# Biochemical characterisation of the human Ccr4-Not complex: core complex assembly and deadenylation activity

**Lorenzo Pavanello**

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

**March 2018**

**Abstract**

Targeted degradation of cytoplasmic mRNA is of fundamental importance for regulated gene expression in eukaryotic cells. The shortening and removal of the poly(A) tail (deadenylation) is the initial and often rate-limiting step in this process. Two key components in deadenylation are the PAN2-PAN3 and the Ccr4-Not complexes. PAN2-PAN3 is a trimer composed by two regulatory subunits (PAN3) and a catalytic enzyme (PAN2). The Ccr4-not complex is composed of at least eight subunits, constituting four modules, that bind the scaffold protein CNOT1: the N-terminal module (CNOT10, CNOT11); the catalytic nuclease module (Caf1, Ccr4); the CNOT9 module (CNOT9); and the NOT-module (CNOT2, CNOT3). A cellular and biochemical characterisation of the PAN2-PAN3 complex was planned, to further understand its role in eukaryotic cells and determine the deadenylation rates. Similarly, we have started to reconstitute the human Ccr4-Not to understand the contributions of the individual modules to the activity of the complex. Initially, we purified a recombinant nuclease module composed of Caf1, Ccr4, and the anti-proliferative protein BTG2. Next, we reconstituted a core complex including the nuclease module, CNOT9 and a segment of CNOT1 encompassing the MIF4G and DUF3819 domains (CNOT1M). Then, we compared the activity of the core complex and the nuclease module using substrates containing short ($A_9$) and longer ($A_{20}$ or $A_{50}$) poly(A) tails, as well as poly(A) tails coated with the poly(A)-binding protein PABPC1. We observed that the CNOT9 module contributes to enhanced deadenylation and increased selectivity towards terminal adenosine residues. Finally, the influence of the melanoma-associated P131L amino acid substitution of CNOT9 was investigated with structural and biochemical approaches.

**Acknowledgements**

I would like to thank my supervisor Dr Sebastiaan Winkler for the scientific discussions, the constant support and the inspiration I received during my PhD studies. I would like to thank Dr Ingrid Dreveny and Emma Gunnell for the great help received with protein crystallography; Dr Hilary Collins for the immunocytochemistry experiments and her expertise with confocal microscopy; Max Ziemann and Prof Peter Fischer for interesting discussions about drug-like molecules modelling and synthesis; Dr Alexey Potapov for discussions about EPR and potential future outlooks. I would like to thank all the members of the Winkler group for the support, training and guidance in the lab, as well as all the people from the Gene Regulation and RNA Biology group, and in particular Bismoy and Barbara, for making the lab a nice place where to work.

I would like to thank all the people working in CBS that I can proudly call friends, too many to fit in one page; your presence and smiles have helped me a lot more than words can describe. A special thank goes to my favourite housemate, gym and joke buddy Edu, one of the most special people I have ever met, in every sense. I would like to say thank you to Fabio and Elisabetta for endless and entertaining discussions about life and studies over legendary coffees. A big thank to Alistair and Agathe for being there when I needed help and for opening their house in a very welcoming way. Thanks to Berto, Marco, Tia and Albe and all my friends in Italy, who gave me an external prospective over the PhD life.

I want to thank my family, mamma, papa', Marta e Luca, Matteo e Roberta, Giorgia and the newborn Gaia. Your support was really important to me and coming back home was very special every time. Finally, and most importantly, I need to thank Sophie. I simply cannot explain how much you meant for me during the last three years. Your love, support and kindness were fundamental for making this journey full of memories that will make me smile forever.

## Table of contents

**Abstract** ..... **ii**

**Acknowledgements** ..... **iii**

**Table of contents** ..... **iv**

**List of figures** ..... **xiii**

**List of tables** ..... **xviii**

**List of abbreviations** ..... **xix**

**Chapter 1. Introduction** ..... **2**

**1.1 mRNA biogenesis and export** ..... **2**

**1.2 Overview of the pathways for cytoplasmic mRNA decay** ..... **4**

**1.3 Deadenylation** ..... **5**

**1.4 PAN2-PAN3 complex** ..... **10**

*1.4.1 The catalytically active enzyme PAN2* ..... 10

*1.4.2 The regulatory subunit PAN3* ..... 13

*1.3.3 Generic recruitment mechanisms* ..... 13

*1.4.4 Targeted recruitment mechanisms* ..... 16

**1.5 Ccr4-Not complex** ..... **17**

*1.5.1 The N-terminal module* ..... 19

*1.5.2 The Nuclease module* ..... 19

*1.5.3 The CNOT9 module* ..... 22

*1.5.4 The NOT module* ..... 24

*1.5.5 Recruitment of the Ccr4-not complex* ..... 26

*1.5.6 Recruitment via PABPC1* ..... 26

*1.5.7 miRNA-mediated recruitment* ..... 29

*1.5.8 Protein-mediated recruitment* ..... 29

**1.6 Decapping ..... 34**

*1.6.1 The Lsm1-7-Pat1 complex* ..... 34

*1.6.2 The Dcp1-Dcp2 complex* ..... 37

**1.7 Degradation ..... 40**

*1.7.1 XRN1* ..... 40

*1.7.2 Exosome* ..... 41

**1.8 Aims of the study ..... 45**

*1.8.1 Objectives* ..... 45

*1.8.2 Specific aims* ..... 46

*1.8.3 Experimental approaches* ..... 47

**Chapter 2. Materials and methods ..... 49**

**2.1 Bacterial growth and transformation ..... 49**

*2.1.1 Reagents, stock solutions and buffers* ..... 49

*2.1.2 Culture of* Escherichia coli ..... 49

*2.1.3 Preparation of* Escherichia coli *competent cells* ..... 50

*2.1.4 Transformation of competent cells* ..... 50

**2.2 Molecular biology ..... 51**

2.2.1 Reagents, stock solutions and buffers ..... 51

*2.2.2 Small scale DNA plasmid preparation* ..... 51

*2.2.3 Large scale plasmid DNA preparation* ..... 51

*2.2.4 Determination of DNA/RNA concentration* ..... 52

*2.2.5 Agarose gel electrophoresis* ..... 52

*2.2.6 DNA extraction and purification from agarose gel* ..... 52

*2.2.7 Annealing of DNA oligo* ..... 53

*2.2.8 Restriction enzyme digestion of DNA* ..... 53

*2.2.9 Removal of 5' phosphate group from digested plasmid DNA* ..... 53

*2.2.10 Ligation of DNA fragments* ..... 53

*2.2.11 Polymerase Chain Reaction (PCR)* ..... 54

*2.2.12 Site-directed mutagenesis reaction* ..... 55

**2.3 Yeast two-hybrid analysis ..... 60**

*2.3.1 Reagents, stock solutions and buffers for use in yeast culture* ..... 60

*2.3.2 Culture of* Saccharomyces cerevisiae *strain YRG2* ..... 60

*2.3.3 Transformation of YRG2 cells* ..... 61

*2.3.4 Yeast two-hybrid β-galactosidase assay* ..... 62

**2.4 Cell culture ..... 63**

*2.4.1 Reagents, stock solutions and buffers for tissue culture* ..... 63

*2.4.2 Retrieving cells from liquid nitrogen storage* ..... 63

*2.4.3 Maintenance of HEK293 cell line* ..... 63

*2.4.4 Freezing cells for long term storage* ..... 64

*2.4.5 Transfection with Calcium Phosphate* ..... 64

*2.4.6 Transfection with GeneJuice* ..... 65

**2.5 Genome editing with CRISPR/Cas9 technology ..... 66**

*2.5.1 gRNAs design and cloning* ..... 66

*2.5.2 HEK293 transfection and generation of stable cell lines* ..... 66

*2.5.3 gRNAs screening for efficiency* ..... 67

*2.5.4 Genome editing testing* ..... 67

**2.6 Protein expression and purification ..... 68**

*2.6.1 Reagents, stock solutions and buffers* ..... 68

*2.6.2 Expression of His-tagged proteins* ..... 68

*2.6.3 Expression of StrepII-tagged proteins* ..... 69

*2.6.4 Purification of His-tagged proteins* ..... 69

*2.6.5 Purification of StrepII-tagged proteins* ..... 70

*2.6.6 Purification by size-exclusion chromatography* ..... 71

*2.6.7 TEV cleavage* ..... 71

**2.7 Protein analysis ..... 72**

*2.7.1 Reagents, stock solutions and buffers for use in protein analysis* ..... 72

*2.7.2 Bradford assay to determine protein concentration* ..... 72

*2.7.3 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis* ..... 73

*2.7.4 Western Blot Analysis* ..... 73

**2.8 Deadenylation assays ..... 75**

*2.8.1 Reagents, stock solutions and buffers* ..... 75

*2.8.2 Denaturing polyacrylamide gel electrophoresis* ..... 75

*2.8.3 Fluorescence-based deadenylase assay* ..... 76

*2.8.4 Coating of poly(A)$_{20}$ and poly(A)$_{50}$ RNA oligos with PABPC1 proteins* ..... 76

**2.9 Differential scanning fluorimetry (DSF) ..... 78**

**Chapter 3. Cellular and biochemical tools to characterise the human PAN2-PAN3 complex ..... 80**

**3.1 Introduction ..... 80**

**3.2 Bioinformatic analysis of PAN2 and PAN3 variants associated with cancer and conserved residues of the PAN2-PAN3 complex .. 81**

**3.3 Generation of PAN2 knockout HEK293 stable cell lines ..... 91**

*3.3.1 Establishing a CRISPR/Cas9 protocol to target genes: a proof of principle* ..... 91

*3.3.2 Targeting the PAN2 gene* ..... 91

**3.4 Investigation of critical mutations in the PAN2-PAN3 complex 96**

*3.4.1 Potential amino acids mediating the PAN2-PAN3 interaction and the PAN3-PAN3 homodimer formation* ..... 96

*3.4.2 Generation of plasmids for yeast two-hybrid experiments* ..... 96

*3.4.3 Identification of residues involved in the PAN2-PAN3 interaction and the PAN3-PAN3 homodimer formation* ..... 96

**3.5 Expression and purification of the human PAN2-PAN3 complex ..... 100**

*3.5.1 Vectors for protein expression in* E. coli *BL21 DE3 strain* ..... 100

*3.5.2 PAN2-PAN3 complex expression in* E. coli *BL21 DE3* ..... 100

*3.5.3 Vectors for protein expression in* E. coli *ArcticExpress DE3* ..... 105

*3.5.4 PAN2-PAN3 complex expression in* E. coli *ArcticExpress DE3* ..... 105

*3.5.5 Vectors for the expression of PAN2 and PAN3 fragments* ..... 108

*3.5.6 Expression of PAN2 and PAN3 fragments in* E. coli *BL21 DE3* ..... 108

*3.5.7 PAN2 and PAN3 fragments expression in* E. coli *ArcticExpress DE3* ..... 108

*3.5.8 Vectors for co-expression of PAN2-PAN3 complex and PABPC1 in* E. coli *BL21 DE3* ..... 112

*3.5.9 PAN2-PAN3 complex and PABP co-expression in* E. coli *BL21 DE3* ..... 112

*3.5.10 Vectors to express catalytically active PAN2 fragments* ..... 114

*3.5.11 PAN2 fragments expression in* E. coli *BL21 DE3* ..... 114

**3.6 Discussion ..... 116**

**Chapter 4. Expression and purification of a human Ccr4-Not core complex composed of BTG2, Caf1, Ccr4, CNOT1M and CNOT9 ..... 120**

**4.1 Introduction ..... 120**

**4.2 Expression and purification of the deadenylase enzyme Caf1 and the BTG2-Caf1 dimer ..... 122**

*4.2.1 Vectors for protein expression in* E. coli *BL21 DE3* ..... 122

*4.2.2 Caf1 and BTG2-Caf1 expression and purification* ..... 123

**4.3 Reconstitution of a nuclease module of the Ccr4-Not complex including Ccr4, Caf1 and BTG2 .....125**

*4.3.1 Vectors for protein co-expression in* E. coli *BL21 DE3* ..... 125

*4.3.2 Nuclease module expression and purification* ..... 125

**4.4 Expression and purification of a CNOT1M/CNOT9 module .....128**

*4.4.1 Vector for protein co-expression in* E. coli *BL21 DE3* ..... 128

*4.4.2 CNOT9 module expression and purification* ..... 128

**4.5 Co-expression and purification of BTG2, Caf1, Ccr4, CNOT1M and CNOT9 .....131**

*4.5.1 Vectors for protein co-expression in* E. coli *BL21 DE3* ..... 131

*4.5.2 Core complex co-expression in* E. coli *BL21 DE3* ..... 131

*4.5.3 Core complex purification with IMAC* ..... 131

*4.5.4 Core complex purification with affinity chromatography using the StrepII tag* ..... 132

