobic pocket 1 and make multiple hydrophobic interactions and two VDW contacts with residues Tyr36, Trp45, Phe94 and Ser106 of the hDcp1aEVH1 protein. In addition, the carbonyl oxygen of Pro108 is hydrogen-bonded to the indole nitrogen of Trp45 in the hDcp1aEVH1 protein. P110 in the LPKP motif contacts residues Phe38, Trp45, Leu96 and Ile104 in hydrophobic pocket 2 through multiple hydrophobic interactions while Lys109 is exposed to the solvent region and forms a hydrogen bond with the carbonyl oxygen of Asn43. In the C-terminal portion of the PNRC2 peptide, Trp114 is the key residue that interacts with the hDcp1aEVH1 protein. This residue inserts its side chains deeply in hydrophobic pocket 2 and makes hydrophobic interactions with the side-chain of Ile104 as well as the methylene groups of Glu87 and Arg98. In addition, the indole nitrogen of Trp114 is hydrogen-bonded to the OE1 atom of Glu87.

Figure 32. Sequence alignment analysis

(A) Sequence alignment of Dcp1 EVH1 domain from *H.sapiens*, *S.cerevisiae*, *S.pombe*, *M. musculus*, *D.melanogaster* and *C. elegans* respectively. Residues of the hDcp1a protein that interacts with the PNRC2 peptide are indicated with a star above the sequence. (B) Sequence alignment of the PNRC2 protein from *H.sapiens*, *B.taurus*, *M. musculus*, *R.norvegicus* and *X. laevis* respectively. Residues of the PNRC2 peptide that interact with the hDcp1a protein are indicated with a star above the sequence and more importantly, they remain invariant across species.

Table 4. Binding network of PNRC2/hDcp1aEVH1 interface

PNRC2 Residues	hDcp1a Residues	Type of interaction	Distance between interacting atoms (Å)
Pro102 (C)	Tyr36 (OH)	VDW	3.62
Ser103(OG)	Tyr36(OH)	Hydrogen bond	3.41
Ser103(O)	Tyr36(OH)	Hydrogen bond	2.60
Ser103(OG)	Phe94(CZ)	VDW	3.47
Ser103(OG)	Trp108(CD1)	VDW	3.33
Ser105(CA)	Tyr36(CE1)	Hydrophobic	3.78
Ser105(CB)	Trp45(O)	VDW	3.42
Ser105(OG)	Lys47(N)	Hydrogen bond	3.66
Leu107(CB)	Trp45(CD1)	Hydrophobic	3.62
Leu107(O)	Trp45(CD1)	VDW	3.59
Leu107(CD1)	Tyr36(CG)	Hydrophobic	3.30
Leu107(CD1)	Ser106(CB)	Hydrophobic	3.63
Leu107(CD2)	Phe94(CE1)	Hydrophobic	3.39
Pro108(CD)	Phe94(CD2)	Hydrophobic	3.81
Pro108(O)	Trp45(CE2)	VDW	3.22
Pro108(O)	Trp45(NE1)	Hydrogen bond	2.39
Lys109(NZ)	Asn43(O)	Hydrogen bond	3.19

PNRC2 Residues (atom label)	hDcp1a Residues (atom label)	Type of interaction	Distance between interacting atoms (Å)
Pro110(CA)	Leu96(CD2)	Hydrophobic	3.57
Pro110(CB)	Ile104(CG2)	Hydrophobic	3.88
Pro110(CD)	Phe38(CE1)	Hydrophobic	3.19
Pro110 (CD)	Trp45 (CZ2)	Hydrophobic	3.74
Pro111(CD)	Gln89(OE1)	VDW	3.57
His113(CE1)	Glu87(OE1)	VDW	3.80
Trp114(CD1)	Arg98(CZ)	Hydrophobic	3.28
Trp114(CD1)	Arg98(NH2)	VDW	3.21
Trp114(CB)	Ile104(CD1)	Hydrophobic	3.66
Trp114(NE1)	Glu87(OE1)	Hydrogen bond	2.53
Trp114(CE2)	Glu87(OE1)	VDW	3.34
Trp114(CZ2)	Glu87(CG)	Hydrophobic	3.37
Trp114(CZ3)	Tyr97(C)	Hydrophobic	3.62

Figure 33. Electrostatic charge distribution hDcp1aEVH1 binding interface with PNRC2

The electrostatic charge distribution (blue positive to red negative) of hDcp1aEVH1 binding interface with the PNRC2 protein represented as stick model (pink). The interface is mostly uncharged and hence it shows that the binding of the hDcp1aEVH1 protein with the PNRC2 protein is predominantly hydrophobic interactions.

3.3.5 A novel recognition mechanism of PNRC2 by hDcp1aEVH1

The binding of proline rich motif of PNRC2, LPKP, to the hDcp1aEVH1 protein retains characteristics that are common to all classes of previously characterised EVH1 domains in complex with their respective PRS peptides (**Figure 34**). First, the motif adopts a PPII helical conformation. Second, it recognises and binds to the same canonical aromatic surface utilised by all classes of EVH1 domain to recognise their peptide ligands.

Nevertheless, the residues of the PNRC2 proline rich motif that interact with the Dcp1aEVH1 protein provide several lines of evidence, which suggest that the PNRC2 peptide may bind to the Dcp1aEVH protein via a novel recognition mechanism. First, as shown in detailed surface representation of the EVH1 interfaces (**Figure 34**), a set of four aromatic residues (Tyr36, Phe38, Trp45 and Phe94) in the hDcp1aEVH1 protein is involved in the recognition of the PNRC2 PRS peptide. In contrast, a different set of three aromatic residues in other EVH1 domains (Tyr16, Trp23 and Phe77 in Mena; Phe14, Trp24 and Phe74 in Homer and Tyr46, Trp54 and Phe104 in N-Wasp) is responsible for recognising the PRS ligands. As a result, the PNRC2 peptide does not share common motifs with the PRS ligands in other EVH1 domains. Second, the length

of PNRC2 peptide (14 residues) is more than twice as long as the minimal PRS peptide (6 residues) required for Mena or Homer EVH1 recognition. Hence, in the structure of the PNRC2 Δ NR-hDcp1aEVH1 complex, the buried surface ($\sim 1450\text{\AA}^2$) is also more than twice that of the Mena (Prehoda et al., 1999); $\sim 710\text{\AA}^2$) or Homer protein (Beneken et al., 2000); $\sim 560\text{\AA}^2$) in complex with its respective PRS peptide. However, the PNRC2 peptide is shorter than the Wasp Inhibitor Peptide (WIP; 25 residues) that binds to the N-Wasp EVH1 protein (Ball et al., 2002). Thus, buried surface of the PNRC2 Δ NR-hDcp1aEVH1 complex is smaller than that of the N-Wasp EVH1 protein with the bound WIP peptide (Volkman et al., 2002); $\sim 2130\text{\AA}^2$). Third, the PNRC2 peptide adopts a novel conformation with its PPII helical conformation distorted by incorporation of turns at both its N and C terminus to appropriately position Pro102, Ser103 and Trp114 for interactions with the hDcp1aEVH1 protein.

Additionally, several features indicate that the interaction between the hDcp1aEVH1 protein and the PNRC2 peptide encompass elements that were previously thought to be unique to all three classes of EVH1 domain. First, the stacking interaction between the methylene group of Arg98 in the hDcp1a protein and Trp114 in the C-terminal portion of the PNRC2 peptide (**Figure 34A**) is similar to the interaction between Arg81 of Mena and Phe1 of the FPPPP peptide (**Figure 34B**). Accordingly, Trp114 of PNRC2 and Phe1 of FPPPP peptide are shown to be critical residues that set the register of binding and maintain an unambiguous directionality of the PRS ligand (see below). Second, the N-terminal portion of the PNRC2 peptide binds to the edge of the hydrophobic pocket 1 of the hDcp1aEVH1 protein in a conformation similar to that observed for the TPPSPF peptide bound to the Homer EVH1 domain protein (Beneken et al., 2000) (**Figures 34A and 34C**). Third, the PNRC2 peptide binds to the hDcp1aEVH1 protein in the same orientation as the WIP peptide bound to the N-Wasp EVH1 protein but in a reverse orientation as compared to the PRS ligands bound to the Mena and Homer proteins (Volkman et al., 2002) (**Figure 31**). Taken together, these results indicate that the hDcp1aEVH1 protein recognises the PNRC2 peptide via a novel recognition mechanism.

Figure 34. Surface views of the EVH1 interfaces

Residues involved in interaction are labelled and shown as stick models. The surface utilised by the EVH1 domain for recognising the PRS ligand is shown in light grey while the coloring scheme for the rest of the surface is as in Figure 31. **(A)** Surface representation of the hDcp1aEVH1 protein with bound PNRC2 peptide (residues 102-115). The LPKP motif of the PNRC2 peptide adopts the PPII conformation and it docks in the conserved aromatic surface, which overlaps with hydrophobic pocket 1 to 2. **(B)** Surface representation of the Mena EVH1 domain in complex with the FPPPP peptide. **(C)** Surface representation of the Homer EVH1 domain in complex with the TPPSPF peptide. **(D)** Surface representation of the N-Wasp EVH1 domain with the bound WIP peptide.

3.4 Mutational effects on PNRC2-hDcp1a interaction

To delineate the role of the residues lining the binding interface between the hDcp1aEVH1 protein and the PNRC2 peptide, we mutated residues in the hDcp1aEVH1-PNRC2 interface and examined the effects of these mutations on their interactions. For the first set of experiments involving the WT PNRC2 protein and various hDcp1a mutants, the loading amount of the WT PNRC2 protein is fixed in all samples (**Figure 35A**). Subsequently, through visualisation of the hDcp1a protein band in SDS PAGE, the binding affinity of the WT PNRC2 protein with the WT hDcp1a and its various mutants can be compared. Similarly for the second set of experiments involving the WT hDcp1a protein with the PNRC2 mutant proteins, the loading amount of the WT hDcp1a protein is fixed (**Figure 35B**). Hence, *In vitro* His tag affinity pull down assays demonstrate that four mutants Y36A, F38A, W45A of the hDcp1aEVH1 protein and W114A of the PNRC2 protein have significant reduction in the interaction between the hDcp1aEVH1 and PNRC2 proteins (**Figure 35**). Importantly, the reduction in interaction leads to decrease in solubility of the four mutants, Y36A, F38A, W45A of hDcp1aEVH1 protein and W114A of the PNRC2 proteins. This results in significantly higher loading volume (50ul) of the four mutants as compared to that of wild type (5ul) to obtain comparable amount of soluble protein during SDS PAGE.

Figure 35. *In vitro* His tag affinity pull down assay

(A) His tagged fusion WT PNRC2 protein was used to pull down tag free WT and mutants hDcp1a proteins. (B) His tagged fusion WT hDcp1a protein was used to pull down tag free WT and mutants PNRC2 protein.

3.5 The NR box of PNRC2 is required for binding to the phosphorylated Upf1

Previous Yeast two-hybrid and coimmunoprecipitation assays showed that the SH3-binding motif and the NR box in the C-terminus of the PNRC2 protein is sufficient for binding to the Upf1 protein (Cho et al., 2009). Furthermore, the PNRC2 protein preferentially interacts with the hyperphosphorylated Upf1 protein compared with the wild-type Upf1 protein (Cho et al., 2009). Additionally, preferential association between the hyperphosphorylated Upf1 and hDcp1a proteins also has been reported (Isken et al., 2008). However, indirect interaction cannot be ruled out as the binding result was obtained using coimmunoprecipitation. Our structure shows that the PRS region immediately downstream of the SH3-binding motif in the PNRC2 protein directly binds to the hDcp1a protein while Zhou et al., (2006) showed that the SH3-binding motif of the PNRC2 protein is essential for binding to the nuclear receptors. Thus, it is proposed that the direct association of the hyperphosphorylated Upf1 protein with the NR box of PNRC2 protein mediates the binding of the Upf1 to hDcp1a proteins.

To examine if the NR box of PNRC2 protein directly binds to the hyperphosphorylated Upf1 protein, we synthesised a double-phosphorylated Upf1 peptide L(pS)QPEL(pS)QDSYLGD, which is derived from the human Upf1 C-terminal SQ-containing peptide LSQPELSQDSYLGD. This peptide has been shown to be the best substrate phosphorylated by the human SMG1 protein (Yamashita et al., 2001). Isothermal titration calorimetry (ITC) was used to analyse the binding of this double-phosphorylated Upf1 peptide to the protein complexes PNRC2 Δ NR-hDcp1aEVH1 and PNRC2 (FL)-hDcp1aEVH1. As a control, the binding of the nonphosphorylated Upf1 peptide to the PNRC2 (FL)-hDcp1aEVH1 protein complex was also examined. The PNRC2 (FL)-hDcp1aEVH1 protein complex binds to the double-phosphorylated Upf1 peptide with an apparent K_d of 0.21 μ M and a binding stoichiometry of 1 (**Figure 36C**). Nevertheless, the same PNRC2 (FL)-hDcp1aEVH1 protein complex shows no binding to the nonphosphorylated Upf1 peptide (**Figure 36B**) while the PNRC2 Δ NR-hDcp1aEVH1 protein complex shows no detectable binding to the phospho-Upf1 peptide (**Figure 36A**). Taken together, these results suggest that the NR box of PNRC2 protein is required for binding to the phosphorylated Upf1 protein in a 1:1 ratio. However, as the structure of PNRC2 (FL)-hDcp1aEVH1–Upf1 protein complex is not available, further binding assay like Surface Plasmon Resonance (SPR) analysis may be done to verify our results.