*4.5.5 Analysis of core complex obtained after co-expression and purification* 132

**4.6 Assembly of the core complex by reconstitution using purified nuclease and CNOT9 modules.....134**

*4.6.1 Core complex isolation by size-exclusion chromatography* ..... 134

*4.6.2 Core complex purification with IMAC* ..... 134

*4.6.3 Core complex purification with StrepII affinity chromatography* ..... 135

**4.8 Discussion.....137**

**Chapter 5. Biochemical activity of a human Ccr4-Not core complex ..... 141**

**5.1 Introduction.....141**

**5.2 Deadenylase activity degrading short poly(A)$_9$-tailed mRNAs ..... 143**

*5.2.1 A fluorescence-based assay to characterise deadenylation activity* ..... 143

*5.2.2 Both the anti-proliferative protein BTG2 and the CNOT1-CNOT9 module enhance deadenylation* ..... 143

**5.3 Deadenylase activity degrading poly(A)$_{20}$-tailed RNA ..... 147**

*5.3.1 A PAGE-based assay to characterise deadenylation activity* ..... 147

*5.3.2 Enzymes and complexes catalytic activity with a poly(A)$_{20}$ substrate* ..... 147

**5.4 Deadenylase activity degrading poly(A)$_{50}$-tailed RNA ..... 151**

*5.4.1 A PAGE-based assay to characterise deadenylation activity* ..... 151

*5.4.2 Enzymes and complexes catalytic activity with a poly(A)$_{50}$ substrate* ..... 151

**5.5 Expression and purification of the human Poly(A)-Binding Protein (PABPC1) ..... 154**

*5.5.1 Vectors for protein expression in* E. coli *BL21 DE3* ..... 154

*5.5.2 PABPC1 (RRM1, RRM1-2 and FL) expression and purification* ..... 154

*5.5.3 poly(A) tails coating with PABPC1 purified proteins* ..... 155

**5.6 Deadenylase activity of Caf1 and Caf1-complexes using poly(A)$_{20}$- PABPC1 substrates ..... 158**

*5.6.1 A PAGE-based assay to characterise deadenylation activity* ..... 158

*5.6.2 Activity of enzymes and complexes using a poly(A)$_{20}$ substrate* ..... 158

**5.7 Deadenylase activity of Caf1 and Caf1-complexes using poly(A)$_{50}$- PABPC1 substrates ..... 163**

*5.7.1 A PAGE-based assay to characterise deadenylation activity* ..... 163

*5.7.2 Activity of enzymes and complexes using a poly(A)$_{50}$ substrate* ..... 163

**5.8 Selectivity for poly(A), poly(U), poly(C) and poly(G) RNA substrates ..... 168**

*5.8.1 A PAGE-based assay to characterise deadenylation activity* ..... 168

*5.8.2 The Ccr4-Not core complex displays greater selectivity towards the poly(A) oligo when compared to the nuclease module* ..... 168

**5.9 Discussion ..... 170**

**Chapter 6. Biophysical and biochemical characterisation of the melanoma-associated P131L variant of CNOT9 ..... 174**

**6.1 Introduction ..... 174**

**6.2 Computational modelling of the CNOT9 P131L variant ..... 176**

**6.3 Expression and purification of CNOT9 P131L protein complexes ..... 178**

*6.3.1 Vector for protein expression in* E. coli *BL21 DE3* ..... 178

*6.3.2 Expression and purification of CNOT9 P131L and CNOT9 P131L-containing complexes* ..... 179

*6.3.3 Ccr4-Not core complex with CNOT9 P131L* ..... 179

**6.4 Structural characterisation of the CNOT9 P131L variant ..... 183**

*6.4.1 Differential scanning fluorimetry* ..... 183

*6.4.2 Crystallisation of the DUF3910-CNOT9 (P131L) complex* ..... 183

**6.5 Characterisation of the activity and selectivity of the Ccr4-Not core complex containing the CNOT9 P131L variant ..... 187**

*6.5.1 Deadenylation assays* ..... 187

*6.5.2 Selectivity assays* ..... 193

**6.6 Discussion ..... 195**

**Chapter 7. Concluding remarks ..... 199**

**7.1 PAN2-PAN3 complex characterisation ..... 199**

**7.2 Reconstitution of a human Ccr4-Not core complex ..... 200**

**7.3 Deadenylation activity of the Ccr4-Not core complex ..... 201**

**7.4 Investigation on CNOT9 P131L variant associated with melanoma ..... 203**

**7.5 Future outlook..... 203**

**References ..... 205**

**Appendix 1. Investigating BTG2-PABPC1 interaction via BoxC domain of BTG2 ..... 220**

**Appendix 2. Characterisation of deadenylation complexes in human cells by immunocytochemistry ..... 223**

**A2.1 Immunofluorescence microscopy ..... 225**

A2.1.2 Reagents, stock solutions and buffers for use in immunofluorescence ..... 225

A2.2.2 Preparation of mammalian cells for immunofluorescence ..... 225

**Appendix 3. Crystallisation of CNOT9 P131L ..... 231**

**Appendix 4. Structural characterisation of the human Ccr4-Not complex by Electron Paramagnetic Resonance..... 235**

**Appendix 5. Incorporation of unnatural amino acids in the deadenylation enzymes of the Ccr4-Not complex..... 237**

**List of figures**

Figure 1.1 - A schematic overview of the 'canonical' pathways of regulated mRNA degradation. ..... 8

Figure 1.2 – The trimeric S. cerevisiae PAN2-(PAN3)2 complex structure. ..... 12

Figure 1.3 – The generic recruitment of the PAN2-PAN3 complex. ..... 15

Figure 1.4 – The Ccr4-Not complex. ..... 18

Figure 1.5 – The deadenylase module of the Ccr4-Not complex. ..... 21

Figure 1.6 – Structural representations of the central domains of CNOT1. ..... 23

Figure 1.7 – The Not-module of the Ccr4-Not complex. ..... 25

Figure 1.8 – Structural representations of Caf1, BTG/Tob proteins and PABPC1. ..... 28

Figure 1.9 – Deadenylation by Ccr4-Not is triggered by three recruitment mechanisms. ..... 31

Figure 1.10 – Roquin1 interactions. ..... 32

Figure 1.11 – Regulatory mechanisms of the Ccr4-Not complex. ..... 33

Figure 1.12 – The Lsm1-7 complex. ..... 36

Figure 1.13 – A proposed model for the step-wise regulation of the mRNA decapping mechanism. ..... 39

Figure 1.14 – Structural representation of the Dm Xrn1 enzyme. ..... 44

Figure 3.1 – Schematic representation of the mutations distribution in the PAN2 and PAN3 genes. ..... 86

Figure 3.2 – Structural representation of the PAN2-PAN3 complex. ..... 87

Figure 3.3 – Conserved residues in the PAN2-PAN3 complex. ..... 89

Figure 3.4 –Schematic diagram indicating the distribution of cancer-associated variants at highly conserved positions in PAN2 and PAN3. ..... 90

Figure 3.5 – Plasmids used for CRISPR/Cas9 genome editing..... 93

Figure 3.6 – Analysis of cell lines edited by CRISPR/Cas9. ..... 94

Figure 3.7 – Analysis of the gRNAs efficiency in a population of cells..... 95

Figure 3.8 – Overview on the yeast two-hybrid experiments..... 98

Figure 3.9 – Analysis of the yeast two-hybrid assays..... 99

Figure 3.10 - Vectors generated for the PAN2-PAN3 complex expression in *E. coli* BL21 DE3. .....102

Figure 3.11 – Expression of the PAN2-PAN3 complex in *E. coli* BL21 DE3. .....103

Figure 3.12 – Western blots analysis of the protein expression cultures 104

Figure 3.13 - Vectors generated for the PAN2-PAN3 complex expression in *E. coli* ArcticExpress DE3. .....106

Figure 3.14 - Expression of the PAN2-PAN3 complex in *E. coli* ArcticExpress DE3.....107

Figure 3.15 –Prediction of the structured regions of both human PAN2 and PAN3 using the I-TASSER server.....109

Figure 3.16 – Vectors generated for the expression of a PAN2-PAN3 complex composed by fragments in *E. coli* BL21 DE3 and ArcticExpress DE3. .....110

Figure 3.17 - Expression of the PAN2-PAN3 complex in *E. coli* BL21 DE3 (A) and ArcticExpress DE3 (B).....111

Figure 3.18 – Vector L07 generated to express PABPC1 in *E. coli* BL21 DE3. .....113

Figure 4.1 – Caf1 expression in *E. coli* BL21 DE3. .....124

Figure 4.2 – Vectors generated for the nuclease module expression in *E. coli* BL21 DE3. ..... 127

Figure 4.3 – Vector generated for the CNOT9 module expression in *E. coli* BL21 DE3. ..... 130

Figure 4.4 – Vectors generated for the co- expression of the Ccr4-Not core complex in *E. coli* BL21 DE3. ..... 133

Figure 4.5 – Outline of the purification strategy designed to isolate the core complex of Ccr4-Not ..... 136

Figure 5.1 – Analysis of the purified deadenylase Caf1; Caf1 in complex with BTG2; the nuclease module (Caf1, Ccr4 and BTG2); the Ccr4-Not core complex (Caf1, Ccr4, BTG2, CNOT1M and CNOT9). ..... 144

Figure 5.2 – Fluorescence-based deadenylation assays. ..... 145

Figure 5.3 – Outline of the in vitro deadenylation reactions designed with the poly(A)-tailed RNA substrate. ..... 148

Figure 5.4 – Deadenylation activities of the Caf1 enzyme and the Caf1-containing complexes. ..... 149

Figure 5.5 –Deadenylation activities of the Caf1 enzyme and the Caf1-containing complexes. ..... 152

Figure 5.6 – PABPC1 expression in *E. coli* BL21 DE3. ..... 156

Figure 5.7 – Electrophoretic mobility shift assay of PABPC1-poly(A) protein:RNA complexes. ..... 157

Figure 5.8 – Comparison of the deadenylation activities of the Caf1 enzyme and the Caf1-containing complexes in presence of a 8-fold excess of PABPC1 (1-190). ..... 159

Figure 5.9 – Comparison of the deadenylation activities of the Caf1 enzyme and the Caf1-containing complexes in presence of a 6-fold excess of PABPC1 (FL). ..... 161

Figure 5.10 –Deadenylation activities of Caf1 and Caf1-containing complexes using PABPC1-poly(A)$_{50}$ substrates. .....164

Figure 5.11 – Deadenylation activities of Caf1 and Caf1-containing complexes using PABPC1-poly(A)50 substrates. .....166

Figure 5.12 – Comparison of the substrate selectivity of the Ccr4-Not core complexes.....169

Figure 6.1 – Structural analysis of the CNOT9 P131L variant. .....177

Figure 6.2 – Vectors generated for the expression of CNOT9 P131L in *E. coli* BL21 DE3. .....181

Figure 6.3 – Expression of CNOT9 (P131L)-containing complexes.....182

Figure 6.4 – Differential scanning fluorimetry (DSF) analysis comparing CNOT9 wild-type with the P131L variant. .....185

Figure 6.5 – Differential scanning fluorimetry (DSF) analysis comparing wild-type CNOT9 and the P131L variant. .....186

Figure 6.6 - Analysis of the purified Ccr4-Not core complexes (Caf1, Ccr4, BTG2, CNOT1M and CNOT9).....188

Figure 6.7 - Comparison of the deadenylation activities of the Cc4-Not core complexes using a poly(A)$_{20}$ oligo as a substrate .....189

Figure 6.8 – Comparison of the deadenylation activities of the Cc4-Not core complexes using a poly(A)$_{20}$ oligo coated with the RRM1-RRM2 domains of PABPC1.....190

Figure 6.9 – Comparison of the deadenylation activities of the Cc4-Not core complexes using a poly(A)$_{20}$ oligo coated with PABPC1.....191

Figure 6.10 - Comparison of the deadenylation activities of the Cc4-Not core complexes using a poly(A)$_{50}$ oligo as a substrate .....192

Figure 6.11 - Comparison of the substrate selectivity of the Ccr4-Not core complexes.....194

Figure A1.1 – Purified complexes including BTG2 showing an altered BoxC .....221

Figure A2.2 – Representative panels comparing HEK293 cells .....228

Figure A2.3 – Ana.ysis of the P-bodies formation in HEK293 cells.....229

Figure A3.1 – Plate crystals obtained from the optimisation step .....233

Figure A5.1 – Analysis of Caf1 (N-term Amber codon) expression .....238

**List of tables**

Table 2.1 – List of primers. ..... 59

Table 2.2 - Mixture for yeast transformation. ..... 62

Table 2.3 - List of plasmids for genome editing. ..... 66

Table 2.4 - List of antibodies used for Western blotting. ..... 74

Table 2.5 - The 20% denaturing gel composition for deadenylation assays. ..... 76