A

B

C

Figure 36. Binding isotherms of ITC titrations

ITC titrations of (A) double-phosphorylated Upf1 peptide to the PNRC2 Δ NR-hDcp1aEVH1 protein complex (B) nonphosphorylated Upf1 peptide to the PNRC2 (FL)-hDcp1aEVH1 protein complex (C) double-phosphorylated Upf1 peptide to the PNRC2 (FL)-hDcp1aEVH1 protein complex. The upper panels show the binding isotherms and the lower panels show the integrated heat for each injection fitted to a single-site model. Accordingly, the binding stoichiometry, binding constant, enthalpy and entropy are shown.

3.6 PNRC2 and hDcp1a synergistically stimulate decapping *in vitro*

The C-terminal regions of the Edc1 and Edc2 proteins contain a proline-rich consensus sequence that binds to the ScDcp1 protein. This allows the ScDcp1 protein to recruit the coactivators Edc1 and Edc2 proteins to the Dcp2 protein for decapping stimulation (Borja et al., 2011; Schwartz et al., 2003). Sequence comparison showed that the PNRC2 peptide bound to the Dcp1aEVH1 protein shares the characteristic of the proline-rich consensus sequence of the Edc1 and Edc2 proteins (**Figure 37A**), suggesting that the PNRC2 protein may stimulate decapping through facilitating the complex formation between the hDcp1a and Dcp2 proteins.

Since the ScDcp1 protein only consists of the EVH1 domain and it is still capable of stimulating decapping *in vitro* (She et al., 2006; She et al., 2004; Sheth and Parker, 2006), we reasoned that the EVH1 domain of the hDcp1a protein might be sufficient to stimulate decapping *in vitro*. As expression of the hDcp1aEVH1 or PNRC2 Δ NR proteins alone in *E. coli* gave rise to insoluble protein, we co-expressed the hDcp1aEVH1 and PNRC2 Δ NR (W114A), a mutant showing residual binding to the hDcp1aEVH1 (**Figure 35B**). The residual interaction between the hDcp1a and PNRC2 Δ NR (W114A) proteins improved the solubility of each individual protein and allowed us to purify the hDcp1aEVH1 and PNRC2 Δ NR (W114A) proteins separately by attaching a His-tag to either the hDcp1aEVH1 or PNRC2 Δ NR (W114A) proteins in two separate coexpression constructs, and then removing the untagged protein by extensive washes of the TALON metal affinity resin column.

Decapping assays demonstrated that the hDcp1aEVH1 protein stimulated decapping moderately (**Figures 37B, lanes 5-7 and 37C, lanes 5-7**) while the PNRC2 Δ NR (W114A) protein stimulated decapping strongly (**Figure 37B, lanes 8-10 and 37C, lanes 8-10**). More importantly, upon addition of the PNRC2 Δ NR-hDcp1aEVH1 protein complex, the efficiency of decapping activity of the hDcp2 protein was markedly enhanced (**Figure 37B, Lanes 11-13 and 37C, lanes 11-13**), suggesting that the PNRC2 and hDcp1aEVH1 protein worked in a cooperative manner to stimulate the Dcp2 protein decapping activity.

Figure 37. The hDcp1a and PNRC2 proteins synergistically stimulate the decapping activity of the hDcp2 protein. (A) Sequence alignment of the proline-rich region of the PNRC2 protein (residue 102 - 115) with the proline-rich sequences of the Edc1 (residue 163-175) and Edc2 proteins (residue 134-145) from *S. cerevisiae*, and human Smad4 protein (residue 290-304) (B) Effects of the hDcp1aEVH1, PNRC2ΔNR (W114A) and PNRC2ΔNR-Dcp1aEVH1 proteins on the decapping activity of the hDcp2(1-245) protein. The enhanced decapping activities were quantified using the amount of cap-labeled RNA relative to that of hDcp2 (1-245) alone (lane 4) from three independent measurements. (C) Time-course decapping reactions with the indicated proteins. The proteins were incubated with cap-labeled RNA at room temperature for the time points indicated.

To gain insight into the mechanism by which the PNRC2 protein stimulates decapping, we examined whether the GST-Dcp2 protein binds to the purified hDcp1aEVH1, PNRC2 Δ NR (W114A) and PNRC2 Δ NR-Dcp1aEVH1 protein complex by His-tag pull-down assays. Western blotting analysis using α -GST antibody showed that the PNRC2 Δ NR-Dcp1aEVH1 protein complex display the strongest interaction to the GST-Dcp2 protein, followed by the PNRC2 Δ NR (W114A) protein. However, it was observed that the hDcp1aEVH1 protein did not bind to the hDcp2 protein (**Figure 38**). Furthermore, the interaction of the PNRC2 Δ NR-Dcp1aEVH1 protein complex or the PNRC2 Δ NR (W114A) protein with the GST-Dcp2 protein is not mediated by the GST protein as the GST protein alone had no interaction with the proteins used here (**Figure 38**).

Figure 38. Interactions of PNRC2 and hDcp1a with Dcp2.

(A) While the hDcp1aEVH1 does not bind the Dcp2 protein, the PNRC2 Δ NR (W114A) binds to Dcp2 protein weakly. More importantly, PNRC2 Δ NR- Dcp1aEVH1 complex bind the Dcp2 protein with high affinity. Purified recombinant His-hDcp1aEVH1, His-PNRC2 Δ NR (W114A) and the hDcp1a EVH1-His-PNRC2 Δ NR complex (top panel) were individually precipitated with GST-Dcp2 (upper panel) or the GST alone (middle panel) and analysed by western blotting with anti-GST antibody. Co-eluted His-tagged proteins were analysed by western blotting with anti-His-tag antibody (bottom panel).

4 DISCUSSION

4.1 Yeast two-hybrid assays

To validate the results of *in vitro* His tag affinity pull down assay (**Figure 35**), yeast two-hybrid assays performed by our collaborator identified one mutation L96A in the hDcp1a protein that abolished the binding to the PNRC2 protein and four hDcp1a mutations Y36A, W45A, F94A and R98A that showed reduced binding to the PNRC2 protein (**Figure 39A**). Moreover, in PNRC2 the mutations W114A and K109A abrogated the binding of PNRC2 to hDcp1a while mutation P108A substantially reduced binding to hDcp1a (**Figure 39B**). Overall, the results from the yeast two-hybrid is in good agreement with our structural data as detailed in the binding network of the PNRC2 protein with hDcp1aEVH1 protein (**Table 4**). More importantly, the three residues of the binding defective PNRC2 mutants, Pro108, Lys109 and Trp114, are shown to form hydrogen bonds with residues from the hDcp1aEVH1 protein. Hence, this shows that hydrogen bonding makes a major contribution to the interaction between the hDcp1aEVH1 and PNRC2 protein.

Based on the binding network analysis (**Table 4**), Pro108 and Trp114 of the PNRC2 proteins are proposed to form the strongest hydrogen bonds with Trp45 and Glu87 of hDcp1aEVH1 protein respectively. This is because the distance between the hydrogen bond formation residues is less than 3Å. To determine if mutations of P108A and W114A in the PNRC2 protein affect its P-body localisation, the endogenous hDcp1a protein was used as a P-body marker. As shown in **Figure 40A**, the signal from the wild-type (WT) PNRC2 protein almost completely overlapped with that from the hDcp1a protein which indicates that the PNRC2 protein is localised to P-bodies. Conversely, the PNRC2 (W114A) mutant, which disrupts the interaction with the hDcp1a protein, was mainly localised to nucleus while the PNRC2 (P108A) mutant, with reduced hDcp1a protein interaction, was localised to both the P-bodies and nucleus. Thus, the results show that the interaction of the PNRC2 protein with the hDcp1a protein is required for its localisation to P-bodies.

To test if interaction of the PNRC2 protein with the hDcp1a protein is required for its ability to promote mRNA degradation, we examined how the P108A and W114A mutations in the PNRC2 protein affected its ability to promote mRNA degradation when tethered to mRNAs through the MS2 RNA binding proteins (**Figure 40**) (Cho et al., 2009). As assessed by quantitative real-time RT-PCR, while we observed that tethering of the PNRC2 (WT) protein strongly reduced mRNA levels, tethering of W114A and P108A PNRC2 mutants reduce the efficiency of mRNA degradation by 6.4 and 2.5-fold, respectively (**Figures 40B and 40C**). The *in vivo* tethering assay results is in good agreement with the results from the *in vitro* decapping assay where the hDcp1aEVH1-PNRC2 Δ NR complex and the hDcp1aEVH1 protein alone enhance the decapping activity of hDcp2 by 6.2 and 2.4 fold respectively (**Figure 37**). Taken together, these results indicate that the binding of the PNRC2 protein to the hDcp1a protein is important for its P-body localisation. In addition, while the *in vitro* decapping assay demonstrate that the absence of the PNRC2 does not lead to a complete abolishment of mRNA decapping activity, both the *in vitro* decapping assay and *in vivo* tethering assay show the PNRC2 protein's ability to promote mRNA degradation.

Figure 39. Yeast two hybrid assay

Yeast two-hybrid analysis of the interaction between the hDcp1a and PNRC2 proteins. (A) Wild-type and seven interface mutants of the hDcp1a in AD vector against the PNRC2 in BD vector. Five hDcp1a mutants Y36A, W45A, F94A, L96A and R98A have reduced binding with the PNRC2 protein. (B) Wild-type and five interface mutants of the PNRC2 in AD vector against the hDcp1a in BD vector. Three PNRC2 mutants P108A, K109A, W114A show significantly reduced binding with hDcp1a

Figure 40. Mutational effects on P-body localisation and NMD.

(A) Immunostaining of PNRC2 WT and its deletion variants. As shown, W114A remained in the nucleus while P108A localised to both P-bodies and nucleus. Conversely, PNRC2 WT was found solely in P bodies. (B) and (C) Tethering of PNRC2 WT and its deletion variants. (B) Western blotting to show the comparable expression of MS2 fusion proteins. (C) qRT-PCR of β -6bs mRNA and MUP mRNA. The levels of β -6bs mRNA were normalised to the level of MUP mRNA. The normalised levels of β -6bs mRNA in the presence of MS2-HA were set to 100%. As shown, tethering of W114A and P108A PNRC2 mutants reduce the efficiency of mRNA degradation by 6.4 and 2.5-fold, respectively

4.2 SiRNA PNRC2 knock out assay

Consistent with the *in vitro* binding results (**Figure 38**), downregulation of the endogenous PNRC2 protein using siRNA (**Figure 41A**), performed by our collaborator, reduced the interaction between the endogenous hDcp1a and FLAG-Dcp2 proteins by about 3-fold (**Figure 41B**), suggesting that the PNRC2 protein plays a role in mediating the association between the hDcp1a and Dcp2 proteins. Taken together, these results indicate that the PNRC2 protein is a *bona fide* decapping coactivator, and acts synergistically with the hDcp1a protein to activate decapping in addition to its adaptor role in NMD.

Figure 41. Immunoprecipitates (IP) of FLAG-Dcp2 using the extracts of HEK293T cells depleted of endogenous PNRC2.

(A) Western blotting of PNRC2. To demonstrate the quantitativity of Western blotting, 3-fold serial dilutions of total-cell extracts were loaded in the three left-most lanes. (B) IP of FLAG-Dcp2. IP was performed using α -FLAG antibody or mouse IgG as a control. The levels of co-immunopurified endogenous Dcp1a were normalised to the level of immunopurified FLAG-Dcp2. The normalised level in presence of control siRNA was set to 1.

4.3 Role of PNRC2-hDcp1a complex in NMD

In mammals, phosphorylation of the Upf1 protein is a critical event, which contributes to NMD in three ways. First, the phosphorylated Upf1 protein binds to eIF3 and interfere with eIF3-mediated conversion of 40S/Met-tRNAi^{Met}/mRNA to translationally competent 80S/Met-tRNAi^{Met}/mRNA (Isken et al., 2008). This leads to translation repression of mRNAs and allows mRNA decay machinery to assemble on the translationally repressed mRNAs. Second, the phosphorylated Upf1 protein binds to the SMG7 protein, which targets the mRNA for decay by an unknown mechanism (Unterholzner and Izaurralde, 2004). Third, Isken et al., (2008) also demonstrated that the phosphorylated Upf1 protein has greater affinity for mRNA decapping enzyme, hDcp1a, even though direct interaction between the two proteins has not been demonstrated.

Here, our results show an alternative mechanism whereby the phosphorylation of the human Upf1 protein leads to NMD. For the first time, we demonstrate that the NR box of the PNRC2 protein binds directly to the phosphorylated Upf1 protein at high affinity *in vitro*. Subsequently, the interaction between PRS of PNRC2 and EVH1 domain of hDcp1a proteins is required to target the phosphorylated Upf1 protein and its bound mRNA into P bodies for NMD. As the hDcp1a protein is involved in decapping, we proposed that NMD substrates targeted to P bodies through this mechanism undergo decapping dependent 5' - 3' decay pathway. Additionally, since the PNRC2 protein is involved in SMD (Cho et al., 2012), we further propose that SMD substrates are targeted to P bodies for decay using identical mechanism.

Our study showed that one function of the phosphorylation of the Upf1 protein is to facilitate its direct interaction with the PNRC2 protein. Coupled with the direct interaction of the PNRC2 with hDcp1a proteins, the mRNA surveillance machinery is linked to the decapping complex. Therefore, the PNRC2 protein transduces the signal of PTC recognition to the decapping complex and triggers mRNA decapping.

4.4 PNRC2 Δ NR-hDcp1aEVH1 complex is a novel class of EVH1 domain

The proline rich motif of the PNRC2 peptide, LPKP, adopts a PPII conformation and binds to the canonical PRS binding site which are the two characteristics that are common to all classes of previously characterised EVH1 domains in complex with their respective PRS peptides. A residue residing in this motif, Pro108, is one of the key specificity determinants (**Figures 39B**) which demonstrates the importance of the binding of proline rich motif of the PNRC2 protein to the canonical PRS binding site of the hDcp1a protein.

Consistent with previous observations (Beneken et al., 2000; Prehoda et al., 1999; Volkman et al., 2002), the diversity of residues flanking the core PRS ligand results in unique ligand binding mode for respective classes of EVH1 domains. Accordingly, the residues flanking the N and C terminus of the proline rich motif of the PNRC2 protein form a novel motif, which adopts a unique conformation with intermediate length as compared to all previously known PRS ligands of respective EVH1 domains. Importantly, Trp114 residing in the C-terminal portion of the PNRC2 peptide is shown to be the second critical residue affecting binding affinity. Interestingly, stacking interaction of Trp114 with Arg98 of hDcp1a is analogous to the stacking interaction between Phe1 of FPPPP peptide and equivalent Arg81 of class I EVH1 domain, Mena. Additionally, Trp114 also determines the binding orientation of the PNRC2 protein which is identical only to class III EVH1 domain's PRS ligand, WIP. Consistent with these observations, point mutations of Pro108 and Trp114 to alanine in the PNRC2 protein showed strong defects in hDcp1a binding, P-body localisation and degradation of tethered mRNAs. Taken together, the results demonstrate that PNRC2 Δ NR-hDcp1aEVH1 complex is a novel class of EVH1 domain.

4.5 Critical Residues for PNRC2 and hDcp1aEVH1 interaction

The contribution of binding interface residues to the association of the PNRC2 to hDcp1aEVH1 proteins have been studied through *in vitro* His tag affinity pull down assay and *in vivo* yeast two-hybrid assay. Both assays confirmed the importance of Trp114 of the PNRC2 protein in mediating the interaction between the PNRC2 and hDcp1aEVH1 proteins.

Unexpectedly, three mutants (F94A, L96A of the hDcp1aEVH1 proteins and K109A of the PNRC2 Δ NR protein) maintain strong interaction in His tag affinity pull down assay but show significantly reduced interaction in yeast two-hybrid assay. Structure analysis of the PNRC2 Δ NR-hDcp1aEVH1 protein complex reveals that Phe94 and Leu96 are part of hydrophobic pocket 1 and they are involved in hydrophobic interaction with Leu107, Pro108 and Pro110 in the LPKP motif of the PNRC2 protein. Additionally, Lys109 of the PNRC2 protein forms a hydrogen bond with the carbonyl oxygen of Asn43. Thus, the result of yeast two-hybrid assay is in agreement with structure analysis. Hence, we suggest that the discrepancy of results between yeast two-hybrid assay and *in vitro* His tag affinity pull down assay is probably due to the overexpression of mutant proteins in the cell lysate. We postulate that high concentration of mutant proteins is able to compensate for the reduction in interaction between the PNRC2 and hDcp1aEVH1 proteins and hence leading to false positives in our *in vitro* His tag affinity pull down assay.

Borja et al., (2011) used a randomised nonapeptide library presented on capsid of filamentous phage M13 to isolate peptides that interact with ScDcp1. After four rounds of panning, they observe the enrichment of 20 sequences that have the motif of FPRP[S/T][F/W]. Hence, they proposed that the preference for a basic residue between the prolines may be a general motif common to coactivators of Dcp1. In support of their finding, our yeast two-hybrid assay shows that Lys109, which lies between two proline residues, is a critical residue that mediates the interaction between the PNRC2 and hDcp1aEVH1 proteins. This demonstrates that the general motif of coactivators of Dcp1 is conserved from *S. cerevisiae* to humans.

4.6 Conserved binding mode of PRS proteins with Dcp1

Due to poor solubility of the Dcp1 protein, Borja et al (2011) failed to identify a definitive interaction mapping of EVH1 domain of the Dcp1 protein to PRS of the Edc1 and Edc2 proteins. However, upon addition of the Edc1 protein, they noted that the most prominent chemical shift changes of the Dcp1 protein occur in the vicinity of the putative PRS binding site. In support of the study, our structure proved that the putative PRS binding site is indeed the binding site for PRS of the PNRC2 protein. Additionally, as sequence alignment of PRS of the PNRC2, Smad4, Edc1 and Edc2 proteins showed that the PRS are highly similar (**Figure 37A**), we proposed that the Smad4, Edc1 and Edc2 proteins might bind Dcp1 protein in a similar fashion as the PNRC2 protein.

In agreement with sequence alignment results, our *in vivo* yeast two-hybrid mutagenesis study also suggests that the Edc1, PNRC2 and Smad4 proteins bind the Dcp1 protein in a similar manner. For example, the mutation of Tyr47 to alanine in the ScDcp1 protein (equivalent to Tyr36 in the hDcp1a protein) disrupted its binding to the Edc1 PRS peptide (Borja et al., 2011). Moreover, analogous to the Pro108 and Trp114 mutations in the PNRC2 protein, the corresponding residues in *S. cerevisiae* Edc1 (Pro170) (Borja et al., 2011) and Smad4 (Trp302) proteins (Bai et al., 2002) are found to be essential for binding to the ScDcp1 and hDcp1a proteins, respectively. The conservation of this aromatic Trp residue in the PRS peptide between the PNRC2 and SMAD4 proteins can be extended to the Edc1 and Edc2 as both proteins contain a Phe in the consensus PRS sequence (**Figure 37A**). These observations suggest that the presence of an aromatic residue (Phe or Trp) in the C-terminus of the PRS peptide could be a general feature for binding to the EVH1 domain of the Dcp1 proteins. Given that the PNRC2, Edc1 and Edc2 proteins stimulate decapping, it is tempting to speculate that the Smad4 protein may have an underappreciated role of being a decapping activator.

4.7 Assembly of distinct decapping complex

In humans, the Dcp1-Dcp2 protein interaction appears weaker and it requires adaptor proteins (Lsm1-7, Edc3, Hedls, DDX6/RCK) to bridge the interaction of Dcp1-Dcp2. Accordingly, it was observed that distinct decapping complexes are assembled to target specific subclass of mRNA. For example, the TTP protein associates with Dcp2 and Dcp1 proteins to form a decapping complex specific for ARE containing mRNA while the hDcp2, Dcp1a, Hedls, Edc3, Lsm1 and Lsm4 form an unique decapping complex for human 3' uridylated mRNA (Fenger-Gron et al., 2005). In support of the model, it was found that the hDcp1a protein can interact with the hDcp2 and Hedls proteins independently of the Edc3 and DDX6/RCK proteins. This suggests the existence of two multimeric complexes consisting of (i) the hDcp1a, Hedls and hDcp2 proteins (ii) the hDcp1a, Edc3 and DDX6/RCK proteins in humans (Tritschler et al., 2009).

In support of previous studies, our *in vitro* decapping assays show that the PNRC2 protein works in synergy with the Dcp1a protein to promote the decapping activity of the Dcp2 protein (**Figure 37B and 37C**). Additionally, our *in vitro* pull down and western blotting analyses indicate that the PNRC2 protein directly binds to the Dcp2 protein albeit with low affinity while the binding of the hDcp1a to Dcp2 proteins is mediated by the PNRC2 protein (**Figure 38**). Consistent with *in vitro* binding results, downregulation of the endogenous PNRC2 via siRNA (**Figure 41B**) also reduce the interaction between endogenous hDcp1a and FLAG-Dcp2. Taken together, our results show that PNRC2 alone is sufficient to mediate the association of the hDcp1a and Dcp2 proteins to form a functional complex with enhanced rate of decapping activity. Additionally, the decapping activity of this novel decapping complex is independent of the presence of additional decapping coactivators (Lsm1-7, Edc3, Hedls, DDX6/RCK). Given the role that PNRC2 plays in NMD and SMD, it is tempting to speculate that the PNRC2-hDcp1a-Dcp2 decapping complex may specifically target NMD and SMD substrates for decay. Furthermore, we also show that the PNRC2 protein stimulates decapping through direct binding to the Dcp2 protein while the hDcp1a protein stimulates decapping through PNRC2-mediated recruitment to the Dcp2 protein. Conversely, *S. cerevisiae* Dcp1 and Dcp2 proteins physically interact with each other to form the holo-decapping

enzyme (Beelman et al., 1996; Sakuno et al., 2004; Steiger et al., 2003). Furthermore, it was demonstrated that the Edc1 and Edc2 proteins stimulate decapping through ScDcp1-mediated recruitment to the Dcp2 protein which was in contrast to the mode of action of the PNRC2 protein (Borja et al., 2011).

Superpositioning of our structure with the SpDcp1-Dcp2 binding interface (2QKL; she et al, 2008) reveals that four out of nine SpDcp1 residues and three out of eight SpDcp2 residues involved in binding are conserved in the Dcp1a and Dcp2 proteins (**Figure 42A**). Thus, the low sequence conservation of Dcp1-Dcp2 binding interface across species explains the relatively weaker interaction between the hDcp1a and Dcp2 proteins. More importantly, our structure (**Figure 42B**) suggests that N terminus of the PNRC2 protein may be responsible for contacting the Dcp2 protein as it is physically closer to the Dcp2 protein.

Figure 42. Superposition of PNRC2 Δ NR-hDcp1aEVH1 complex to SpDcp1-SpDcp2 interface.

(A) Sequence alignment of Dcp2 binding region in the SpDcp1p, ScDcp1p and hDcp1a protein; Sequence alignment of Dcp1 binding region in the SpDcp2p, ScDcp2p and hDcp2 proteins. Invariant residues are shown in red. (B) Conserved residues that are involved in Dcp1-Dcp2 interaction are labelled and represented as sticks. The SpDcp2, SpDcp1, hDcp1a and PNRC2 proteins are displayed in blue, wheat, pale green and pink respectively.

4.8 Synthegradases

Two recent reports have demonstrated that transcription factors and DNA promoters can directly influence the relative stability of transcripts that they produce (Bregman et al., 2011; Trcek et al., 2011), suggesting that transcription and mRNA decay are coordinated to each other to regulate the level of an mRNA. In support of these observations, the mRNA decapping factors hDcp1a, Dcp2 and Edc3 and exonuclease Xrn2 proteins were reported to function in widespread premature termination of RNA polymerase II transcription (Brannan et al., 2012). Additionally, besides its role in mRNA decay, the hDcp1a (also called SMIF) protein has been shown to function as a crucial transcriptional co-activator in TGF β signalling (Bai et al., 2002). In this study, we demonstrate that the PNRC2 protein stimulates mRNA decapping. Therefore both the hDcp1a and PNRC2 are bifunctional proteins, which regulate both mRNA decay and transcription by acting as a decapping activator and a transcriptional coactivator. Proteins that play a dual role in activating transcription and mRNA decay simultaneously are known as “synthegradases” (Bregman et al., 2011).

The conflicting functions inherent in both the hDcp1a and PNRC2 proteins suggest the existence of a control mechanism to ensure that stimulation of transcription and decay of mRNA do not occur simultaneously. In this study, our structural data has provided some clues to this control mechanism. Our results show that part of the SH3-binding motif (residues 99-105) of the PNRC2 protein binds to the hDcp1aEVH1 surface with the side-chains of the proline residues (Pro102 and Pro104) exposed to the solvent region (**Figure 34A**). This suggests that SH3-binding motif does not contribute much to the binding specificity towards hDcp1a. Consistent with this observation, the SH3 binding motif of PNRC2 has been shown to be essential only for its interaction with the nuclear receptors (Zhou et al., 2006). Importantly, the overlap between the hDcp1a binding site and SH3 binding motif suggests that when the PNRC2 protein is in complex with the hDcp1a protein to promote decay of mRNA, it is unable to bind nuclear receptors to activate transcription of mRNA. Alternatively, as interaction with the hDcp1a protein leads to sequestering of the PNRC2 protein in P bodies, the PNRC2 protein is also inhibited spatially from interaction with nuclear receptors in the nucleus.

Similarly, as discussed above (**section 4.6**), results from our study propose that the Smad4 protein is a potential competitor of PNRC2 in contacting the PRS binding site of hDcp1a EVH1 domain. Thus, this suggests that when Dcp1a-Dcp2 proteins is in complex with PNRC2 to carry out mRNA decapping, the hDcp1a protein is unable to bind the Smad4 protein to activate TGF β pathway to initiate transcription of its target genes.

4.9 hDcp1a is a potential control point that determines cell survival

The Grb2 is an adaptor protein that binds to growth factor-activated receptor via its SH2 domain and simultaneously form complex with SOS (guanine nucleotide exchange factor) with its SH3 domain. Hence, this brings SOS into close proximity with the Ras protein and activates the Ras signalling pathway. The Ras signalling pathway activate a cascade of kinases that phosphorylates many transcriptional factors involved in cell cycle regulation and leads to cell proliferation and differentiation. Additionally, the Ras protein is a common oncogene in human cancer as overactive Ras protein often leads to cancer.

In addition to its function as a nuclear receptor coactivator, the PNRC2 protein isoform, PNRC, is found to interact with SH3 domains of Grb2 via its N and C terminus SH3 binding motifs (Zhou et al., 2004). It was shown that the PNRC protein binds to Grb2 protein and suppress Grb2 mediated Ras activation. In support of the model, the PNRC protein expression is found to be significantly lower in breast cancer tissue as compared to noncancer tissue. Furthermore, overexpression of the PNRC protein was shown to inhibit cell growth in HeLa cells.