Table 3.1 - PAN2 mutations collected from the COSMIC database. ..... 84

Table 3.2 – PAN3 mutations collected from the COSMIC database. ..... 85

Table 3.3 - List of pQE80L vectors generated to express different PAN2 fragments. ..... 115

Table A2.1 - List of the antibodies used in the immunofluorescence microscopy experiments. ..... 227

Table A3.1 - Optimisation plate set up after 'hit' in the screening plate... 232

## List of abbreviations

**Ago** Argonaute

**ARE** A-U rich element

**ARE-BP** ARE binding protein

**ATP** Adenosine triphosphate

**BSA** Bovine serum albumin

**Cas9** CRISPR associated protein 9

**CCR4** Carbon catabolite repression 4

**CRISPR** Clustered regularly interspaced short palindromic repeats

**Dcp1** Decapping enzyme 1

**Dcp2** Decapping enzyme 2

**DMSO** Dimethyl sulfoxide

**DNA** Deoxyribonucleic acid

**DUF3819** Domain of unknown function 3819

***E. Coli*** *Escherichia coli*

**EDTA** Ethylenediaminetetraacetic acid

**EDC3** Enhancer of decapping-3

**EEP** Exonuclease-endonuclease-phosphatase

**eIF4** eukaryotic initiation factor 4

**Flc** Fluorescein

**FRET** *Förster* resonance energy transfer

**GW182/TNRC6** Glycine-Tryptophan protein of 182 kDa/ Trinucleotide Repeat Containing 6A

**HEAT** Huntingtin, Elongation factor 3 (EF3), protein phosphatase 2A (PP2A), TOR1

**HEK293** Human embryonic kidney 293

**IMAC** Immobilised metal affinity chromatography

**IPTG** Isopropyl β-D-1-thiogalactopyranoside

**LB** Lysogeny Broth

**LRR** Leucine-rich repeat

**MCS** Multi cloning sites

**$m^7G$** 7-methyl-guanosine

**MIF4G** Middle domain of eukaryotic initiation factor 4G

**mRNA** Messenger RNA

**miRNA** MicroRNA

**miRISC** MicroRNA induced silencing complex

**NMD** Nonsense mediated decay

**NSD** Nonstop decay

**PABPC1** Poly(A) binding protein cytoplasmic-1

**PAN2** Poly(A) specific ribonuclease subunit 2

**PAN3** Poly(A) specific ribonuclease subunit 3

**PARN** Poly(A) specific ribonuclease

**PCR** Polymerase Chain Reaction

**Poly(A)** Poly adenosine tail

**RNA** Ribonucleic acid

**rRNA** Ribosomal RNA

**RRM** RNA recognition motif

**RQCD1** Required for cell differentiation 1

**SDS-PAGE** Sodium dodecyl sulfate polyacrylamide gel electrophoresis

**SEC** Size exclusion chromatography

**ssRNA** single stranded RNA

**TAMRA** Tetramethylrhodamine

**TTP** Tristetraprolin

**UTR** Untranslated region

**WD-40** tryptophan-aspartic acid 40 amino acids

**XRN1** 5'-3' Exoribonuclease 1

# Chapter 1

## Introduction

# Chapter 1. Introduction

## 1.1 mRNA biogenesis and export

Messenger RNA (mRNA) is the key molecule to translate the information coded by genes into proteins. mRNAs need to be accurately synthesised, as any error could lead to non-functional, misregulated, or potentially toxic proteins. In eukaryotic cells, mRNA synthesis and export to the cytoplasm are coupled, and several protein complexes are involved.

In the nucleus, RNA Polymerase II transcribes all genes that encode for proteins. RNA Pol II associates with several general transcription factors in the pre-initiation complex, which is regulated by the mediator (Woychik & Hampsey 2002; Conaway & Conaway 2011). After the pre-initiation complex is recruited on the promoter, TFIIH, which contains a DNA helicase, unwinds the double-strand DNA and Pol II starts synthesising RNA. After the first few bases, Pol II is released from the promoter and the pre-initiation complex disassembles, and the mediator recruits the next Pol II (Yudkovsky et al. 2000). Once entering the elongation step RNA Pol II is highly processive and can achieve many kilobases of transcribed RNA without dissociating (Singh & Padgett 2009). RNA transcription termination is precisely regulated to facilitate Pol II re-engagement on a different promoter and avoid mis-regulation of downstream genes. Termination can be achieved by Pol II stalling and conformational change leading to dissociation during 3' polyadenylation (Greenblatt et al. 1993). An alternative route is the 'torpedo model', where a 5'-3' exonuclease Xrn2, which enters the cleaved RNA, displaces the elongating Pol II that continues downstream RNA synthesis (Proudfoot 2016).

Co-transcriptionally, three main mRNA maturation steps take place: splicing, 5' capping and 3' polyadenylation (Buratowski 2009; Kuehner et al. 2011; Egloff & Murphy 2008). They are all required for mRNA export in

the cytoplasm, and to ensure mRNA stability and efficient translation. The key player of these processes is the C-terminal domain (CTD) of Rpb1, a subunit of RNA Pol II. Its deletion impairs them but has no effect on transcription rates (McCracken et al. 1997). Interestingly, the phosphorylation status of this domain, which contains a seven-amino acid repeat, has direct consequences on the activity of Pol II and is found next to the channel where the newly synthesised RNA is threaded (Cramer 2001; Fabrega et al. 2003). After several nucleotides are synthesised, the $m^7G$ cap is added to RNA trascripts by two enzymes called Human Capping Enzyme (HCE) and RNA 7-methyltransferase (MT) (Hirose & Manley 2000), recruited by the CTD of Pol II. Then, it is tethered by a cap-binding complex to prevent degradation by exonucleases (Izaurralde et al. 1994). This structure is also involved in mRNA export and translation (Proudfoot et al. 2002). The spliceosome, an RNA-protein machinery, is responsible for precisely removing introns, the non-coding regions of genes (Will & Lührmann 2011). RNA splicing is a highly regulated process, that involves several conserved elements such as 5' and 3' splice sites, a branch-point and a poly-pyrimidine sequence. The CTD of Pol II helps organising the splicing procedure by directly recruiting splicing co-factors, and its activity is phosphorylation-dependent (Das et al. 2007). Finally, the poly(A)-tail is added to the 3'-end of mRNAs. Before its addition, the pre-mRNA is cleaved by a protein machinery that includes a cleavage and polyadenylation specificity factor (CPSF) that recognises conserved sequence elements and, indirectly regulated by Pol II CTD, cleaves the substrate (Glover-Cutter et al. 2008). Then, the poly(A) tail is addedby a nuclear poly(A) polymerase; its length can vary between 40-50 nucleotides to over one hundred, but generally shorter than previously thought (H. Chang et al. 2014). Polyadenylation is stimulated by transcription and its length is related to mRNA stability (H. Chang et al. 2014) and potentially enhanced translation (Burgess & Gray 2010). mRNA export to the cytoplasm, where translation

occurs, is executed by the nuclear pore complex, in a process mediated by several transport co-factors, such as SR proteins which are also involved in splicing (Köhler & Hurt 2007; Huang et al. 2003). An additional control mechanism for mRNA stability and translation is cytoplasmic polyadenylation, or the elongation of poly(A)-tails in the cytoplasm, which generally triggers translation activation in response to specific signals (Charlesworth et al. 2013). Largely, this process is regulated by CPEB (cytoplasmic polyadenylation element binding) proteins and the best-characterised mechanisms are involved in oocytes development in *Xenopus* (Villalba et al. 2011).

### 1.2 Overview of the pathways for cytoplasmic mRNA decay

Post-transcriptional gene regulation is one of the most important ways for the cells to determine the expression rates of specific genes. By finely regulating the mRNAs half-lives, thanks to multi-step decay pathways controlled by specialised multimeric complexes, cells are able to respond to extra-cellular stimuli in a precise fashion.

A quality control machinery is also constantly checking the newly synthesized mRNAs. For instance, mRNAs displaying premature stop codons are targeted by the non-sense mediated decay (NMD) machinery, which promotes accelerated degradation to avoid the synthesis of potentially toxic truncated proteins (Brogna & Wen 2009; Schweingruber et al. 2013). Similarly, transcripts lacking translation stop codons are targeted by a nonstop decay (NSD) pathway and efficiently destroyed by the cytoplasmic exosome (Vasudevan et al. 2002; Schoenberg & Maquat 2012). Finally, no-go decay (NGD) targets long mRNAs displaying secondary structures that prevent ribosomes to slide along the nucleic acid sequence (Doma & Parker 2006; Schoenberg & Maquat 2012). These mRNAs are cleaved by endonucleases and subsequently degraded by the exosome (3'-

5') and the exoribonuclease XRN1 (5'-3'). The focus of this introduction is the canonical pathway for cytoplasmic mRNA decay in eukaryotes.

Over two decades of research on mRNA degradation have helped identifying the main pathways for mRNA decay. In eukaryotes, mature mRNAs are transported from the nucleus to the cytoplasm in a circle-loop conformation. A poly(A) tail (70-120 nucleotides) is added co-transcriptionally to the 3'-end (Zhao et al. 1999) and a $m^7G$ (7-methylguanylate) cap is tethered to virtually all mRNAs. Such 'closed' conformation prevents premature or non-specific degradation and enhances translation rates and efficiency (Garneau et al. 2007). Interestingly, this circular structure is promoted by eIF4F (eukaryotic translation initiation factor 4F), which is able to bind the $m^7G$ cap, the poly(A) tail and, by recruiting eIF4G, the poly(A) binding protein (PABPC1) (Tarun & Sachs 1996; Imataka et al. 1998; Wells et al. 1998; Safaee et al. 2012). This underlines how transcription and mRNA maturation in the nucleus, and mRNA translation and turnover in the cytoplasm are closely related.

### 1.3 Deadenylation

Given the cross-sectional importance of the poly(A) tail, it is not surprising that the poly(A) tail-shortening process, or deadenylation, is a major discriminant between decay pathways for mRNA decay. Poly(A)-tail removal is a crucial step that triggers mRNA turnover (Fig. 1.1). However, it is also the rate limiting step and can be reverted when particular transcripts contain conserved sequences for readenylation (Weill et al. 2012). Several RNA-binding proteins are associated with targeted deadenylation, such as Roquin, TTP and Nanos (Leppek et al. 2013; Sgromo, Raisch, Bawankar, et al. 2017; Sandler et al. 2011; Fabian et al. 2013; Suzuki et al. 2010; Suzuki et al. 2012; Joly et al. 2013; Bhandari et al. 2014), as well as the miRNA machinery (Fabian 2011; Chekulaeva et al. 2011).

Interestingly, PABPC1 is also directly linked to poly(A) tail removal (Stupfler et al. 2016), which suggests a two-faced role of this key protein.

Shortening poly(A) tails of mRNA molecules triggers degradation through either one of two distinct routes (Fig. 1.1). The $m^7G$ cap is removed by a decapping complex and the mRNAs is destroyed by a conserved 5'-3' XRN1 exonuclease (Braun et al. 2012; C. Te Chang et al. 2014; Valkov et al. 2016). Alternatively, mRNAs are degraded in a 3'-5' direction by a large exosome complex and the 5'-cap is hydrolysed by a decapping enzyme (Chen et al. 2005; Milac et al. 2014; Zhou et al. 2015).

MicroRNAs (miRNAs) are short, non-coding RNA molecules of paramount importance in gene expression regulation in eukaryotes. They are usually 20-22 nucleotides long, and different families of miRNAs are based on the sequence of nucleotide 2-8, which are the most important for mRNA targeting. miRNAs are processed by RNase III proteins such as Drosha and Dicer, and are usually located within introns or non-coding transcripts, although a small percentage can be found in exons (Ha & Kim 2014). Within the mRNA targeting process, miRNAs provide the target specificity by base-pairing, usually at the 3'-UTR of mRNAs (Bartel 2009), while AGO proteins are the effectors, by recruiting co-factors that triggers mRNA deadenylation and decay, and translation repression (Huntzinger & Izaurralde 2011). All steps of miRNA biogenesis are finely regulated, to ensure no off-target effects are triggered. miRNAs are transcribed by the RNA polymerase II, which produces a pri-miRNA (1 kb) showing an hairpin structure (Lee et al. 2002). Then, the micro-processor complex, including Drosha and DGCR8, cleaves the pri-miRNA into a pre-miRNA (Lee et al. 2003). The pre-miRNA is then transported to the cytoplasm via the exportin 5 (EXP5) and cleaved by Dicer into an RNA duplex (Hutvágner et al. 2001). Finally, the duplex is loaded onto the AGO protein, forming the pre-RISC complex. After unwinding of the duplex, the mature RISC is formed, ready to target specific

mRNAs (Kawamata & Tomari 2010). Additional non canonical pathways have been recently identified for miRNA synthesis and maturation, that by-pass Drosha and Dicer (Ha & Kim 2014).

Another crucial aspect to consider for understanding the mRNA decay process is its localisation within the cell. In eukaryotes, cytoplasmic foci called processing bodies (P-bodies) have been discovered and characterised as the centres for mRNA degradation (Parker & Sheth 2007; Eulalio, Behm-Ansmant & Izaurralde 2007; Chen 2013). P-bodies are cytoplasmic foci where mRNAs that are translationally repressed are co-localised with several complexes belonging to the mRNA deadenylation, decapping and decay, and the translation repression pathways. Their size and protein composition seem to vary rapidly in response to stress and other external stimuli (Parker & Sheth 2007). Several studies have collected evidence for a potential dynamic cycle between P-bodies and polysomes. mRNAs that enter P-bodies can either be degraded or stored and re-enter the translation mechanisms (Brengues et al. 2005; Bhattacharyya et al. 2006; Andrei et al. 2005).

As discussed in several studies, most of the enzymatic complexes involved in mRNA decay are co-localised in the P-bodies (Eulalio, Behm-Ansmant, Schweizer, et al. 2007). Interestingly, some enzymes seem to have not only a functional role but also a structural implication within the P-bodies (Zheng et al. 2008). For instance, there is evidence for a potential interaction of the two main deadenylation complexes *in vivo*; however, only indirect links between key proteins could be determined (Zheng et al. 2008). Future studies are needed to shed light on these interactions, together with an investigation of the precise role of each enzyme within the P-bodies.

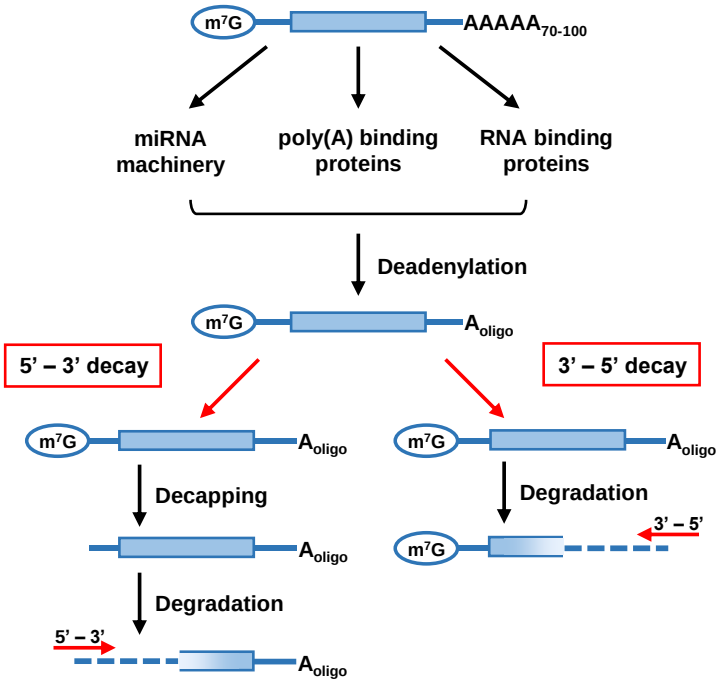



**Figure 1.1** - A schematic overview of the 'canonical' pathways of regulated mRNA degradation. The initial deadenylation step is common to both pathways. It triggers decapping and the 5'-3' decay cascade or direct degradation in the 3'-5' pathway.

In eukaryotes, the mRNA poly(A) tail shortening is the main function of two multimeric complexes: the Ccr4-Not and the PAN2-PAN3 complex (Wahle & Winkler 2013). Extensive studies have established targeted recruitment mechanisms for the Ccr4-Not complex, and the activity of both complexes is triggered in miRNA-mediated mRNA degradation. However, no other targeted recruitment mechanism has been discovered for the PAN2-PAN3 complex to our knowledge, suggesting a broader activity range for this complex. In a biphasic model PAN2-PAN3 is responsible for the initial deadenylation step, whereas Ccr4-Not is the complex removing the last 20-30 nucleotides. An alternative hypothesis could see a poly(A) tail maintenance role for the PAN2-PAN3 complex, that would act in synergy with the poly(A) polymerase (PAP), responsible for readenylation. On the other hand, Ccr4-Not is recruited for regulated mRNA deadenylation.

## 1.4 PAN2-PAN3 complex

The PAN2-PAN3 complex is a highly conserved multisubunit machinery (Wahle & Winkler 2013). Many of its features have been extensively investigated in the last two decades, both *in vitro* and *in vivo* (Uchida et al. 2004). Structural studies have recently explored the architecture of this atypically asymmetric trimeric complex (PAN2-PAN3/PAN3) in model organisms such as *N. crassa* (Christie et al. 2013; Jonas et al. 2014), *D. melanogaster* (Christie et al. 2013), *C. thermophilum* (Wolf et al. 2014) and *S. cerevisiae* (Schäfer et al. 2014). However, no crystal structure of the human PAN2-PAN3 complex is available to date due to the low solubility of these proteins in the most used expression systems.

### *1.4.1 The catalytically active enzyme PAN2*

PAN2 is a 135 kDa cytoplasmic exoribonuclease that belongs to the DEDD/RNAseD superfamily of exonucleases, which display these four residues in the active site, co-ordinating two divalent metal ions. It is $Mg^{2+}$ dependent and releases 5'-AMP upon hydrolysis of the poly(A) mRNA tail (Uchida et al. 2004). Interestingly, PAN2 has been shown to have no ability to directly bind RNA, and in absence of PAN3 has almost negligible catalytic activity (Wolf et al. 2014; Schäfer et al. 2014). Upon binding of its complex partner, it shows high deadenylation activity and substrate specificity (Wolf et al. 2014). PAN2 encompasses several domains and conserved regions (Uchida et al. 2004). At its N-terminal region, it contains a conserved WD40 domain, a rigid protein platform which mediates specific interactions with PAN3 (Jonas et al. 2014; Schäfer et al. 2014) (Fig. 1.2). A linker region downstream the WD40 domain is perhaps the most significant sequence for the complex formation, as underlined in different studies (Wolf et al. 2014; Jonas et al. 2014; Schäfer et al. 2014). This sequence contains several conserved residues responsible for binding a coiled coil formed by the PAN3 homodimer (Fig. 1.2). In addition, upon binding PAN3 C-terminal

domain, the linker sterically prevents a second WD40 domain to bind; therefore, a second PAN2 protein cannot bind the PAN3 homodimer (Jonas et al. 2014). A central region of PAN2 is a ubiquitin specific protease domain (USP); however, lacking essential residues in the catalytic triad, it shows no activity (Quesada et al. 2004; Jonas et al. 2014; Schäfer et al. 2014). The USP domain engages extensive interactions with the PAN2 C-terminus exonuclease domain (Schäfer et al. 2014) (Fig. 1.2). The conserved RNase domain shares an overall similar structure with other exo-RNases of the DEDD superfamily but presents differences at the interface with the USP domain (Jonas et al. 2014; Schäfer et al. 2014). The active site is altogether different from the consensus structure of the family; while it still coordinates magnesium ions to catalyse the reaction, it is deep and engages the poly(A) nucleotides via a conserved aromatic residue (Schäfer et al. 2014), a mechanisms displayed for instance by Ccr4, which belongs to a different superfamily (EEP - exonuclease/endonuclease/phosphatase) (Wang et al. 2010). Despite this similarity, Ccr4 and PAN2 show different substrate preferences in *S. cerevisiae* (Sun et al. 2013).

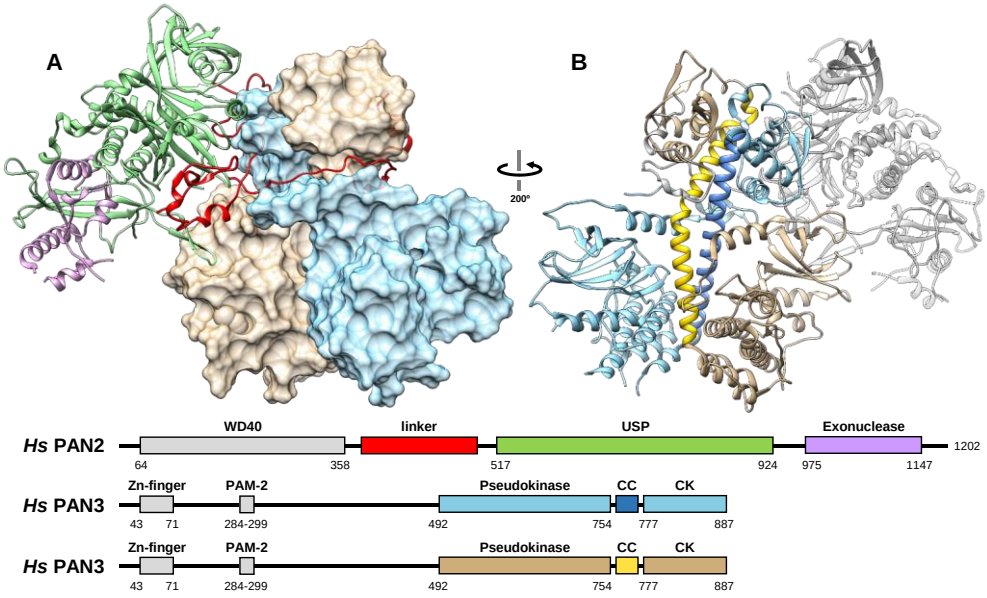



**Figure 1.2** – The trimeric *S. cerevisiae* PAN2-$(PAN3)_2$ complex structure. The linker region (red) wraps around the asymmetric PAN3 homodimer (tan and sky blue). The coiled-coil regions of the two PAN3 proteins (yellow and blue) prove the asymmetry of the homodimer, as the axis of the 'external' PAN3 is bent towards PAN2 (grey) (Schäfer et al. 2014).

### *1.4.2 The regulatory subunit PAN3*

PAN3 is a 95 kDa protein that forms an asymmetric homodimer when assembled in the yeast PAN2-PAN3 complex (Schäfer et al. 2014). It displays several conserved motifs. *S. cerevisiae* PAN3 presents a CCCH-type zinc-finger domain on the N-terminus, that has been shown to bind mRNAs poly(A) tail (Wolf et al. 2014). A PAM-2 motif is present downstream, within an unstructured region; PAM-2 motifs are highly conserved structures that have been largely associated with protein-protein interactions with the C-terminus of PABP across a number of species (Siddiqui et al. 2007) (Fig. 1.2). The central region of PAN3 contains a pseudokinase (PK) domain. This shows a different structure folding and lacks a few catalytic residues when compared to other kinases, but seems to have retained the ability to bind ATP in a $Mg^{2+}$-dependent way, although displays no activity *in vitro* (Christie et al. 2013). A coiled-coil region between the PK and the C-terminal domain (CTD) is responsible for the assembly of the asymmetric homodimer in various organisms (Christie et al. 2013; Jonas et al. 2014; Wolf et al. 2014; Schäfer et al. 2014) (Fig. 1.2). Upon formation of the asymmetric homodimer, specific residues of PAN3 in the coiled coil have been identified to engage conserved interactions with PAN2 and with the CTD of PAN3, showing a great reduction of the overall energy of the complex (Christie et al. 2013; Jonas et al. 2014; Wolf et al. 2014; Schäfer et al. 2014).

### *1.3.3 Generic recruitment mechanisms*

The recently discovered asymmetric conformation of the PAN2-PAN3 complex gives a few insights on its role in mRNA deadenylation in eukaryotes. Interestingly, PAN2-PAN3 displays initial high deadenylation rates, which drop once the poly(A) tail is 20-30 nucleotides long. The resolved structure of the PAN2-PAN3 complex has shed a light on this peculiar catalytic activity. The PAN3 homodimer seems to have two ways of binding target mRNA. In its N-terminus, the zinc-finger domain ensures

specificity for poly(A) tail, while the PAM-2 motif binds PABP proteins. In the C-terminus, the pseudokinase domain displays ATP-binding residues, that seem to have a high effect on mRNA deadenylation *in vivo* (Christie et al. 2013). A different explanation for the ATP-binding pocket is for PAN3 to be able to recognise RNA molecules, and therefore to stimulate PAN2 activity. ATP presence inhibits the overall activity of the complex *in vitro* (Schäfer et al. 2014), possibly by competing with RNA to bind PAN3. Therefore, the pseudokinase domain is likely to provide low-specificity RNA binding (Wolf et al. 2014). Finally, the C-terminal region of the PAN3 homodimer, by recruiting the WD40 and the linker region of PAN2, allows PAN3 to direct the active exonuclease domain of PAN2 towards the substrate (Fig. 1.3). To support this hypothesis, it has been shown that PAN3 can bind PAN2 and PABP simultaneously (Uchida et al. 2004; Siddiqui et al. 2007; Huang et al. 2013). Interestingly, the presence of PABP on the mRNA is not essential for deadenylation *in vitro* (Schäfer et al. 2014). Once the poly(A) tail is shortened, the zinc-finger reaches the 3'-UTR region of the mRNA, for which has a low affinity, and this is likely to disengage the complex from the substrate. This may be a further control system to enhance PAN2 selectivity towards its substrate.