Sequence alignment of PNRC with PNRC2 proteins (**Figure 43**) reveals that the PNRC protein has the exact PRS sequence that binds the hDcp1a protein. This shows that the PNRC protein may be an interaction partner of the hDcp1aEVH1 protein. Given that the N terminus and C terminus SH3 binding motifs of PNRC are required for binding Grb2, we speculate that binding of the PNRC to Dcp1aEVH1 proteins will abolish the interaction between the PNRC and Grb2 proteins. This implies that the Grb2 protein is no longer sequestered and it can activate the Ras signalling pathway to stimulate cell proliferation and differentiation.

As described above, we proposed that the Smad4 protein binds to the same site on hDcp1a as PNRC protein (**Section 4.6**). Thus, the Smad4 and PNRC proteins are proposed to be binding competitors to the hDcp1aEVH1

protein. In previous study, it was demonstrated that the Smad4 protein binds to the hDcp1a protein to activate the TGF β pathway, which arrests cell cycle at G1 phase to stop cell proliferation and activate apoptosis. This has an interesting implication. Depending on the binding partners (Smad4 or PNRC), the hDcp1a protein may serve as a control point, which determines if the cell proliferates or undergoes apoptosis. This proposed function of the hDcp1a protein may be useful in the treatment of cancer.

Figure 43. Sequence alignment of PNRC with PNRC2

The PNRC protein has identical PRS sequence (residues 288-300) as the PNRC2 protein, which suggests that it is a interaction partner of the Dcp1aEVH1 protein.

4.10 CIKS is a potential interaction partner of hDcp1a

The CIKS protein is proposed to be an adaptor protein that can bind to regulatory sub-unit NEMO/IKK γ and SAPK/JNK kinases to activate both complexes (Leonardi et al., 2000; Mauro et al., 2003). The activation of IKK leads to the activation of NF- κ B protein, which plays an important role in stress response, inflammation and regulation of apoptosis. In a separate study, the authors observed that overexpression of the hDcp1a protein led to suppression of NF- κ B nuclear activity (Rzeczkowski et al., 2011). Subsequently, alignment of proline rich sequence of the PNRC2 protein with the CIKS protein shows that the CIKS protein share similar sequence with the PNRC2 protein (**Figure 44**). More importantly, critical binding residues, Pro108, Lys109 and Trp114, of the PNRC2 protein are all conserved in equivalent residues of the CIKS protein. Thus, this suggests a potential interaction between the CIKS protein with EVH1 domain of the hDcp1a protein. Thus, we speculate that the mechanism of hDcp1a suppression of NF- κ B activity may be due to the protein sequestering the CIKS protein, which leads to inactivation of IKK. Nevertheless, the interaction between the hDcp1a and CIKS proteins remains to be established and it will be interesting to observe if the interaction affects the activation of NF- κ B pathway.

Figure 44. Sequence alignment of Ciks with PNRC2.

As shown, the human Ciks protein (381-395) shares high sequence conservation with the prs of the PNRC2 protein. This suggests that the Ciks protein may be a potential interaction partner of the hDcp1a protein.

4.11 C-terminal extension of Dcp1a

One of the main differences in the yeast Dcp1 from the hDcp1a protein is the C terminal extension, which is also present in *D. melanogaster* Dcp1. Previous study showed that residues from C-terminal extension (539-582 aa) undergo trimerisation to form a novel antiparallel assembly that comprised of three kinked α helices. Functional data provide evidences that the trimerisation domain (TD) is required for the hDcp1a protein to interact with either the Dcp2 or Hedls proteins. Additionally, the domain facilitates hDcp1a recruitment into P bodies and it is required for efficient decapping *in vivo* (Tritschler et al., 2009).

Conversely, our *in vitro* His tag pull down assay and *in vivo* immunoprecipitation results show that the hDcp1aEVH1 protein is able to bind Dcp2 protein in the presence of PNRC2 protein. Thus, this shows that the hDcp1a protein can interact with the Dcp2 protein in a TD independent manner, which suggests that TD and the PNRC2 protein may act independently to facilitate the binding of the hDcp1a to Dcp2 proteins.

In a separate study, Ser315 at C terminal extension of the hDcp1a protein is phosphorylated by c-Jun N terminal kinase (JNK), which is stimulated transiently by interleukin 1 (IL-1) or extendedly by translational stress (Rzeczkowski et al., 2011). Subsequently, it was shown that the phosphorylation of Ser315 of the hDcp1a protein affects the localisation of the hDcp1a protein in P bodies and the formation of P bodies. Additionally, it was observed that phosphorylation of Ser 315 does not affect decapping activity. This is in agreement with our results as we demonstrated that the hDcp1aEVH1 and the PNRC2 proteins are sufficient in stimulating the decapping activity of the Dcp2 protein.

Besides decapping, phosphorylation of Ser315 of the hDcp1a protein also mediates IL-1 gene response and specifically IL-1 induced expression of IL-8. Mutation of Ser315 to alanine completely suppresses the secretion of IL-8, which demonstrates the importance of C terminal hDcp1a protein in effecting IL-1 response. Further promoter gene reporter study suggests that the silencing of IL-1 gene response is due to suppression of transcriptional activity of a major regulator of IL-1 induced genes, NF- κ B, by wild type or mutant hDcp1a protein.

Nevertheless, the suppression effect of the hDcp1a protein was independent from phosphorylation of Ser 315. Interestingly, coimmunoprecipitation assays using NF- κ B or hDcp1a antibodies suggest that both proteins interact directly. As postulated above (**Section 4.10**), it is possible that the CIKS protein may mediate the interaction between the hDcp1a and NF- κ B proteins.

Based on our results, we speculate that the C-terminal extension of hDcp1a protein and its role in facilitating decapping may be due to the weaker binding between the hDcp1a and Dcp2 proteins. More importantly, the involvement of C terminus of hDcp1a in IL-1 signalling pathway is novel. This suggests that Dcp1 has evolved in higher eukaryotes to participate in novel functions besides decapping. We propose that the alternative decapping enzyme, Nudt16, present in mammals may have partially substituted the decapping function of the hDcp1a protein. Thus, the functionally released hDcp1a protein may evolve to acquire novel functions.

5 Conclusion and future direction

5.1 Conclusion

As stated in the rationales of this thesis (**Section 1.5**), an important issue in mammalian NMD is to understand the mechanism utilised by the phosphorylated Upf1 protein to recruit downstream decapping enzyme to initiate the degradation of mRNAs. In this thesis, the atomic structure of the hDcp1aEVH1 protein in complex with the PNRC2 peptide and subsequent ITC analysis of PNRC2 (FL) with the hyperphosphorylated Upf1 peptide have shown for the first time that the PNRC2 protein binds directly to the hDcp1a and Upf1 proteins. Supplemented with functional data from assays, this thesis has shown compelling evidence for the following model. The hyperphosphorylated Upf1 protein binds to NR box at the C terminus of the PNRC2 protein. Consequentially, the PNRC2 protein targets the Upf1 bound transcript to P body via its interaction with the EVH1 domain of hDcp1a protein. In the P body, the PNRC2 protein mediates the formation of hDcp1a-Dcp2 decapping complex to remove the 5' cap of the transcript so that it can be subjected to 5' - 3' decay by the Xrn1 exonuclease.

Additionally, our structure shows that the hDcp1aEVH1 protein binds the PNRC2 protein through a novel mechanism, which was further confirmed by *in vitro* and *in vivo* mutagenesis studies. Hence this establishes the hDcp1aEVH1 protein as a novel class of EVH1 domain. Moreover, given the conflicting roles of the PNRC2 and hDcp1a proteins in promoting the transcription and decay of mRNAs, we have identified the structure basis of a switch that prevents simultaneous stimulation of both processes.

Finally, previous studies have shown that metazoan Dcp1 and Dcp2 proteins require additional proteins (Lsm1-7, Edc3, Hedls, DDX6/RCK) for decapping complex formation. Nevertheless, we have shown that the PNRC2 protein alone is able to mediate the formation of hDcp1a-Dcp2 complex and stimulate decapping *in vitro*. Thus, this lends support that distinct decapping complexes can be assembled to target different subsets of mRNAs.

5.2 Future direction

The results from this thesis have highlighted a number of issues that remain to be solved. First, the molecular interaction of the hyperphosphorylated Upf1 protein with PNRC2 protein has not been solved. Hence, the structure of the hyperphosphorylated Upf1-PNRC2-hDcp1a protein complex will be useful in studying the functional importance of phosphorylation of the Upf1 protein in metazoan cells.

Second, the discovery of a novel decapping complex, PNRC2-hDcp1a-Dcp2, shows that distinct decapping complexes can be assembled. It will be interesting to elucidate the binding mechanism of the PNRC2-hDcp1a-Dcp2 decapping complex and determine the mode of stimulation used by the PNRC2 protein in decapping.

Third, we have shown that the binding of the PNRC2 protein to hDcp1a protein synergistically stimulates the decapping activity of the hDcp2 protein. Thus, with our structural data, a PNRC2 protein mimic can be designed to inhibit the binding of the hDcp1a protein with the PNRC2 protein and hence reduce the rate of decapping activity by the hDcp2 protein. This may be beneficial to patients suffering from Ulrich's disease (**Section 1.1.10**) where NMD aggravates the disease.

Sequence alignment of the PNRC2, Edc1 and Edc2 proteins has also identified three potential interaction partners (Smad4, PNRC and Ciks) for the hDcp1aEVH1 protein. The three potential interaction partners may have similar function as the PNRC2 protein and serve as decapping activators that enhance decapping activity of the Dcp1-Dcp2 decapping complex. More importantly, the Smad4, PNRC and Ciks proteins are involved in the TGF β , Ras and IL-1 signalling pathways respectively. As these signalling pathways are implicated in apoptosis, cancer and inflammation response, potential interaction of these proteins with hDcp1a protein suggests that the hDcp1a protein may mediate the progression of these signalling pathways.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213-221.

Altamura, N., Groudinsky, O., Dujardin, G., and Slonimski, P.P. (1992). NAM7 nuclear gene encodes a novel member of a family of helicases with a Zn-ligand motif and is involved in mitochondrial functions in *Saccharomyces cerevisiae*. *J Mol Biol* 224, 575-587.

Amrani, N., Dong, S., He, F., Ganesan, R., Ghosh, S., Kervestin, S., Li, C., Mangus, D.A., Spatrick, P., and Jacobson, A. (2006). Aberrant termination triggers nonsense-mediated mRNA decay. *Biochem Soc Trans* 34, 39-42.

Anders, K.R., Grimson, A., and Anderson, P. (2003). SMG-5, required for *C.elegans* nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. *EMBO J* 22, 641-650.

Andrei, M.A., Ingelfinger, D., Heintzmann, R., Achsel, T., Rivera-Pomar, R., and Luhrmann, R. (2005). A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. *RNA* 11, 717-727.

Azzalin, C.M., and Lingner, J. (2006). The double life of UPF1 in RNA and DNA stability pathways. *Cell Cycle* 5, 1496-1498.

Azzalin, C.M., Reichenbach, P., Khoriantuli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 318, 798-801.

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

Baggs, J.E., and Green, C.B. (2003). Nocturnin, a deadenylase in *Xenopus laevis* retina: a mechanism for posttranscriptional control of circadian-related mRNA. *Curr Biol* 13, 189-198.

Bai, R.Y., Koester, C., Ouyang, T., Hahn, S.A., Hammerschmidt, M., Peschel, C., and Duyster, J. (2002). SMIF, a Smad4-interacting protein that functions as a co-activator in TGFbeta signalling. *Nat Cell Biol* 4, 181-190.

Ball, L.J., Jarchau, T., Oschkinat, H., and Walter, U. (2002). EVH1 domains: structure, function and interactions. *FEBS Lett* 513, 45-52.

Banerjee, S.S., Feinberg, M.W., Watanabe, M., Gray, S., Haspel, R.L., Denking, D.J., Kawahara, R., Hauner, H., and Jain, M.K. (2003). The Kruppel-like factor KLF2 inhibits peroxisome proliferator-activated receptor-gamma expression and adipogenesis. *J Biol Chem* 278, 2581-2584.

Barbee, S.A., Estes, P.S., Cziko, A.M., Hillebrand, J., Luedeman, R.A., Collier, J.M., Johnson, N., Howlett, I.C., Geng, C., Ueda, R., et al. (2006). Staufen- and FMRP-containing neuronal RNPs are structurally and functionally related to somatic P bodies. *Neuron* 52, 997-1009.

Beckerle, M.C. (1998). Spatial control of actin filament assembly: lessons from *Listeria*. *Cell* 95, 741-748.

Beelman, C.A., and Parker, R. (1994). Differential effects of translational inhibition in cis and in trans on the decay of the unstable yeast MFA2 mRNA. *J Biol Chem* 269, 9687-9692.

Beelman, C.A., Stevens, A., Caponigro, G., LaGrandeur, T.E., Hatfield, L., Fortner, D.M., and Parker, R. (1996). An essential component of the decapping enzyme required for normal rates of mRNA turnover. *Nature* 382, 642-646.

Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., and Izaurralde, E. (2007a). A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J* 26, 1591-1601.

Behm-Ansmant, I., and Izaurralde, E. (2006). Quality control of gene expression: a stepwise assembly pathway for the surveillance complex that triggers nonsense-mediated mRNA decay. *Genes Dev* 20, 391-398.

Behm-Ansmant, I., Kashima, I., Rehwinkel, J., Sauliere, J., Wittkopp, N., and Izaurralde, E. (2007b). mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. *FEBS Lett* 581, 2845-2853.

Behm-Ansmant, I., Rehwinkel, J., and Izaurralde, E. (2006). MicroRNAs silence gene expression by repressing protein expression and/or by promoting mRNA decay. *Cold Spring Harb Symp Quant Biol* 71, 523-530.

Belgrader, P., Cheng, J., Zhou, X., Stephenson, L.S., and Maquat, L.E. (1994). Mammalian nonsense codons can be cis effectors of nuclear mRNA half-life. *Mol Cell Biol* 14, 8219-8228.

Beneken, J., Tu, J.C., Xiao, B., Nuriya, M., Yuan, J.P., Worley, P.F., and Leahy, D.J. (2000). Structure of the Homer EVH1 domain-peptide complex reveals a new twist in polyproline recognition. *Neuron* 26, 143-154.

Bessman, M.J., Frick, D.N., and O'Handley, S.F. (1996). The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J Biol Chem* 271, 25059-25062.

Bhattacharya, A., Czaplinski, K., Trifillis, P., He, F., Jacobson, A., and Peltz, S.W. (2000). Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA* 6, 1226-1235.

Boeck, R., Lapeyre, B., Brown, C.E., and Sachs, A.B. (1998). Capped mRNA degradation intermediates accumulate in the yeast *spb8-2* mutant. *Mol Cell Biol* 18, 5062-5072.

Borja, M.S., Piotukh, K., Freund, C., and Gross, J.D. (2011). Dcp1 links coactivators of mRNA decapping to Dcp2 by proline recognition. *RNA* 17, 278-290.

Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Seraphin, B. (2000). A Sm-like protein complex that participates in mRNA degradation. *EMBO J* 19, 1661-1671.

Bowler, M.W., Guijarro, M., Petitdemange, S., Baker, I., Svensson, O., Burghammer, M., Mueller-Dieckmann, C., Gordon, E.J., Flot, D., McSweeney, S.M., et al. (2010). Diffraction cartography: applying microbeams to macromolecular crystallography sample evaluation and data collection. *Acta Crystallogr D Biol Crystallogr* 66, 855-864.

Brakeman, P.R., Lanahan, A.A., O'Brien, R., Roche, K., Barnes, C.A., Huganir, R.L., and Worley, P.F. (1997). Homer: a protein that selectively binds metabotropic glutamate receptors. *Nature* 386, 284-288.

Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J., et al. (2012). mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. *Mol Cell* 46, 311-324.

Bregman, A., Avraham-Kelbert, M., Barkai, O., Duek, L., Guterman, A., and Choder, M. (2011). Promoter elements regulate cytoplasmic mRNA decay. *Cell* 147, 1473-1483.

Brengues, M., Teixeira, D., and Parker, R. (2005). Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science* 310, 486-489.

Brocke, K.S., Neu-Yilik, G., Gehring, N.H., Hentze, M.W., and Kulozik, A.E. (2002). The human intronless melanocortin 4-receptor gene is NMD insensitive. *Hum Mol Genet* 11, 331-335.

Brown, C.E., and Sachs, A.B. (1998). Poly(A) tail length control in *Saccharomyces cerevisiae* occurs by message-specific deadenylation. *Mol Cell Biol* 18, 6548-6559.

Brumbaugh, K.M., Otterness, D.M., Geisen, C., Oliveira, V., Brognard, J., Li, X., Lejeune, F., Tibbetts, R.S., Maquat, L.E., and Abraham, R.T. (2004). The mRNA

surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. *Mol Cell* 14, 585-598.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54, 905-921.

Bruno, I., and Wilkinson, M.F. (2006). P-bodies react to stress and nonsense. *Cell* 125, 1036-1038.

Cao, D., and Parker, R. (2003). Computational modeling and experimental analysis of nonsense-mediated decay in yeast. *Cell* 113, 533-545.

Carastro, L.M., Tan, C.K., Selg, M., Jack, H.M., So, A.G., and Downey, K.M. (2002). Identification of delta helicase as the bovine homolog of HUPF1: demonstration of an interaction with the third subunit of DNA polymerase delta. *Nucleic Acids Res* 30, 2232-2243.

Carrier, M.F., Ducruix, A., and Pantaloni, D. (1999). Signalling to actin: the Cdc42-N-WASP-Arp2/3 connection. *Chem Biol* 6, R235-240.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760-763.

Chakraborty, T., Ebel, F., Domann, E., Niebuhr, K., Gerstel, B., Pistor, S., Temm-Grove, C.J., Jockusch, B.M., Reinhard, M., Walter, U., et al. (1995). A focal adhesion factor directly linking intracellularly motile *Listeria monocytogenes* and *Listeria ivanovii* to the actin-based cytoskeleton of mammalian cells. *EMBO J* 14, 1314-1321.

Chan, W.K., Huang, L., Gudikote, J.P., Chang, Y.F., Imam, J.S., MacLean, J.A., 2nd, and Wilkinson, M.F. (2007). An alternative branch of the nonsense-mediated decay pathway. *EMBO J* 26, 1820-1830.

- Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 76, 51-74.
- Chen, C.Y., and Shyu, A.B. (2003). Rapid deadenylation triggered by a nonsense codon precedes decay of the RNA body in a mammalian cytoplasmic nonsense-mediated decay pathway. *Mol Cell Biol* 23, 4805-4813.
- Chen, Z., Smith, K.R., Batterham, P., and Robin, C. (2005). Smg1 nonsense mutations do not abolish nonsense-mediated mRNA decay in *Drosophila melanogaster*. *Genetics* 171, 403-406.
- Cheng, J., and Maquat, L.E. (1993). Nonsense codons can reduce the abundance of nuclear mRNA without affecting the abundance of pre-mRNA or the half-life of cytoplasmic mRNA. *Mol Cell Biol* 13, 1892-1902.
- Cheng, Z., Muhlrads, D., Lim, M.K., Parker, R., and Song, H. (2007). Structural and functional insights into the human Upf1 helicase core. *EMBO J* 26, 253-264.
- Chiu, S.Y., Lejeune, F., Ranganathan, A.C., and Maquat, L.E. (2004). The pioneer translation initiation complex is functionally distinct from but structurally overlaps with the steady-state translation initiation complex. *Genes Dev* 18, 745-754.
- Chiu, S.Y., Serin, G., Ohara, O., and Maquat, L.E. (2003). Characterization of human Smg5/7a: a protein with similarities to *Caenorhabditis elegans* SMG5 and SMG7 that functions in the dephosphorylation of Upf1. *RNA* 9, 77-87.
- Cho, H., Kim, K.M., Han, S., Choe, J., Park, S.G., Choi, S.S., and Kim, Y.K. (2012). Staufen1-Mediated mRNA Decay Functions in Adipogenesis. *Mol Cell* 46, 495-506.
- Cho, H., Kim, K.M., and Kim, Y.K. (2009). Human proline-rich nuclear receptor coregulatory protein 2 mediates an interaction between mRNA surveillance machinery and decapping complex. *Mol Cell* 33, 75-86.

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, 998-1016.

Chowdhury, A., and Tharun, S. (2008). Lsm1 mutations impairing the ability of the Lsm1p-7p-Pat1p complex to preferentially bind to oligoadenylated RNA affect mRNA decay in vivo. *RNA* 14, 2149-2158.

Chowdhury, A., and Tharun, S. (2009). Activation of decapping involves binding of the mRNA and facilitation of the post-binding steps by the Lsm1-7-Pat1 complex. *RNA* 15, 1837-1848.

Chu, C.Y., and Rana, T.M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *PLoS Biol* 4, e210.

Cohen, L.S., Mikhli, C., Jiao, X., Kiledjian, M., Kunkel, G., and Davis, R.E. (2005). Dcp2 Decaps m²,2,7GpppN-capped RNAs, and its activity is sequence and context dependent. *Mol Cell Biol* 25, 8779-8791.

Coller, J., and Parker, R. (2004). Eukaryotic mRNA decapping. *Annu Rev Biochem* 73, 861-890.

Coller, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. *Cell* 122, 875-886.

Coller, J.M., Gray, N.K., and Wickens, M.P. (1998). mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. *Genes Dev* 12, 3226-3235.

Coller, J.M., Tucker, M., Sheth, U., Valencia-Sanchez, M.A., and Parker, R. (2001). The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. *RNA* 7, 1717-1727.

Conti, E., and Izaurralde, E. (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr Opin Cell Biol* 17, 316-325.

Cosson, B., Berkova, N., Couturier, A., Chabelskaya, S., Philippe, M., and Zhouravleva, G. (2002). Poly(A)-binding protein and eRF3 are associated in vivo in human and *Xenopus* cells. *Biol Cell* 94, 205-216.

Cougot, N., Babajko, S., and Seraphin, B. (2004). Cytoplasmic foci are sites of mRNA decay in human cells. *J Cell Biol* 165, 31-40.

Couttet, P., and Grange, T. (2004). Premature termination codons enhance mRNA decapping in human cells. *Nucleic Acids Res* 32, 488-494.

Culbertson, M.R., and Leeds, P.F. (2003). Looking at mRNA decay pathways through the window of molecular evolution. *Curr Opin Genet Dev* 13, 207-214.

Czaplinski, K., Ruiz-Echevarria, M.J., Paushkin, S.V., Han, X., Weng, Y., Perlick, H.A., Dietz, H.C., Ter-Avanesyan, M.D., and Peltz, S.W. (1998). The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes Dev* 12, 1665-1677.

Czaplinski, K., Weng, Y., Hagan, K.W., and Peltz, S.W. (1995). Purification and characterization of the Upf1 protein: a factor involved in translation and mRNA degradation. *RNA* 1, 610-623.

Danckwardt, S., Neu-Yilik, G., Thermann, R., Frede, U., Hentze, M.W., and Kulozik, A.E. (2002). Abnormally spliced beta-globin mRNAs: a single point mutation generates transcripts sensitive and insensitive to nonsense-mediated mRNA decay. *Blood* 99, 1811-1816.

De la Fortelle E, B.G. (1997). Maximum-likelihood heavy-atom parameter refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Method. *Methods Enzymol* 276, 472-494.

Decker, C.J., and Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev* 7, 1632-1643.

Decker, C.J., Teixeira, D., and Parker, R. (2007). Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in *Saccharomyces cerevisiae*. *J Cell Biol* 179, 437-449.

Denning, G., Jamieson, L., Maquat, L.E., Thompson, E.A., and Fields, A.P. (2001). Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein. *J Biol Chem* 276, 22709-22714.

Derry, J.M., Ochs, H.D., and Francke, U. (1994). Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. *Cell* 78, 635-644.

Deshmukh, M.V., Jones, B.N., Quang-Dang, D.U., Flinders, J., Floor, S.N., Kim, C., Jemielity, J., Kalek, M., Darzynkiewicz, E., and Gross, J.D. (2008). mRNA decapping is promoted by an RNA-binding channel in Dcp2. *Mol Cell* 29, 324-336.

Dlakic, M. (2000). Functionally unrelated signalling proteins contain a fold similar to Mg²⁺-dependent endonucleases. *Trends Biochem Sci* 25, 272-273.

Dunckley, T., and Parker, R. (1999). The DCP2 protein is required for mRNA decapping in *Saccharomyces cerevisiae* and contains a functional MutT motif. *EMBO J* 18, 5411-5422.

Dunckley, T., Tucker, M., and Parker, R. (2001). Two related proteins, Edc1p and Edc2p, stimulate mRNA decapping in *Saccharomyces cerevisiae*. *Genetics* 157, 27-37.

Eberle, A.B., Lykke-Andersen, S., Muhlemann, O., and Jensen, T.H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* 16, 49-55.

Eberle, A.B., Stalder, L., Mathys, H., Orozco, R.Z., and Muhlemann, O. (2008). Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biol* 6, e92.

El-Hakeh, J., Rosenzweig, S., Oleastro, M., Basack, N., Berozdnic, L., Molina, F., Rivas, E.M., Zelazko, M., and Danielian, S. (2002). Wiskott-Aldrich syndrome in Argentina: 17 unique, including nine novel, mutations. *Hum Mutat* 19, 186-187.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.F., Doerks, T., Dorner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes Dev* 21, 2558-2570.

Fenger-Gron, M., Fillman, C., Norrild, B., and Lykke-Andersen, J. (2005). Multiple processing body factors and the ARE binding protein TTP activate mRNA decapping. *Mol Cell* 20, 905-915.

Ferraiuolo, M.A., Basak, S., Dostie, J., Murray, E.L., Schoenberg, D.R., and Sonenberg, N. (2005). A role for the eIF4E-binding protein 4E-T in P-body formation and mRNA decay. *J Cell Biol* 170, 913-924.

Ferraiuolo, M.A., Lee, C.S., Ler, L.W., Hsu, J.L., Costa-Mattioli, M., Luo, M.J., Reed, R., and Sonenberg, N. (2004). A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. *Proc Natl Acad Sci U S A* 101, 4118-4123.

Fischer, N., and Weis, K. (2002). The DEAD box protein Dhh1 stimulates the decapping enzyme Dcp1. *EMBO J* 21, 2788-2797.

Fortes, P., Inada, T., Preiss, T., Hentze, M.W., Mattaj, I.W., and Sachs, A.B. (2000). The yeast nuclear cap binding complex can interact with translation factor eIF4G and mediate translation initiation. *Mol Cell* 6, 191-196.

Franks, T.M., and Lykke-Andersen, J. (2008). The control of mRNA decapping and P-body formation. *Mol Cell* 32, 605-615.

Fribourg, S., Gatfield, D., Izaurralde, E., and Conti, E. (2003). A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat Struct Biol* 10, 433-439.