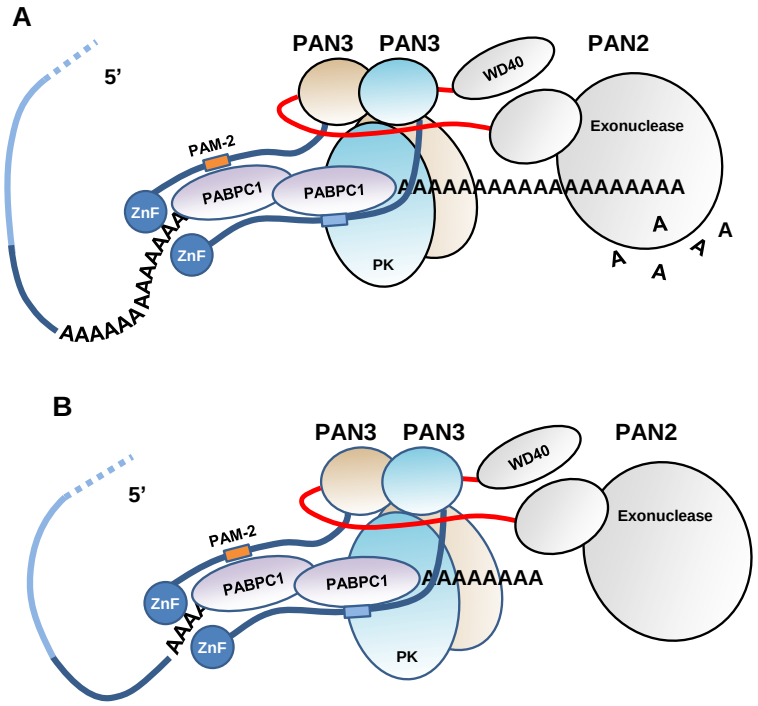



**Figure 1.3** – The generic recruitment of the PAN2-PAN3 complex. (A) PABPC1 stimulates the complex recruitment on the poly(A) tail via the PAM-2 motifs of PAN3. Additional specificity is given by the two zinc-fingers (ZnF) of the homodimer, that recognise the mRNA tail. (B) Both zinc-fingers can 'sense' the 3'-UTR region and cause the arrest of the machinery and the end of the deadenylation process.

### *1.4.4 Targeted recruitment mechanisms*

PAN3 was previously known to directly interact with the miRNA-mediated mRNA decay pathway via GW182/TNRC6 (Braun et al. 2011; Chekulaeva et al. 2011; Fabian 2011), although the structural characterisation of this interaction was unclear. Recently, the PAN3 homodimer assembly, which is not strictly required to bind PAN2 (Christie et al. 2013), has been found to form a W-pocket at the interface, involving well characterised residues in the pseudokinase domain. This pocket facilitates the recruitment of the PAN2-PAN3 by GW182 (Christie et al. 2013) and therefore establishes a link between this complex and the miRNA-driven mRNA decay machinery (Christie et al. 2013). Finally, it is plausible for PAN3 to undertake conformational changes that could either induce allosteric regulation mechanisms for the complex or cause its disassembly with drastic effects on the activity. Given the high impact of PAN3 on the activity of this deadenylation complex, it would be interesting to further investigate potential additional roles of PAN3 in the downstream steps of mRNA decay in eukaryotes.

## 1.5 Ccr4-Not complex

The Ccr4-Not (carbon catabolite repression – negative on TATA-less) complex is a large multisubunit mRNA decay machine conserved across the eukaryotes (Winkler & Balacco 2013). In *S. cerevisiae*, this L-shaped complex has a native molecular weight of approximately 1 MDa (Nasertorabi et al. 2011; J. Chen et al. 2001) (Fig. 1.4). The human complex contains five canonical subunits: CNOT1 (i) is the largest protein, CNOT2 (ii) and CNOT3 (iii) form the NOT domain and the catalytic subunits of the complex are Caf1/CNOT7/POP2 (Ccr4 associated factor 1) (iv) and Ccr4/CNOT6 (v) (Panepinto et al. 2013) (Fig. 1.4). CNOT1 acts as a scaffold; its absence causes complex dissociation *in vivo*, resulting in lower deadenylation activity and P-bodies formation (Ito, Takahashi, et al. 2011). However, along with its key structural role, it is also directly involved in several recruitment mechanisms. CNOT1 presents three main regions, which reveal the modular nature of this complex (Bawankar et al. 2013). The N-terminus, rich in HEAT repeats, is bound by CNOT11, that engages interactions with CNOT10 (Bawankar et al. 2013) (Fig. 1.4). These proteins form the N-terminal module, which is not required for viability in *S. cerevisiae*. A central region of CNOT1 contains two conserved domains; DUF3819 (domain of unknown function) binds CNOT9/ RQCD1/Caf40 (Chen et al. 2014; Mathys et al. 2014) and forms the CNOT9 module (Fig. 1.4) while MIF4G (Middle domain of eukaryotic Initiation Factor 4G) is recognised by Caf1 that tethers Ccr4 to the complex (Bawankar et al. 2013; Basquin et al. 2012; Petit et al. 2012) and forms the nuclease module (Fig. 1.4). Lastly, the unstructured C-terminus of CNOT1 is recognised by CNOT2 and CNOT3 and forms the Not-module (Boland et al. 2013) (Fig. 1.4).

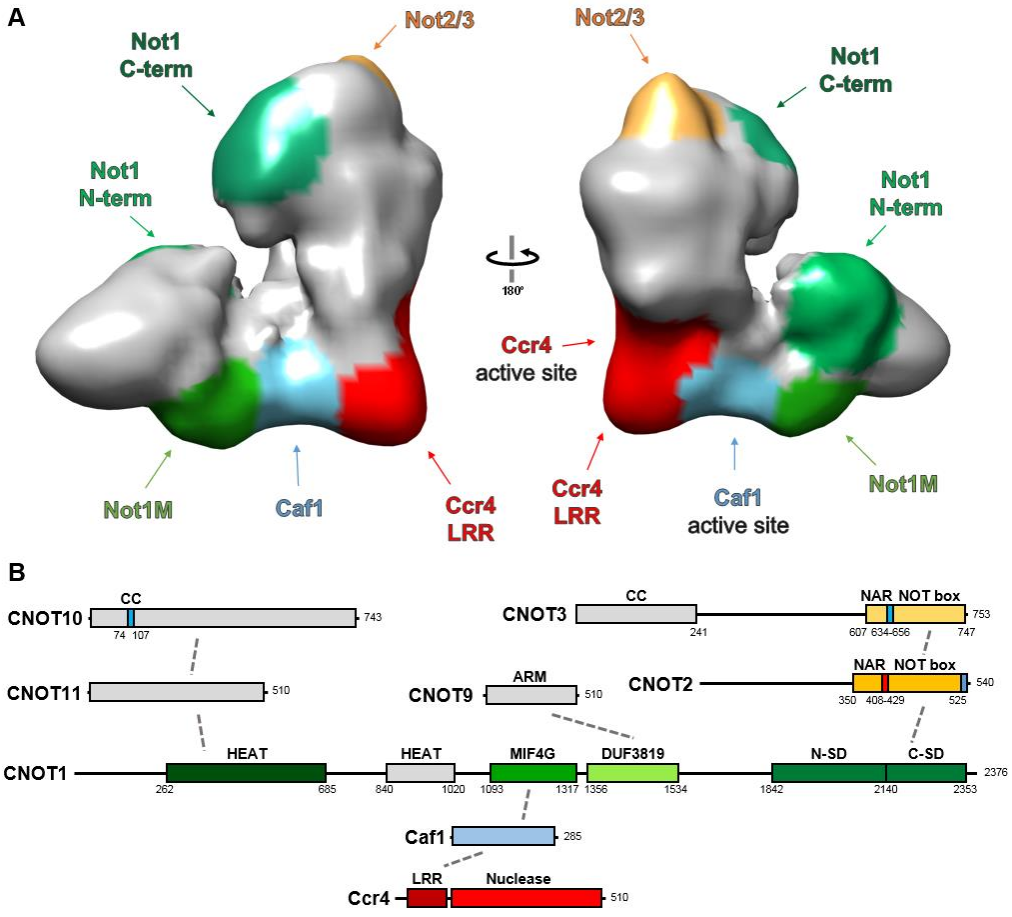



**Figure 1.4** – The Ccr4-Not complex. (A) The electron microscopy density map of the *S. cerevisiae* Ccr4-Not complex (Nasertorabi et al. 2011). Due to the low-resolution, the relative positions of only a few subunits can be identified. (B) A schematic overview of the interactions within the human Ccr4-Not complex.

### *1.5.1 The N-terminal module*

CNOT10 and CNOT11 are non-canonical subunits of the Ccr4-Not complex. CNOT11/C2ORF29 is an established component of the human Ccr4-Not complex (Lau et al. 2009). Its structural role is to tether CNOT10 to the N-terminus of CNOT1, rich in HEAT repeats (Wahle & Winkler 2013; Mauxion et al. 2013; Bawankar et al. 2013). However, depletion of CNOT11 has shown to have no effect on deadenylation rates in vivo (Mauxion et al. 2013). Conversely, CNOT10 depletion significantly affects deadenylation in lower eukaryotes, where it also alters the stability of the complex (Färber et al. 2013).

### *1.5.2 The Nuclease module*

The nuclease module of the Ccr4-Not complex is composed by two deadenylase enzymes Caf1 and Ccr4. Caf1 recognises the MIF4G domain of CNOT1 and tethers Ccr4 (Fig. 1.5). Ccr4/CNOT6 is one of the two active deadenylases of the complex. It is a $Mg^{2+}$-dependent EEP (endonuclease, exonuclease-phosphatase) enzyme (Winkler & Balacco 2013). There are two human paralog genes, *CNOT6/Ccr4a* and *CNOT6L/Ccr4b* (78% identical and 88% similar), mutually exclusive in the Ccr4-Not complex (Mittal et al. 2011; Lau et al. 2009; Winkler & Balacco 2013). Ccr4 contains the C-terminal deadenylase domain and an N-terminal LRR domain (leucine-rich repeat) that binds Caf1 active site, and not directly CNOT1 (Lau et al. 2009; Winkler & Balacco 2013). Caf1 is the second deadenylase enzyme of the Ccr4-Not complex (Wahle & Winkler 2013; Winkler & Balacco 2013). It belongs to the RNase D family, displaying a DEDD active site which coordinates two divalent metal ions (Fig. 1.5). In human, there are two paralog genes for this protein, *CNOT7* and *CNOT8* (76% identical and 89% similar at the amino acid level), encoding for Caf1a and Caf1b, mutually exclusive in the complex (Winkler & Balacco 2013; Doidge et al. 2012; Lau et al. 2009). Caf1 binds Not1 on its MIF4G domain (Basquin et al. 2012; Petit

et al. 2012; Bawankar et al. 2013) (Fig. 1.5). Interestingly, the active sites of Ccr4 and Caf1 point away when they are in complex, suggesting a spatial organisation, possibly triggered by co-factors, that would explain an apparent redundancy (Wahle & Winkler 2013) (Fig. 1.5). A second explanation could be an allosteric regulation for the two deadenylases, that act in a cooperative fashion (Maryati et al. 2015). Ccr4 active domain shares its structure with other components of the EEP superfamily, and has a strong preference for poly(A) polymers *in vitro* (Wang et al. 2010). Both CNOT6 and CNOT6L have a detectable effect on cell growth (Mittal et al. 2011); however, *ΔCNOT6L* cells display a stronger effect (Mittal et al. 2011) possibly related to only partially overlapping targets. Finally, CNOT6 and CNOT6L show different tissue expression levels (Chen et al. 2011), supporting potential roles in tumourigenesis that need to be clarified (Gutierrez-Camino et al. 2014). Similarly, Caf1 represses by deadenylation the anti-proliferative gene PMP22 in MCF7 cells (Aslam et al. 2009).