Frischmeyer, P.A., and Dietz, H.C. (1999). Nonsense-mediated mRNA decay in health and disease. *Hum Mol Genet* 8, 1893-1900.

Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., and Conti, E. (2005). SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. *Mol Cell* 17, 537-547.

Gao, M., Wilusz, C.J., Peltz, S.W., and Wilusz, J. (2001). A novel mRNA-decapping activity in HeLa cytoplasmic extracts is regulated by AU-rich elements. *EMBO J* 20, 1134-1143.

Gao, Q., Das, B., Sherman, F., and Maquat, L.E. (2005). Cap-binding protein 1-mediated and eukaryotic translation initiation factor 4E-mediated pioneer rounds of translation in yeast. *Proc Natl Acad Sci U S A* 102, 4258-4263.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* 8, 113-126.

Gatfield, D., and Izaurralde, E. (2004). Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in *Drosophila*. *Nature* 429, 575-578.

Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. *EMBO J* 22, 3960-3970.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol Cell* 20, 65-75.

Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W., and Kulozik, A.E. (2003). Y14 and hUpf3b form an NMD-activating complex. *Mol Cell* 11, 939-949.

Gertler, F.B., Niebuhr, K., Reinhard, M., Wehland, J., and Soriano, P. (1996). Mena, a relative of VASP and *Drosophila* Enabled, is implicated in the control of microfilament dynamics. *Cell* 87, 227-239.

Gong, C., Kim, Y.K., Woeller, C.F., Tang, Y., and Maquat, L.E. (2009). SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNAs. *Genes Dev* 23, 54-66.

Gong, C., and Maquat, L.E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* 470, 284-288.

Gonzalez, C.I., Ruiz-Echevarria, M.J., Vasudevan, S., Henry, M.F., and Peltz, S.W. (2000). The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. *Mol Cell* 5, 489-499.

Gorlich, D., Kraft, R., Kostka, S., Vogel, F., Hartmann, E., Laskey, R.A., Mattaj, I.W., and Izaurralde, E. (1996). Importin provides a link between nuclear protein import and U snRNA export. *Cell* 87, 21-32.

Greer, W.L., Shehabeldin, A., Schulman, J., Junker, A., and Siminovitch, K.A. (1996). Identification of WASP mutations, mutation hotspots and genotype-phenotype disparities in 24 patients with the Wiskott-Aldrich syndrome. *Hum Genet* 98, 685-690.

Hall, G.W., and Thein, S. (1994). Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: a mechanism for the phenotype of dominant beta-thalassemia. *Blood* 83, 2031-2037.

Hatfield, L., Beelman, C.A., Stevens, A., and Parker, R. (1996). Mutations in trans-acting factors affecting mRNA decapping in *Saccharomyces cerevisiae*. *Mol Cell Biol* 16, 5830-5838.

He, F., Brown, A.H., and Jacobson, A. (1996). Interaction between Nmd2p and Upf1p is required for activity but not for dominant-negative inhibition of the nonsense-mediated mRNA decay pathway in yeast. *RNA* 2, 153-170.

He, F., and Jacobson, A. (1995). Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. *Genes Dev* 9, 437-454.

He, F., and Jacobson, A. (2001). Upf1p, Nmd2p, and Upf3p regulate the decapping and exonucleolytic degradation of both nonsense-containing mRNAs and wild-type mRNAs. *Mol Cell Biol* 21, 1515-1530.

He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., and Jacobson, A. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol Cell* 12, 1439-1452.

Heikkinen, H.L., Llewellyn, S.A., and Barnes, C.A. (2003). Initiation-mediated mRNA decay in yeast affects heat-shock mRNAs, and works through decapping and 5'-to-3' hydrolysis. *Nucleic Acids Res* 31, 4006-4016.

Higgs, H.N., and Pollard, T.D. (1999). Regulation of actin polymerization by Arp2/3 complex and WASp/Scar proteins. *J Biol Chem* 274, 32531-32534.

Hodgkin, J., Papp, A., Pulak, R., Ambros, V., and Anderson, P. (1989). A new kind of informational suppression in the nematode *Caenorhabditis elegans*. *Genetics* 123, 301-313.

Holbrook, J.A., Neu-Yilik, G., Hentze, M.W., and Kulozik, A.E. (2004). Nonsense-mediated decay approaches the clinic. *Nat Genet* 36, 801-808.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J Mol Biol* 233, 123-138.

Hoshino, S., Imai, M., Kobayashi, T., Uchida, N., and Katada, T. (1999). The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3'-Poly(A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein. *J Biol Chem* 274, 16677-16680.

Hosoda, N., Kim, Y.K., Lejeune, F., and Maquat, L.E. (2005). CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. *Nat Struct Mol Biol* 12, 893-901.

Hsu, C.L., and Stevens, A. (1993). Yeast cells lacking 5'→3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Mol Cell Biol* 13, 4826-4835.

Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* 14, 2609-2617.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L.E. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* 106, 607-617.

Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W., and Maquat, L.E. (2008). Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell* 133, 314-327.

Isken, O., and Maquat, L.E. (2007). Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev* 21, 1833-1856.

Iwasaki, S., Takeda, A., Motose, H., and Watanabe, Y. (2007). Characterization of Arabidopsis decapping proteins AtDCP1 and AtDCP2, which are essential for post-embryonic development. *FEBS Lett* 581, 2455-2459.

Izaurralde, E., Lewis, J., Gamberi, C., Jarmolowski, A., McGuigan, C., and Mattaj, I.W. (1995). A cap-binding protein complex mediating U snRNA export. *Nature* 376, 709-712.

Jakymiw, A., Lian, S., Eystathioy, T., Li, S., Satoh, M., Hamel, J.C., Fritzler, M.J., and Chan, E.K. (2005). Disruption of GW bodies impairs mammalian RNA interference. *Nat Cell Biol* 7, 1267-1274.

James, V.H., McNeill, J.M., Lai, L.C., Newton, C.J., Ghilchik, M.W., and Reed, M.J. (1987). Aromatase activity in normal breast and breast tumor tissues: in vivo and in vitro studies. *Steroids* 50, 269-279.

Jankowsky, E., Gross, C.H., Shuman, S., and Pyle, A.M. (2001). Active disruption of an RNA-protein interaction by a DExH/D RNA helicase. *Science* 291, 121-125.

- Jones, B.N., Quang-Dang, D.U., Oku, Y., and Gross, J.D. (2008). A kinetic assay to monitor RNA decapping under single- turnover conditions. *Methods Enzymol* 448, 23-40.
- Jordan, V.C., and Brodie, A.M. (2007). Development and evolution of therapies targeted to the estrogen receptor for the treatment and prevention of breast cancer. *Steroids* 72, 7-25.
- Kadlec, J., Guilligay, D., Ravelli, R.B., and Cusack, S. (2006). Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. *RNA* 12, 1817-1824.
- Kadlec, J., Izaurralde, E., and Cusack, S. (2004). The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat Struct Mol Biol* 11, 330-337.
- Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev* 20, 355-367.
- Kato, A., Ozawa, F., Saitoh, Y., Hirai, K., and Inokuchi, K. (1997). *vesl*, a gene encoding VASP/Ena family related protein, is upregulated during seizure, long-term potentiation and synaptogenesis. *FEBS Lett* 412, 183-189.
- Kaygun, H., and Marzluff, W.F. (2005). Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. *Nat Struct Mol Biol* 12, 794-800.
- Keeling, K.M., Lanier, J., Du, M., Salas-Marco, J., Gao, L., Kaenjak-Angeletti, A., and Bedwell, D.M. (2004). Leaky termination at premature stop codons antagonizes nonsense-mediated mRNA decay in *S. cerevisiae*. *RNA* 10, 691-703.
- Kertesz, S., Kerenyi, Z., Merai, Z., Bartos, I., Palfy, T., Barta, E., and Silhavy, D. (2006). Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucleic Acids Res* 34, 6147-6157.

Khanna, R., and Kiledjian, M. (2004). Poly(A)-binding-protein-mediated regulation of hDcp2 decapping in vitro. *EMBO J* 23, 1968-1976.

Kim, Y.K., Furic, L., Desgroseillers, L., and Maquat, L.E. (2005). Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. *Cell* 120, 195-208.

Kim, Y.K., Furic, L., Parisien, M., Major, F., DesGroseillers, L., and Maquat, L.E. (2007). Staufen1 regulates diverse classes of mammalian transcripts. *EMBO J* 26, 2670-2681.

Kolluri, R., Shehabeldin, A., Peacocke, M., Lamhonwah, A.M., Teichert-Kuliszewska, K., Weissman, S.M., and Siminovitch, K.A. (1995). Identification of WASP mutations in patients with Wiskott-Aldrich syndrome and isolated thrombocytopenia reveals allelic heterogeneity at the WAS locus. *Hum Mol Genet* 4, 1119-1126.

Koonin, E.V. (1992). A new group of putative RNA helicases. *Trends Biochem Sci* 17, 495-497.

Kozlov, G., De Crescenzo, G., Lim, N.S., Siddiqui, N., Fantus, D., Kahvejian, A., Trempe, J.F., Elias, D., Ekiel, I., Sonenberg, N., et al. (2004). Structural basis of ligand recognition by PABC, a highly specific peptide-binding domain found in poly(A)-binding protein and a HECT ubiquitin ligase. *EMBO J* 23, 272-281.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440, 637-643.

Kshirsagar, M., and Parker, R. (2004). Identification of Edc3p as an enhancer of mRNA decapping in *Saccharomyces cerevisiae*. *Genetics* 166, 729-739.

Kufel, J., Bousquet-Antonelli, C., Beggs, J.D., and Tollervey, D. (2004). Nuclear pre-mRNA decapping and 5' degradation in yeast require the Lsm2-8p complex. *Mol Cell Biol* 24, 9646-9657.

Kunz, J.B., Neu-Yilik, G., Hentze, M.W., Kulozik, A.E., and Gehring, N.H. (2006). Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. *RNA* 12, 1015-1022.

Kuzmiak, H.A., and Maquat, L.E. (2006). Applying nonsense-mediated mRNA decay research to the clinic: progress and challenges. *Trends Mol Med* 12, 306-316.

LaGrandeur, T.E., and Parker, R. (1998). Isolation and characterization of Dcp1p, the yeast mRNA decapping enzyme. *EMBO J* 17, 1487-1496.

Lai, P.F., Mohamed, F., Monge, J.C., and Stewart, D.J. (2003). Downregulation of eNOS mRNA expression by TNFalpha: identification and functional characterization of RNA-protein interactions in the 3'UTR. *Cardiovasc Res* 59, 160-168.

Lanier, L.M., Gates, M.A., Witke, W., Menzies, A.S., Wehman, A.M., Macklis, J.D., Kwiatkowski, D., Soriano, P., and Gertler, F.B. (1999). Mena is required for neurulation and commissure formation. *Neuron* 22, 313-325.

Laurent, V., Loisel, T.P., Harbeck, B., Wehman, A., Grobe, L., Jockusch, B.M., Wehland, J., Gertler, F.B., and Carlier, M.F. (1999). Role of proteins of the Ena/VASP family in actin-based motility of *Listeria monocytogenes*. *J Cell Biol* 144, 1245-1258.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J* 19, 6860-6869.

Leeds, P., Peltz, S.W., Jacobson, A., and Culbertson, M.R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev* 5, 2303-2314.

Leeds, P., Wood, J.M., Lee, B.S., and Culbertson, M.R. (1992). Gene products that promote mRNA turnover in *Saccharomyces cerevisiae*. *Mol Cell Biol* 12, 2165-2177.

Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L.E. (2002). The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J* 21, 3536-3545.

Lejeune, F., Li, X., and Maquat, L.E. (2003). Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. *Mol Cell* 12, 675-687.

Lejeune, F., and Maquat, L.E. (2005). Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol* 17, 309-315.

Lelivelt, M.J., and Culbertson, M.R. (1999). Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol Cell Biol* 19, 6710-6719.

Leonard, D., Ajuh, P., Lamond, A.I., and Legerski, R.J. (2003). hLodestar/HuF2 interacts with CDC5L and is involved in pre-mRNA splicing. *Biochem Biophys Res Commun* 308, 793-801.

Leonardi, A., Chariot, A., Claudio, E., Cunningham, K., and Siebenlist, U. (2000). CIKS, a connection to I κ B kinase and stress-activated protein kinase. *Proc Natl Acad Sci USA* 97, 10494-10499.

Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci USA* 100, 189-192.

Lewis, J.D., and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. *Eur J Biochem* 247, 461-469.

Li, R. (1997). Bee1, a yeast protein with homology to Wiscott-Aldrich syndrome protein, is critical for the assembly of cortical actin cytoskeleton. *J Cell Biol* 136, 649-658.

Lin, M.D., Jiao, X., Grima, D., Newbury, S.F., Kiledjian, M., and Chou, T.B. (2008). *Drosophila* processing bodies in oogenesis. *Dev Biol* 322, 276-288.

Liu, H., and Kiledjian, M. (2006). Decapping the message: a beginning or an end. *Biochem Soc Trans* 34, 35-38.

Liu, H., Rodgers, N.D., Jiao, X., and Kiledjian, M. (2002). The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. *EMBO J* 21, 4699-4708.

Liu, J., and Doetsch, P.W. (1998). Escherichia coli RNA and DNA polymerase bypass of dihydrouracil: mutagenic potential via transcription and replication. *Nucleic Acids Res* 26, 1707-1712.

Liu, J., Rivas, F.V., Wohlschlegel, J., Yates, J.R., 3rd, Parker, R., and Hannon, G.J. (2005). A role for the P-body component GW182 in microRNA function. *Nat Cell Biol* 7, 1261-1266.

Lu, Q., Nakamura, J., Savinov, A., Yue, W., Weisz, J., Dabbs, D.J., Wolz, G., and Brodie, A. (1996). Expression of aromatase protein and messenger ribonucleic acid in tumor epithelial cells and evidence of functional significance of locally produced estrogen in human breast cancers. *Endocrinology* 137, 3061-3068.

Lykke-Andersen, J. (2002). Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. *Mol Cell Biol* 22, 8114-8121.

Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* 293, 1836-1839.

Lykke-Andersen, J., and Wagner, E. (2005). Recruitment and activation of mRNA decay enzymes by two ARE-mediated decay activation domains in the proteins TTP and BRF-1. *Genes Dev* 19, 351-361.

Machesky, L.M., Reeves, E., Wientjes, F., Mattheyse, F.J., Grogan, A., Totty, N.F., Burlingame, A.L., Hsuan, J.J., and Segal, A.W. (1997). Mammalian actin-related protein 2/3 complex localizes to regions of lamellipodial protrusion and is composed of evolutionarily conserved proteins. *Biochem J* 328 (Pt 1), 105-112.

Maderazo, A.B., Belk, J.P., He, F., and Jacobson, A. (2003). Nonsense-containing mRNAs that accumulate in the absence of a functional nonsense-mediated mRNA decay pathway are destabilized rapidly upon its restitution. *Mol Cell Biol* 23, 842-851.

Maquat, L.E., and Li, X. (2001). Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. *RNA* 7, 445-456.

Mauro, C., Vito, P., Mellone, S., Pacifico, F., Chariot, A., Formisano, S., and Leonardi, A. (2003). Role of the adaptor protein CIKS in the activation of the IKK complex. *Biochem Biophys Res Commun* 309, 84-90.

McLennan, A.G. (2006). The Nudix hydrolase superfamily. *Cell Mol Life Sci* 63, 123-143.

Medghalchi, S.M., Frischmeyer, P.A., Mendell, J.T., Kelly, A.G., Lawler, A.M., and Dietz, H.C. (2001). Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum Mol Genet* 10, 99-105.

Mendell, J.T., ap Rhys, C.M., and Dietz, H.C. (2002). Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts. *Science* 298, 419-422.

Mendell, J.T., Medghalchi, S.M., Lake, R.G., Noensie, E.N., and Dietz, H.C. (2000). Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes. *Mol Cell Biol* 20, 8944-8957.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* 36, 1073-1078.

Mildvan, A.S., Xia, Z., Azurmendi, H.F., Saraswat, V., Legler, P.M., Massiah, M.A., Gabelli, S.B., Bianchet, M.A., Kang, L.W., and Amzel, L.M. (2005). Structures and mechanisms of Nudix hydrolases. *Arch Biochem Biophys* 433, 129-143.

Millard, T.H., and Machesky, L.M. (2001). The Wiskott-Aldrich syndrome protein (WASP) family. Trends Biochem Sci 26, 198-199.

Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). SnB: crystal structure determination via shake-and-bake. J Appl Cryst 27, 613-621.

Miller, W.R. (1996). Aromatase inhibitors--where are we now? Br J Cancer 73, 415-417.

Mitchell, P., and Tollervey, D. (2003). An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. Mol Cell 11, 1405-1413.

Mitrovich, Q.M., and Anderson, P. (2000). Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in *C. elegans*. Genes Dev 14, 2173-2184.

Morrison, M., Harris, K.S., and Roth, M.B. (1997). smg mutants affect the expression of alternatively spliced SR protein mRNAs in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 94, 9782-9785.

Moser, M.J., Holley, W.R., Chatterjee, A., and Mian, I.S. (1997). The proofreading domain of *Escherichia coli* DNA polymerase I and other DNA and/or RNA exonuclease domains. Nucleic Acids Res 25, 5110-5118.

Muhlemann, O., Eberle, A.B., Stalder, L., and Zamudio Orozco, R. (2008). Recognition and elimination of nonsense mRNA. Biochim Biophys Acta 1779, 538-549.

Muhlrad, D., Decker, C.J., and Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. Genes Dev 8, 855-866.

Muhlrad, D., Decker, C.J., and Parker, R. (1995). Turnover mechanisms of the stable yeast PGK1 mRNA. Mol Cell Biol 15, 2145-2156.

Muhlrad, D., and Parker, R. (1994). Premature translational termination triggers mRNA decapping. *Nature* 370, 578-581.

Muhlrad, D., and Parker, R. (1999). Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* 5, 1299-1307.

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 Dev* 22, 50-65.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53, 240-255.

Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* 23, 198-199.

Neidhardt, F.C., Bloch, P.L., and Smith, D.F. (1974). Culture medium for enterobacteria. *J Bacteriol* 119, 736-747.

Neu-Yilik, G., Gehring, N.H., Thermann, R., Frede, U., Hentze, M.W., and Kulozik, A.E. (2001). Splicing and 3' end formation in the definition of nonsense-mediated decay-competent human beta-globin mRNPs. *EMBO J* 20, 532-540.

Niebuhr, K., Ebel, F., Frank, R., Reinhard, M., Domann, E., Carl, U.D., Walter, U., Gertler, F.B., Wehland, J., and Chakraborty, T. (1997). A novel proline-rich motif present in ActA of *Listeria monocytogenes* and cytoskeletal proteins is the ligand for the EVH1 domain, a protein module present in the Ena/VASP family. *EMBO J* 16, 5433-5444.

Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K.R., Grimson, A., Hachiya, T., Hentze, M.W., Anderson, P., and Ohno, S. (2003). Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. *Mol Cell* 12, 1187-1200.

Page, M.F., Carr, B., Anders, K.R., Grimson, A., and Anderson, P. (1999). SMG-2 is a phosphorylated protein required for mRNA surveillance in *Caenorhabditis elegans* and related to Upf1p of yeast. *Mol Cell Biol* 19, 5943-5951.

Painter, J., and Merritt, E.A. (2006a). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62, 439-450.

Painter, J., and Merritt, E.A. (2006b). TLSMD web server for the generation of multi-group TLS models. *Journal of Applied Crystallography* 39, 109-111.

Pal, M., Ishigaki, Y., Nagy, E., and Maquat, L.E. (2001). Evidence that phosphorylation of human Upf1 protein varies with intracellular location and is mediated by a wortmannin-sensitive and rapamycin-sensitive PI 3-kinase-related kinase signaling pathway. *RNA* 7, 5-15.

Palacios, I.M., Gatfield, D., St Johnston, D., and Izaurralde, E. (2004). An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature* 427, 753-757.

Parker, R., and Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol* 11, 121-127.

Peltz, S.W., Donahue, J.L., and Jacobson, A. (1992). A mutation in the tRNA nucleotidyltransferase gene promotes stabilization of mRNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol* 12, 5778-5784.

Perrakis, A., Harkiolaki, M., Wilson, K.S., and Lamzin, V.S. (2001). ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr* 57, 1445-1450.

Petrella, E.C., Machesky, L.M., Kaiser, D.A., and Pollard, T.D. (1996). Structural requirements and thermodynamics of the interaction of proline peptides with profilin. *Biochemistry* 35, 16535-16543.

Piccirillo, C., Khanna, R., and Kiledjian, M. (2003). Functional characterization of the mammalian mRNA decapping enzyme hDcp2. *RNA* 9, 1138-1147.

Pilkington, G.R., and Parker, R. (2008). Pat1 contains distinct functional domains that promote P-body assembly and activation of decapping. *Mol Cell Biol* 28, 1298-1312.

Pollard, T.D. (1995). Actin cytoskeleton. Missing link for intracellular bacterial motility? *Curr Biol* 5, 837-840.

Pollard, T.D., Blanchoin, L., and Mullins, R.D. (2000). Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annu Rev Biophys Biomol Struct* 29, 545-576.

Prehoda, K.E., Lee, D.J., and Lim, W.A. (1999). Structure of the enabled/VASP homology 1 domain-peptide complex: a key component in the spatial control of actin assembly. *Cell* 97, 471-480.

Pulak, R., and Anderson, P. (1993). mRNA surveillance by the *Caenorhabditis elegans* smg genes. *Genes Dev* 7, 1885-1897.

Rajhans, R., Nair, H.B., Nair, S.S., Cortez, V., Ikuko, K., Kirma, N.B., Zhou, D., Holden, A.E., Brann, D.W., Chen, S., et al. (2008). Modulation of in situ estrogen synthesis by proline-, glutamic acid-, and leucine-rich protein-1: potential estrogen receptor autocrine signaling loop in breast cancer cells. *Mol Endocrinol* 22, 649-664.

Rehwinkel, J., Letunic, I., Raes, J., Bork, P., and Izaurralde, E. (2005). Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA* 11, 1530-1544.

Rehwinkel, J., Raes, J., and Izaurralde, E. (2006). Nonsense-mediated mRNA decay: Target genes and functional diversification of effectors. *Trends Biochem Sci* 31, 639-646.

Reichenbach, P., Hoss, M., Azzalin, C.M., Nabholz, M., Bucher, P., and Lingner, J. (2003). A human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed. *Curr Biol* 13, 568-574.

- Reijns, M.A., Alexander, R.D., Spiller, M.P., and Beggs, J.D. (2008). A role for Q/N-rich aggregation-prone regions in P-body localization. *J Cell Sci* 121, 2463-2472.
- Richter, J.D., and Sonenberg, N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* 433, 477-480.
- Rissland, O.S., and Norbury, C.J. (2009). Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. *Nat Struct Mol Biol* 16, 616-623.
- Rohatgi, R., Nollau, P., Ho, H.Y., Kirschner, M.W., and Mayer, B.J. (2001). Nck and phosphatidylinositol 4,5-bisphosphate synergistically activate actin polymerization through the N-WASP-Arp2/3 pathway. *J Biol Chem* 276, 26448-26452.
- Romao, L., Inacio, A., Santos, S., Avila, M., Faustino, P., Pacheco, P., and Lavinha, J. (2000). Nonsense mutations in the human beta-globin gene lead to unexpected levels of cytoplasmic mRNA accumulation. *Blood* 96, 2895-2901.
- Rosen, E.D., and MacDougald, O.A. (2006). Adipocyte differentiation from the inside out. *Nat Rev Mol Cell Biol* 7, 885-896.
- Ruiz-Echevarria, M.J., and Peltz, S.W. (2000). The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. *Cell* 101, 741-751.
- Rzeczkowski, K., Beuerlein, K., Muller, H., Dittrich-Breiholz, O., Schneider, H., Kettner-Buhrow, D., Holtmann, H., and Kracht, M. (2011). c-Jun N-terminal kinase phosphorylates DCP1a to control formation of P bodies. *J Cell Biol* 194, 581-596.
- Sachs, A.B., Davis, R.W., and Kornberg, R.D. (1987). A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. *Mol Cell Biol* 7, 3268-3276.

Sakuno, T., Araki, Y., Ohya, Y., Kofuji, S., Takahashi, S., Hoshino, S., and Katada, T. (2004). Decapping reaction of mRNA requires Dcp1 in fission yeast: its characterization in different species from yeast to human. *J Biochem* 136, 805-812.

Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S., and Seraphin, B. (1999). Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J* 18, 3451-3462.

Schwartz, D., Decker, C.J., and Parker, R. (2003). The enhancer of decapping proteins, Edc1p and Edc2p, bind RNA and stimulate the activity of the decapping enzyme. *RNA* 9, 239-251.

Schwartz, D.C., and Parker, R. (1999). Mutations in translation initiation factors lead to increased rates of deadenylation and decapping of mRNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol* 19, 5247-5256.

Schwartz, D.C., and Parker, R. (2000). mRNA decapping in yeast requires dissociation of the cap binding protein, eukaryotic translation initiation factor 4E. *Mol Cell Biol* 20, 7933-7942.

Serin, G., Gersappe, A., Black, J.D., Aronoff, R., and Maquat, L.E. (2001). Identification and characterization of human orthologues to *Saccharomyces cerevisiae* Upf2 protein and Upf3 protein (*Caenorhabditis elegans* SMG-4). *Mol Cell Biol* 21, 209-223.

She, M., Decker, C.J., Chen, N., Tumati, S., Parker, R., and Song, H. (2006). Crystal structure and functional analysis of Dcp2p from *Schizosaccharomyces pombe*. *Nat Struct Mol Biol* 13, 63-70.

She, M., Decker, C.J., Sundramurthy, K., Liu, Y., Chen, N., Parker, R., and Song, H. (2004). Crystal structure of Dcp1p and its functional implications in mRNA decapping. *Nat Struct Mol Biol* 11, 249-256.

She, M., Decker, C.J., Svergun, D.I., Round, A., Chen, N., Muhlrads, D., Parker, R., and Song, H. (2008). Structural basis of dcp2 recognition and activation by dcp1. *Mol Cell* 29, 337-349.

Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* 300, 805-808.

Sheth, U., and Parker, R. (2006). Targeting of aberrant mRNAs to cytoplasmic processing bodies. *Cell* 125, 1095-1109.

Shibuya, T., Tange, T.O., Sonenberg, N., and Moore, M.J. (2004). eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol* 11, 346-351.

Shibuya, T., Tange, T.O., Stroupe, M.E., and Moore, M.J. (2006). Mutational analysis of human eIF4AIII identifies regions necessary for exon junction complex formation and nonsense-mediated mRNA decay. *RNA* 12, 360-374.