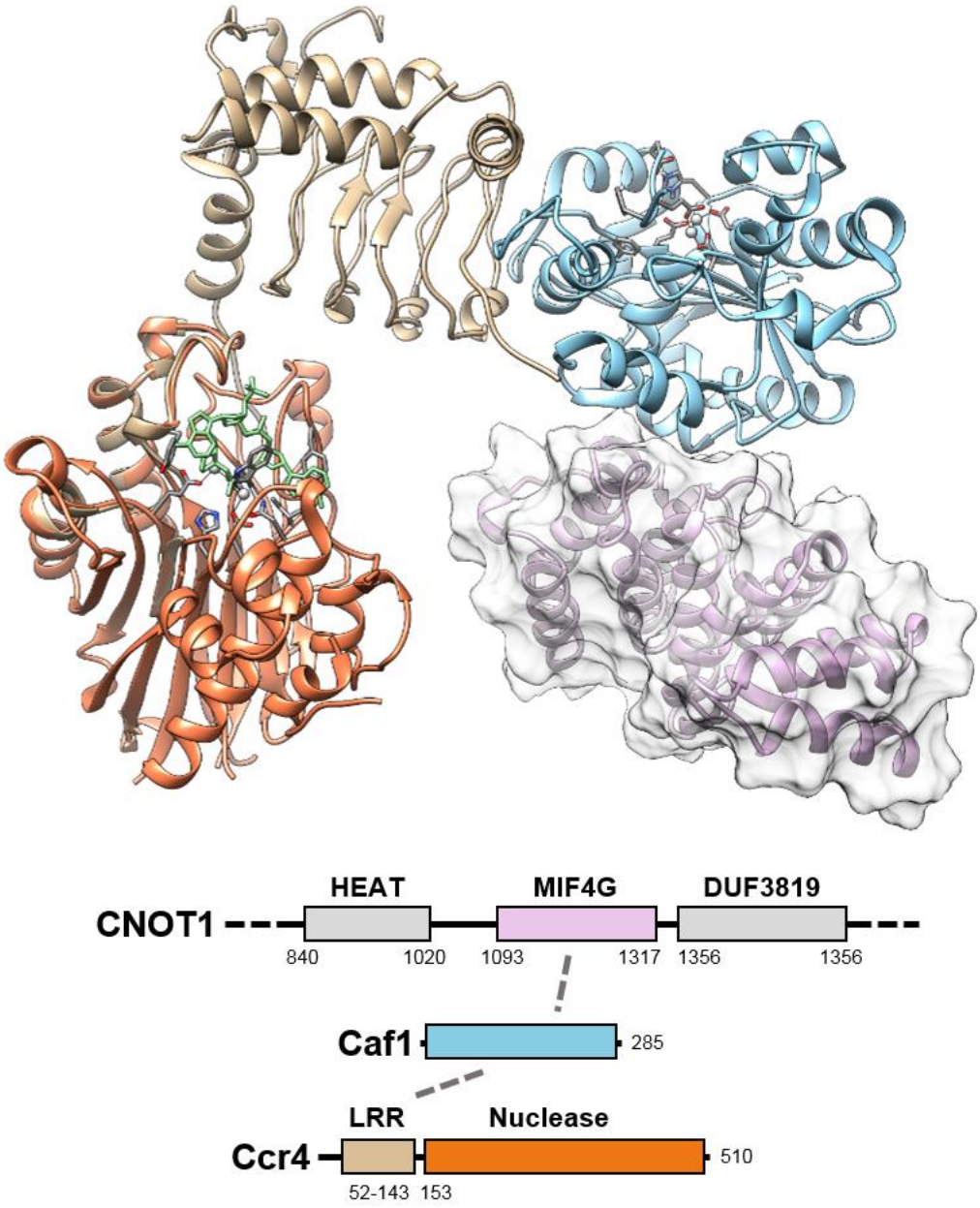



**Figure 1.5** – The deadenylase module of the *S. cerevisiae* Ccr4-Not complex. The active deadenylases Caf1 (blue) and human Ccr4 (red) (Wang et al. 2010) are tethered to the complex via the interaction between Caf1 and the MIF4G domain of CNOT1 (pink) (Basquin et al. 2012).

### *1.5.3 The CNOT9 module*

CNOT9, also known as RQCD1 (Required for Cell Differentiation 1) or Caf40, is one of the most important canonical non-catalytic subunit of the Ccr4-Not complex. CNOT9 is a regulatory subunit acting as a transcriptional co-factor in embryos development (Hiroi et al. 2002) and is involved in growth control and cells differentiation (Mahajan et al. 2002). More recently, CNOT9 has been associated with tumourigenesis (Ajiro et al. 2009; Ajiro et al. 2010). It contains no active site and almost the entire length of the protein is represented by six Armadillo-like repeats (ARM). These highly conserved structures promote protein-protein interactions and could also facilitate CNOT9 recruitment to RNA-binding complexes, as for other proteins with similar structures (Kim et al. 2012). Interestingly, the ARM repeats allow CNOT9 to either form homodimers or to associate with the DUF3819 domain of CNOT1 (Chen et al. 2014; Mathys et al. 2014; Garces et al. 2007) (Figure 1.6). Moreover, upon binding the DUF3819 domain of CNOT1, a structural rearrangement of CNOT9 exposes tryptophan-binding pockets to engage GW182, indicating a key regulatory role in miRNA-mediated mRNA decay (Chen et al. 2014).

Recently, a recurrent CNOT9 P131L variant has been identified in a whole-exome sequencing study of primary cutaneous melanomas (Wong et al. 2015). Whether this missense mutation is having an impact on the structure or the biochemical activity of this protein is still unclear; part of this study focused on shedding a light on this variant.

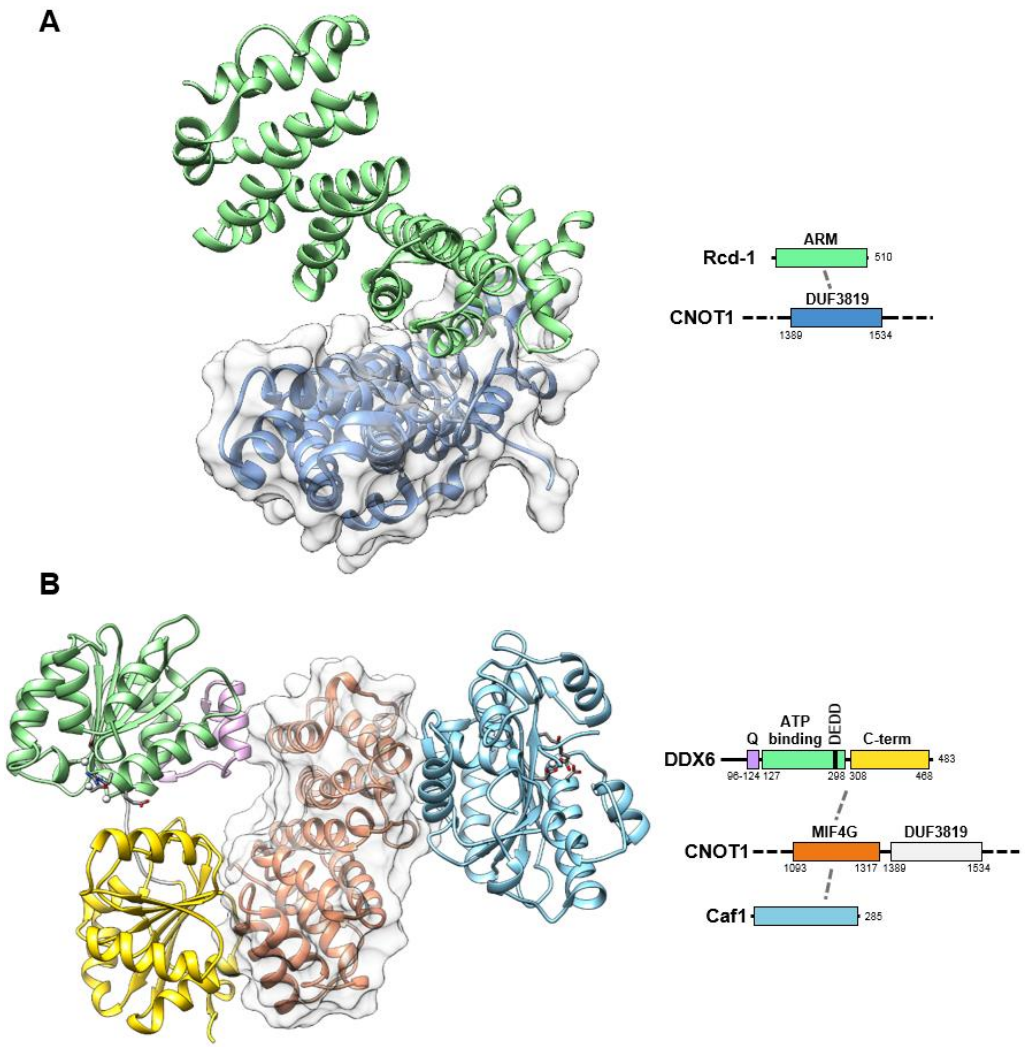



**Figure 1.6** – Structural representations of the central domains of CNOT1. (A) The human CNOT9 module of the Ccr4-Not complex (Chen et al. 2014). A conserved domain of unknown function (DUF3819, in blue) is recognised by the ARM region of CNOT9 (green). (B) The human MIF4G domain (orange) of CNOT1 tethers Caf1 (cyan) and DDX6 (left, multiple domains) simultaneously (Chen et al. 2014; Petit et al. 2012).

### *1.5.4 The NOT module*

The NOT module of the Ccr4-Not complex is composed by the CNOT2-CNOT3 heterodimer. These two conserved proteins share a similar structure; they display a Not-box at their C-terminus (Zwartjes et al. 2004), which is responsible, within a larger region, for the heterodimer assembly (Bawankar et al. 2013; Boland et al. 2013) (Fig. 1.7). More specifically, CNOT2 binds CNOT1, and tethers CNOT3 to the complex (Lau et al. 2009; Ito, Inoue, et al. 2011; Boland et al. 2013). Neither of them displays activity *in vitro*, but they are involved in different mRNA decay regulatory processes in eukaryotes (Temme et al. 2010; Neely et al. 2010; Morita et al. 2011). In particular, CNOT3 appears to be involved in tumour suppression and several human diseases (Neely et al. 2010; Díaz-Peña et al. 2012), although its role in mRNA decay needs further clarification. Conversely, CNOT2 involvement in transcription regulation is more evident (Jayne et al. 2006). Indeed, CNOT2 depletion has a strong effect on complex formation *in vivo*, affects mRNA deadenylation and causes a reduction of the number of P-bodies (Ito, Inoue, et al. 2011).

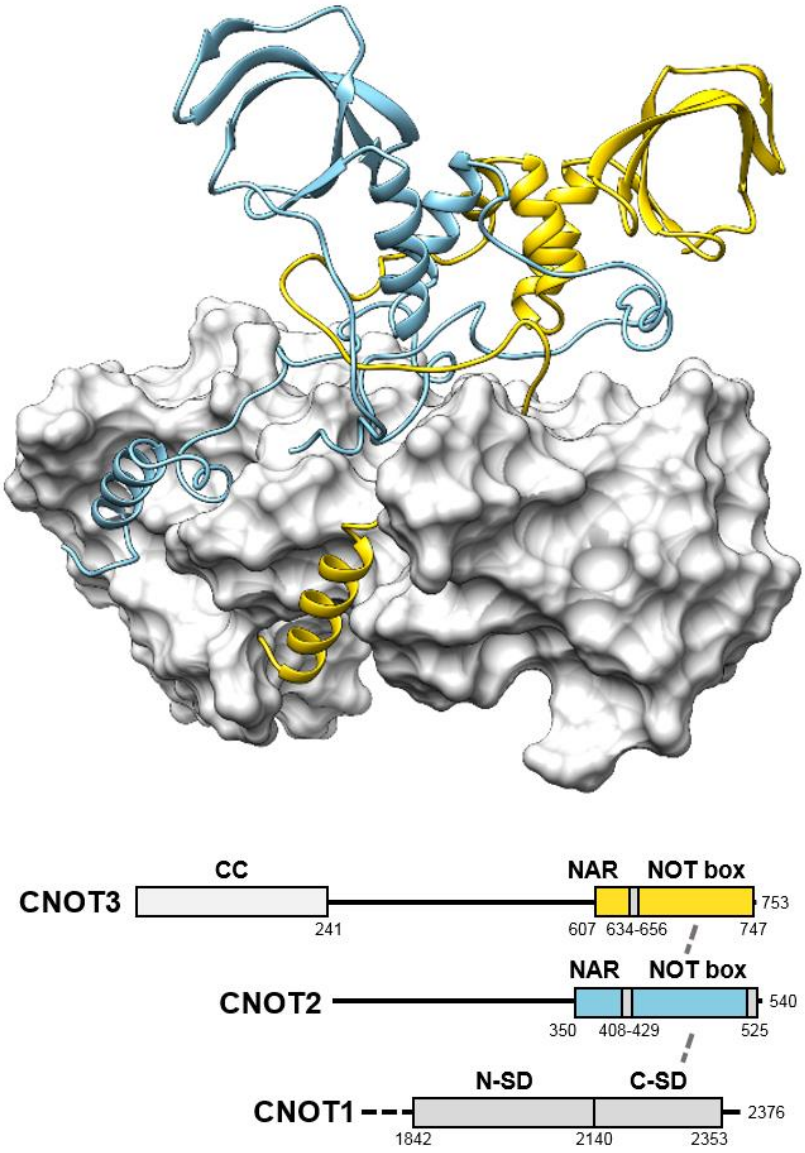



**Figure 1.7** – The Not-module of the human Ccr4-Not complex. The C-terminal region of CNOT1 (grey) tethers the NOT-BOX domains of the heterodimer formed by CNOT2 (light blue) and CNOT3 (yellow) (Boland et al. 2013).

### *1.5.5 Recruitment of the Ccr4-not complex*

The mechanisms to target the Ccr4-Not complex to a specific mRNA have been extensively studied in the last decade. There are three main mechanisms; (i) a generic recruitment via PABPC1, (ii) a targeted recruitment mediated by RNA-binding proteins and (iii) the miRNA-mediated recruitment. Recently, an additional recruitment mechanism was established for the decay of $m^6A$-containing mRNAs. The Ccr4-Not complex is tethered to these substrates via a direct interaction between its C-terminal domain and the N-terminal region of YTHDF2, which recognises $m^6A$ modifications, found to promote deadenylation (Du et al. 2016).

### *1.5.6 Recruitment via PABPC1*

The first type of recruitment is mediated by PABPC1, virtually present on all mRNAs (Deo et al. 1999). PABPC1 is generally recognised by a PAM-2 motif, which is not present in any of the Ccr4-Not subunits. However, the interaction with PABPC1 is ensured by the BTG/Tob proteins, co-factors of the Ccr4-Not complex, via BTG/APRO (AntiPROliferative) domains that bind PABPC1 RRMs (RNA recognition motif) (Stupfler et al. 2016).

The BTG/Tob family is composed of six proteins: BTG1, BTG2/Tis21/PC3, BTG3/Ana, BTG4/PC3B, Tob1 and Tob2. They share a similar structure, a conserved portion of 104-106 amino acids identified as the BTG domain, corresponding to the N-terminal of the protein; the globular BTG domain displays five α-helices and four β-sheets (Mauxion et al. 2009). The N-terminal domains are conserved between Tob1 and Tob2, and BTG1 and BTG2. The BTG domain, which shares no similarity with any other known sequence (Mauxion et al. 2009), has shown protein-protein interactions with the deadenylase subunits of the Ccr4-Not complex (Horiuchi et al. 2009) (Fig. 1.8A). This interaction is required for their anti-proliferative activity (Doidge et al. 2012; Ezzeddine et al. 2012; Horiuchi et al. 2009).

Specifically, the heterodimer formed by BTG1/BTG2 and Caf1 has been associated with enhanced deadenylation activity in presence of PABPC1, which is not detected with an altered BoxC of BTG2 that disrupts the BTG2-PABPC1 interaction (Stupfler et al. 2016). In particular, the first two RRM of PABPC1 are sufficient for simultaneous binding of BTG2 and the poly(A) tail (Stupfler et al. 2016). Conversely to BTG2, Tob displays nanomolar affinity for the C-terminal region of PABPC1, via two conserved PAM2 motifs present within its unstructured C-terminus (Ruan et al. 2010). While it is known that BTG1 and BTG2 bind PABPC1 with their BoxC domain and Tob1 and Tob2 interact with PABPC1 with their PAM2 motifs, it is still unclear what regions of BTG3 and BTG4 are engaged in PABPC1 binding.

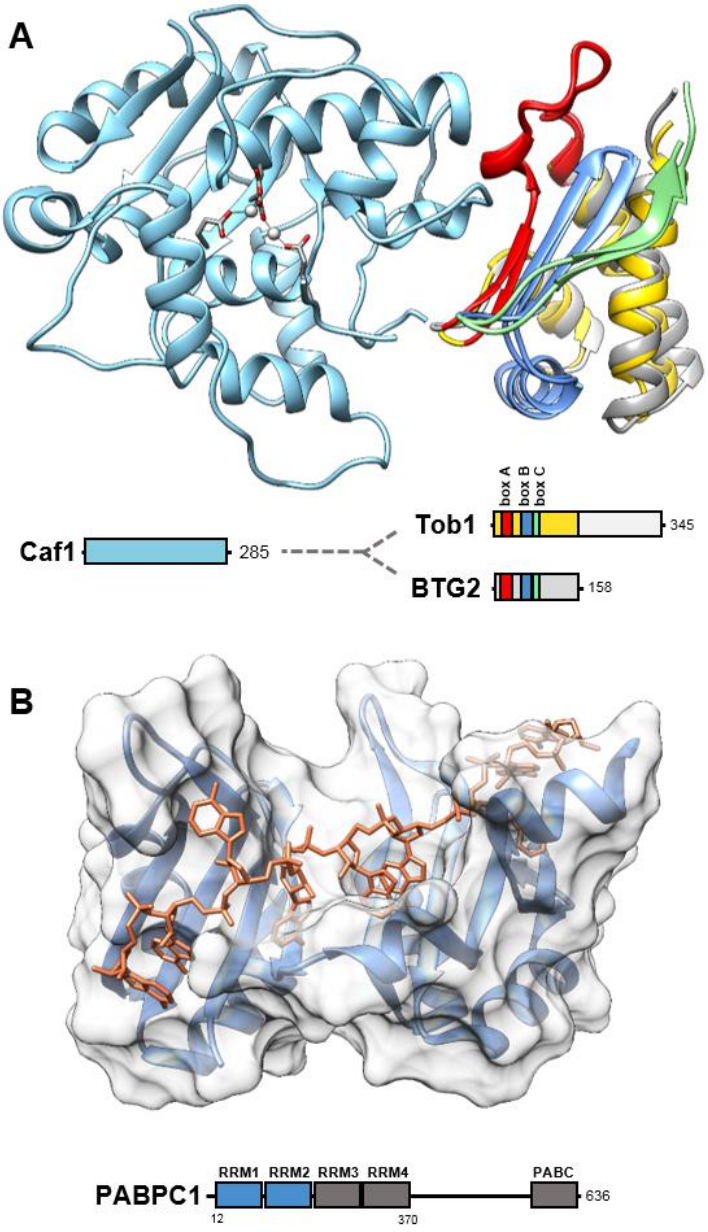



**Figure 1.8** – Structural representations of human Caf1, BTG/Tob proteins and PABPC1. Caf1 (cyan) is recruited and directed to specific targets via its interaction with the antiproliferative proteins of the BTG/Tob family (BTG2 in grey and Tob1 in yellow), mediated by the boxes A (red), B (sky blue) and C (green) (Horiuchi et al. 2009; Yang et al. 2008). (B) BTG2 directly interacts (crystal not available) with specific domains of human PABPC1, RRM1 and RRM2 (blue), also responsible for a cooperative binding of the poly(A) tails (orange) of virtually all mRNAs in the cytoplasm (Safaee et al. 2012).

### *1.5.7 miRNA-mediated recruitment*

CNOT1 also mediates the Ccr4-Not recruitment in the miRNA-driven mRNA decay pathways. GW182, an effector of the miRISC complex, directly interacts with CNOT1, to tether the complex to the Argonaute protein (Braun et al. 2011; Fabian 2011; Chekulaeva et al. 2011). In particular, GW182 contains two LWG-repeats motifs in the C-terminal domain, CIM-1 and CIM-2, both binding CNOT1 (Fabian 2011) (Fig. 1.9B). Therefore, GW182 can be considered the link between miRNA regulation and mRNA degradation. Once a transcript is targeted by a particular miRNA bound to an Argonaute protein, GW182 is recruited by AGO1 and leads to mRNA target decay by activating the Ccr4-Not complex (Behm-Ansmant et al. 2006). Depletion of AGOs or GW182 causes similar changes in target transcripts levels, suggesting analogous impacts on the miRNAs regulation mechanism (Fabian 2011; Chekulaeva et al. 2011; Braun et al. 2011). Specifically, when CNOT9 is docked on the DUF3819 of CNOT1, conformational changes expose W-pockets on the surface of CNOT9 that are recognised by GW182 (Chen et al. 2014).

### *1.5.8 Protein-mediated recruitment*

There are several examples of mechanisms to recruit the Ccr4-Not complex on specific mRNAs by RNA-binding proteins (Fig. 1.9C). For instance, AREs (A/U rich elements), present in most mRNA displaying high turnover rates, are recognised by tristetraprolin (TTP), which binds the MIF4G domain of CNOT1 to direct Caf1 activity (Sandler et al. 2011; Fabian et al. 2013) (Figure 10B). The Not-module, composed by the C-terminus of CNOT1, CNOT2 and CNOT3, is a binding partner for Nanos proteins, which mediate translation repression and murine germ cell development (Suzuki et al. 2012; Suzuki et al. 2010; Bhandari et al. 2014). Nanos1 binds CNOT1 via a conserved N-terminal CNOT1-interacting motif (NIM), that is composed of aromatic residues inserted in a hydrophobic pocket of CNOT1 C-terminal

domain (Bhandari et al. 2014) (Figure 1.11A). Recently, Roquin1 and 2 have been shown to promote mRNA degradation by recognising stem-loop structures in the 3'UTR of mRNA target via their conserved ROQ domain (Schlundt et al. 2014; Janowski et al. 2016) (Fig. 1.10). In particular, Roquin1 recruits the Ccr4-Not complex with its non-conserved C-terminal region (Sgromo, Raisch, Bawankar, et al. 2017) (Fig. 1.10). The mechanism is better characterised in *D. melanogaster*, where the CNOT9/Rcd-1/Caf40 binding motif (CBM) of Roquin recognises a conserved surface region of CNOT9 (Sgromo, Raisch, Bawankar, et al. 2017). Other proteins are known to bind target mRNAs. For instance, PUM1 (Pumilio) proteins that recognise the 3' UTR of specific mRNAs are associated with poly(A) shortening and mRNA decay (Goldstrohm et al. 2006) (Fig. 1.11). Remarkably, PUM1 is known to directly interact with Caf1 and therefore to recruit the Ccr4-Not complex in yeast and humans (Goldstrohm et al. 2006). PUM1 displays an interesting structure; its C-terminus is composed by several α-helices that form a concave space where it can bind 15-20 ribonucleotides (X. Wang et al. 2002). Pumilio and Nanos directly interact in *D. melanogaster*, where Nanos enhances Pumilio RNA binding and facilitates mRNA decay (Weidmann et al. 2016).

In yeast, the activity of the Ccr4-Not complex is regulated by Mmi1, a mRNA-binding protein often found associated with the complex, which is able to bind DSR motifs (Stowell et al. 2016). Several proteins are co-purified with Ccr4-Not, and although regulation mechanisms might be similar to Mmi1, further investigation is required to address them.