Silva, A.L., Pereira, F.J., Morgado, A., Kong, J., Martins, R., Faustino, P., Liebhaber, S.A., and Romao, L. (2006). The canonical UPF1-dependent nonsense-mediated mRNA decay is inhibited in transcripts carrying a short open reading frame independent of sequence context. *RNA* 12, 2160-2170.

Simon, E., Camier, S., and Seraphin, B. (2006). New insights into the control of mRNA decapping. *Trends Biochem Sci* 31, 241-243.

Singh, G., Jakob, S., Kleedehn, M.G., and Lykke-Andersen, J. (2007). Communication with the exon-junction complex and activation of nonsense-mediated decay by human Upf proteins occur in the cytoplasm. *Mol Cell* 27, 780-792.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol* 6, e111.

Snapper, S.B., Takeshima, F., Anton, I., Liu, C.H., Thomas, S.M., Nguyen, D., Dudley, D., Fraser, H., Purich, D., Lopez-Illasaca, M., et al. (2001). N-WASP deficiency reveals distinct pathways for cell surface projections and microbial actin-based motility. *Nat Cell Biol* 3, 897-904.

- Song, M.G., and Kiledjian, M. (2007). 3' Terminal oligo U-tract-mediated stimulation of decapping. *Rna* 13, 2356-2365.
- Song, M.G., Li, Y., and Kiledjian, M. (2010). Multiple mRNA decapping enzymes in mammalian cells. *Mol Cell* 40, 423-432.
- Stalder, L., and Muhlemann, O. (2008). The meaning of nonsense. *Trends Cell Biol* 18, 315-321.
- Stalder, L., and Muhlemann, O. (2009). Processing bodies are not required for mammalian nonsense-mediated mRNA decay. *RNA* 15, 1265-1273.
- Steiger, M., Carr-Schmid, A., Schwartz, D.C., Kiledjian, M., and Parker, R. (2003). Analysis of recombinant yeast decapping enzyme. *Rna* 9, 231-238.
- Stevens, A. (1980). Purification and characterization of a *Saccharomyces cerevisiae* exoribonuclease which yields 5'-mononucleotides by a 5' leads to 3' mode of hydrolysis. *J Biol Chem* 255, 3080-3085.
- Stoecklin, G., Mayo, T., and Anderson, P. (2006). ARE-mRNA degradation requires the 5'-3' decay pathway. *EMBO Rep* 7, 72-77.
- Sun, X., Moriarty, P.M., and Maquat, L.E. (2000). Nonsense-mediated decay of glutathione peroxidase 1 mRNA in the cytoplasm depends on intron position. *EMBO J* 19, 4734-4744.
- Sun, X., Perlick, H.A., Dietz, H.C., and Maquat, L.E. (1998). A mutated human homologue to yeast Upf1 protein has a dominant-negative effect on the decay of nonsense-containing mRNAs in mammalian cells. *Proc Natl Acad Sci U S A* 95, 10009-10014.
- Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J., and Soret, J. (2001). SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J* 20, 1785-1796.

Symmons, M.F., Williams, M.G., Luisi, B.F., Jones, G.H., and Carpousis, A.J. (2002). Running rings around RNA: a superfamily of phosphate-dependent RNases. *Trends Biochem Sci* 27, 11-18.

Symons, M., Derry, J.M., Karlak, B., Jiang, S., Lemahieu, V., McCormick, F., Francke, U., and Abo, A. (1996). Wiskott-Aldrich syndrome protein, a novel effector for the GTPase CDC42Hs, is implicated in actin polymerization. *Cell* 84, 723-734.

Tarun, S.Z., Jr., and Sachs, A.B. (1996). Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J* 15, 7168-7177.

Taylor, M.J., and Peculis, B.A. (2008). Evolutionary conservation supports ancient origin for Nudt16, a nuclear-localized, RNA-binding, RNA-decapping enzyme. *Nucleic Acids Res* 36, 6021-6034.

Teixeira, D., and Parker, R. (2007). Analysis of P-body assembly in *Saccharomyces cerevisiae*. *Mol Biol Cell* 18, 2274-2287.

Teixeira, D., Sheth, U., Valencia-Sanchez, M.A., Brengues, M., and Parker, R. (2005). Processing bodies require RNA for assembly and contain nontranslating mRNAs. *RNA* 11, 371-382.

Tharun, S. (2009). Roles of eukaryotic Lsm proteins in the regulation of mRNA function. *Int Rev Cell Mol Biol* 272, 149-189.

Tharun, S., He, W., Mayes, A.E., Lennertz, P., Beggs, J.D., and Parker, R. (2000). Yeast Sm-like proteins function in mRNA decapping and decay. *Nature* 404, 515-518.

Tharun, S., and Parker, R. (1999). Analysis of mutations in the yeast mRNA decapping enzyme. *Genetics* 151, 1273-1285.

Tharun, S., and Parker, R. (2001). Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol Cell* 8, 1075-1083.

Theriot, J.A. (1994). Actin filament dynamics in cell motility. *Adv Exp Med Biol* 358, 133-145.

Thore, S., Mauxion, F., Seraphin, B., and Suck, D. (2003). X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. *EMBO Rep* 4, 1150-1155.

Tilney, L.G., and Portnoy, D.A. (1989). Actin filaments and the growth, movement, and spread of the intracellular bacterial parasite, *Listeria monocytogenes*. *J Cell Biol* 109, 1597-1608.

Trcek, T., Larson, D.R., Moldon, A., Query, C.C., and Singer, R.H. (2011). Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. *Cell* 147, 1484-1497.

Tritschler, F., Braun, J.E., Motz, C., Igreja, C., Haas, G., Truffault, V., Izaurralde, E., and Weichenrieder, O. (2009). DCP1 forms asymmetric trimers to assemble into active mRNA decapping complexes in metazoa. *Proc Natl Acad Sci U S A*.

Tritschler, F., Eulalio, A., Truffault, V., Hartmann, M.D., Helms, S., Schmidt, S., Coles, M., Izaurralde, E., and Weichenrieder, O. (2007). A divergent Sm fold in EDC3 proteins mediates DCP1 binding and P-body targeting. *Mol Cell Biol* 27, 8600-8611.

Troyanovsky, B., Levchenko, T., Mansson, G., Matvijenko, O., and Holmgren, L. (2001). Angiomotin: an angiostatin binding protein that regulates endothelial cell migration and tube formation. *J Cell Biol* 152, 1247-1254.

Tu, J.C., Xiao, B., Yuan, J.P., Lanahan, A.A., Leoffert, K., Li, M., Linden, D.J., and Worley, P.F. (1998). Homer binds a novel proline-rich motif and links group 1 metabotropic glutamate receptors with IP3 receptors. *Neuron* 21, 717-726.

Tucker, M., Valencia-Sanchez, M.A., Staples, R.R., Chen, J., Denis, C.L., and Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell* 104, 377-386.

Unterholzner, L., and Izaurralde, E. (2004). SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. *Mol Cell* 16, 587-596.

Usuki, F., Yamashita, A., Higuchi, I., Ohnishi, T., Shiraishi, T., Osame, M., and Ohno, S. (2004). Inhibition of nonsense-mediated mRNA decay rescues the phenotype in Ullrich's disease. *Ann Neurol* 55, 740-744.

van Dijk, E., Cougot, N., Meyer, S., Babajko, S., Wahle, E., and Seraphin, B. (2002). Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J* 21, 6915-6924.

van Dijk, E., Le Hir, H., and Seraphin, B. (2003). DcpS can act in the 5'-3' mRNA decay pathway in addition to the 3'-5' pathway. *Proc Natl Acad Sci USA* 100, 12081-12086.

Villa, A., Notarangelo, L., Macchi, P., Mantuano, E., Cavagni, G., Brugnani, D., Strina, D., Patrosso, M.C., Ramenghi, U., Sacco, M.G., et al. (1995). X-linked thrombocytopenia and Wiskott-Aldrich syndrome are allelic diseases with mutations in the WASP gene. *Nat Genet* 9, 414-417.

Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B., and Mattaj, I.W. (1996). A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. *J Cell Biol* 133, 5-14.

Volkman, B.F., Prehoda, K.E., Scott, J.A., Peterson, F.C., and Lim, W.A. (2002). Structure of the N-WASP EVH1 domain-WIP complex: insight into the molecular basis of Wiskott-Aldrich Syndrome. *Cell* 111, 565-576.

Wang, J., Vock, V.M., Li, S., Olivas, O.R., and Wilkinson, M.F. (2002a). A quality control pathway that down-regulates aberrant T-cell receptor (TCR) transcripts by a mechanism requiring UPF2 and translation. *J Biol Chem* 277, 18489-18493.

Wang, W., Cajigas, I.J., Peltz, S.W., Wilkinson, M.F., and Gonzalez, C.I. (2006). Role for Upf2p phosphorylation in *Saccharomyces cerevisiae* nonsense-mediated mRNA decay. *Mol Cell Biol* 26, 3390-3400.

Wang, Z., Jiao, X., Carr-Schmid, A., and Kiledjian, M. (2002b). The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc Natl Acad Sci U S A* 99, 12663-12668.

Wang, Z., and Kiledjian, M. (2001). Functional link between the mammalian exosome and mRNA decapping. *Cell* 107, 751-762.

Weischenfeldt, J., Damgaard, I., Bryder, D., Theilgaard-Monch, K., Thoren, L.A., Nielsen, F.C., Jacobsen, S.E., Nerlov, C., and Porse, B.T. (2008). NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev* 22, 1381-1396.

Welch, E.M., Barton, E.R., Zhuo, J., Tomizawa, Y., Friesen, W.J., Trifillis, P., Paushkin, S., Patel, M., Trotta, C.R., Hwang, S., et al. (2007). PTC124 targets genetic disorders caused by nonsense mutations. *Nature* 447, 87-91.

Weng, Y., Czaplinski, K., and Peltz, S.W. (1996). Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Mol Cell Biol* 16, 5477-5490.

Wichroski, M.J., Robb, G.B., and Rana, T.M. (2006). Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies. *PLoS Pathog* 2, e41.

Wilschanski, M., Yahav, Y., Yaacov, Y., Blau, H., Bentur, L., Rivlin, J., Aviram, M., Bdolah-Abram, T., Bebok, Z., Shushi, L., et al. (2003). Gentamicin-induced correction of CFTR function in patients with cystic fibrosis and CFTR stop mutations. *N Engl J Med* 349, 1433-1441.

Wilusz, C.J., Gao, M., Jones, C.L., Wilusz, J., and Peltz, S.W. (2001). Poly(A)-binding proteins regulate both mRNA deadenylation and decapping in yeast cytoplasmic extracts. *RNA* 7, 1416-1424.

Wilusz, C.J., and Wilusz, J. (2004). Bringing the role of mRNA decay in the control of gene expression into focus. *Trends Genet* 20, 491-497.

Wittmann, J., Hol, E.M., and Jack, H.M. (2006). hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol Cell Biol* 26, 1272-1287.

Wu, J., Srinivasan, S.V., Neumann, J.C., and Lingrel, J.B. (2005). The KLF2 transcription factor does not affect the formation of preadipocytes but inhibits their differentiation into adipocytes. *Biochemistry* 44, 11098-11105.

Xu, J., Yang, J.Y., Niu, Q.W., and Chua, N.H. (2006). Arabidopsis DCP2, DCP1, and VARICOSE form a decapping complex required for postembryonic development. *Plant Cell* 18, 3386-3398.

Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., and Ohno, S. (2001). Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev* 15, 2215-2228.

Yu, J.H., Yang, W.H., Gulick, T., Bloch, K.D., and Bloch, D.B. (2005). Ge-1 is a central component of the mammalian cytoplasmic mRNA processing body. *RNA* 11, 1795-1802.

Zhou, D., Chen, B., Ye, J.J., and Chen, S. (2004). A novel crosstalk mechanism between nuclear receptor-mediated and growth factor/Ras-mediated pathways through PNRC-Grb2 interaction. *Oncogene* 23, 5394-5404.

Zhou, D., and Chen, S. (2001). PNRC2 is a 16 kDa coactivator that interacts with nuclear receptors through an SH3-binding motif. *Nucleic Acids Res* 29, 3939-3948.

Zhou, D., Quach, K.M., Yang, C., Lee, S.Y., Pohajdak, B., and Chen, S. (2000). PNRC: a proline-rich nuclear receptor coregulatory protein that modulates transcriptional activation of multiple nuclear receptors including orphan receptors SF1 (steroidogenic factor 1) and ERRalpha1 (estrogen related receptor alpha-1). *Mol Endocrinol* 14, 986-998.

Zhou, D., Shen, R., Ye, J.J., Li, Y., Tsark, W., Isbell, D., Tso, P., and Chen, S. (2008). Nuclear receptor coactivator PNRC2 regulates energy expenditure and adiposity. *J Biol Chem* 283, 541-553.

Zhou, D., Ye, J.J., Li, Y., Lui, K., and Chen, S. (2006). The molecular basis of the interaction between the proline-rich SH3-binding motif of PNRC and estrogen receptor alpha. *Nucleic Acids Res* 34, 5974-5986.

Zhu, Q., Watanabe, C., Liu, T., Hollenbaugh, D., Blaese, R.M., Kanner, S.B., Aruffo, A., and Ochs, H.D. (1997). Wiskott-Aldrich syndrome/X-linked thrombocytopenia: WASP gene mutations, protein expression, and phenotype. *Blood* 90, 2680-2689.

Zicha, D., Allen, W.E., Brickell, P.M., Kinnon, C., Dunn, G.A., Jones, G.E., and Thrasher, A.J. (1998). Chemotaxis of macrophages is abolished in the Wiskott-Aldrich syndrome. *Br J Haematol* 101, 659-665.