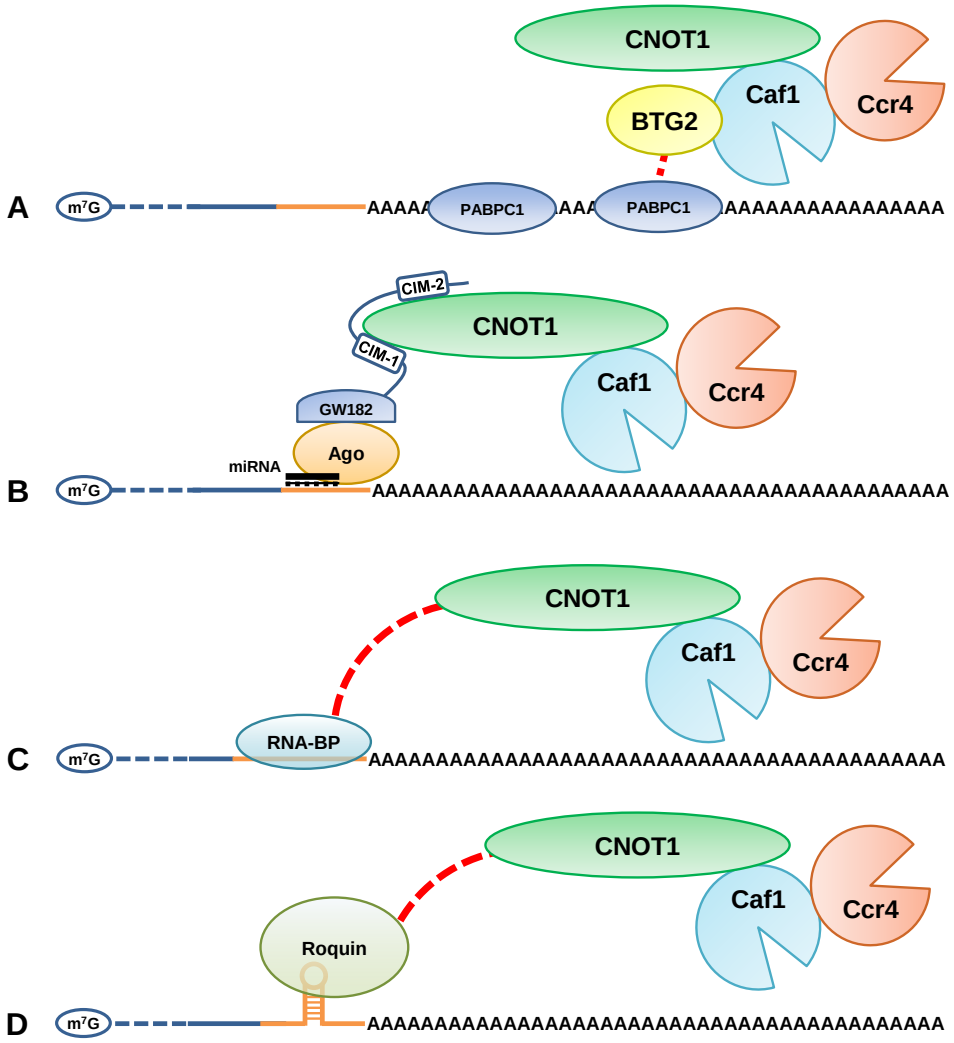



**Figure 1.9** – Deadenylation by Ccr4-Not is triggered by three recruitment mechanisms. The 'canonical' recruitment via PABPC1 (A), miRNA mediated mRNA decay involving GW182 (B), several different RNA-binding proteins (C) and the Roquin protein recognising hairpins on the 3'UTR (D).

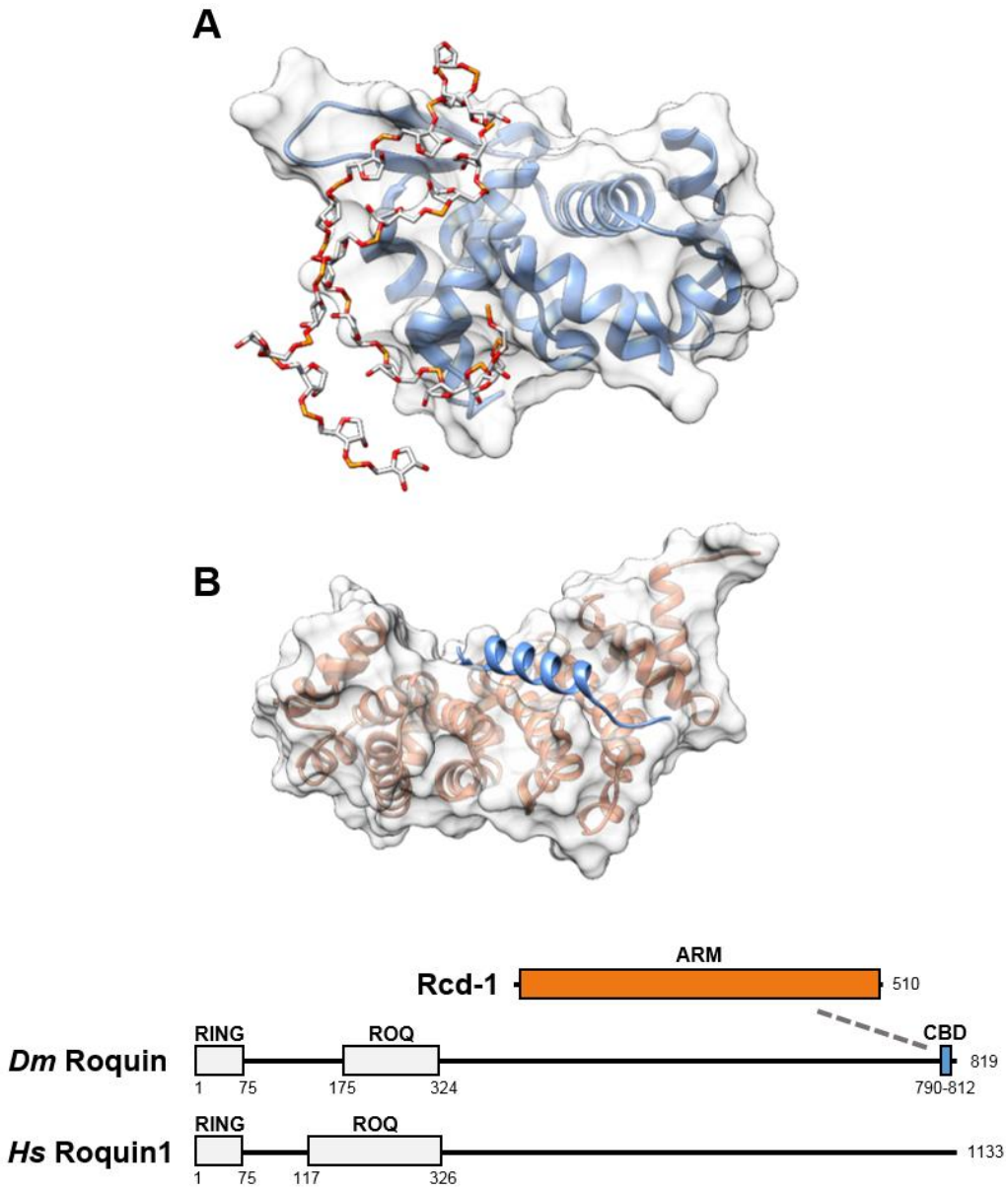



**Figure 1.10** – Roquin1 interactions. (A) Human Roquin1 (grey) recruits the Ccr4-Not complex to mRNA targets that display stem-loop structures on their 3' UTRs (Tan et al. 2014). (B) The structure of a C-terminal fragment of *D. melanogaster* Roquin (light blue) interacting with human Rcd-1 (orange) was recently resolved, unveiling the interactions between the regulatory protein and the Ccr4-Not complex (Sgromo, Raisch, Bawankar, et al. 2017).

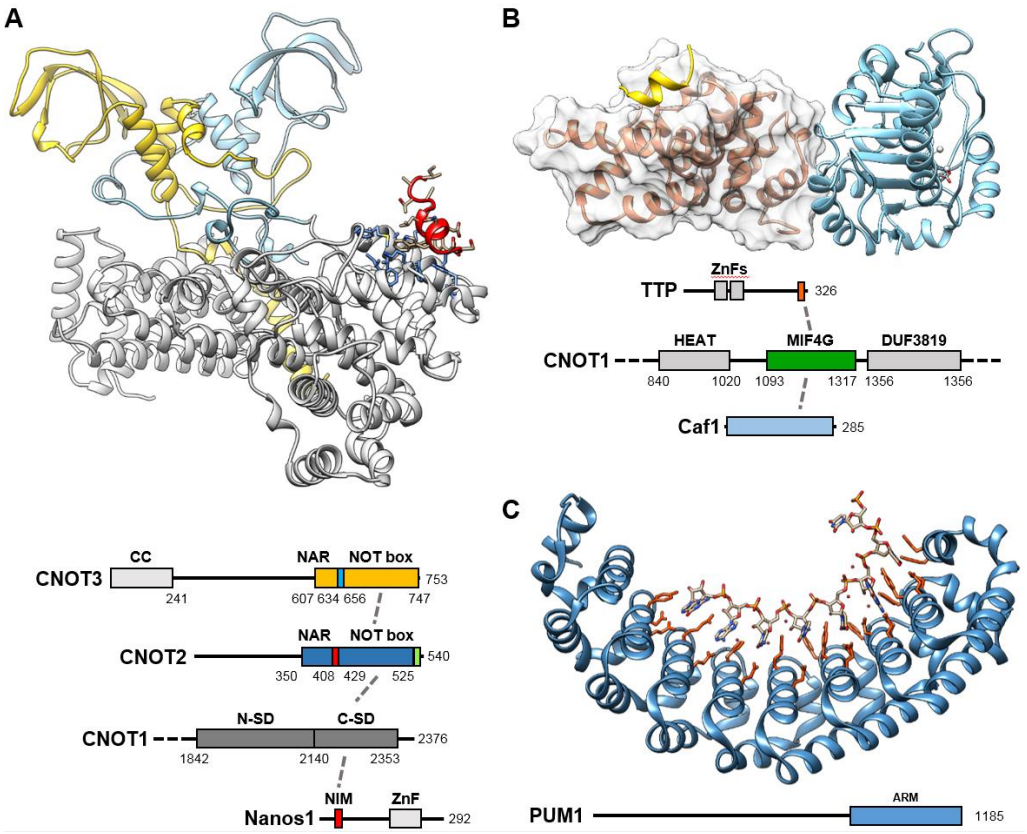



**Figure 1.11** – Regulatory mechanisms of the Ccr4-Not complex. (A) Structure showing the docking position of a human Nanos1 fragment (red) (Bhandari et al. 2014) onto the human NOT module of the Ccr4-Not complex, composed by the C-terminus of CNOT1 (gray), CNOT2 (blue) and CNOT3 (orange) (Boland et al. 2013). (B) Structure showing the relative positions of a human TTP fragment (orange), MIF4G domain of CNOT1 (green) and Caf1 (light blue) (Fabian et al. 2013; Petit et al. 2012). TTP recognises ARE elements on the 3'UTR of specific mRNAs and by tethering the MIF4G domain of CNOT1 indirectly recruits Caf1 for deadenylation. (C) Cartoon showing the structure of *D. melanogaster* Pumilio (blue) binding RNA (Weidmann et al. 2016).

## 1.6 Decapping

The 5' end cap present on virtually all mRNAs in human cells, with few exceptions, is a nucleotide-based structure responsible for mRNA protection and circularisation. It consists of a guanine linked to the mRNA 5'UTR via an unusual 5'–5' triphosphate bond resistant to most ribonucleases; after capping, the nucleotide is methylated on position 7 by a methyltransferase, leading to the $m^7G$ mature cap. mRNA decapping is the removal of this structure by hydrolysis from the 5' end of mRNA destined to decay.

### *1.6.1 The Lsm1-7-Pat1 complex*

The Lsm1-7 complex plays a key role with Pat1 in the 5'-3' end mRNA degradation pathway by linking deadenylation with decapping (Bouveret et al. 2000; Chowdhury et al. 2007; Haas et al. 2010; Sharif & Conti 2013). The Lsm1-7 complex subunits form a heptameric ring (Fig. 1.12A) that binds short poly(A) oligonucleotides *in vivo* (Tharun et al. 2000; Tharun & Parker 2001) and colocalises to P-bodies, the cytoplasmic foci for mRNA processing (Pilkington & Parker 2008; Haas et al. 2010; Ozgur et al. 2010; Ingelfinger et al. 2002). Each subunits displays a conserved Sm-like domain, flanked by amino acid chains at the N-terminus and C-terminus. The Lsm1 subunit, only present in the cytoplasmic complex, displays a longer C-terminal extension (Fig. 1.12A). The structural feature, not present in the other six subunits, is composed by a long α-helix that has contacts with both sides of the ring (Sharif & Conti 2013). This structure contributes to RNA binding (Tharun et al. 2005; Chowdhury & Tharun 2008; Chowdhury et al. 2012; Chowdhury et al. 2014), although the α-helix possibly prevents the RNA chain to enter the channel (Sharif & Conti 2013). Pat1 is tethered to the complex by the Lsm2 and Lsm3 subunits (Sharif & Conti 2013; Wu et al. 2014) and actively participates to RNA binding (Chowdhury et al. 2014) (Fig. 1.12B). Pat1 is an important regulator in the degradation pathway

given its recruitment capabilities. Pat1 interactions have been detected with XRN1 in yeast (Bouveret et al. 2000; Chowdhury et al. 2007) and with the Ccr4-Not complex in *D. melanogaster* (Haas et al. 2010). Intriguingly, the MIF4G domain of CNOT1M binds and stimulates DDX6, an RNA helicase involved in miRNA regulation, which also activates the decapping pathway and represses translation (Chen et al. 2014; Mathys et al. 2014) (Figure 12). In addition, direct links have been established between Pat1 and the decapping complex Dcp1-Dcp2 (Pilkington & Parker 2008; Ozgur et al. 2010; Nissan et al. 2010; Ingelfinger et al. 2002) and with the decapping activator DDX6 (Haas et al. 2010; Braun et al. 2010; Nissan et al. 2010; Ozgur et al. 2010; Sharif et al. 2013), suggesting multimeric complexes forming on mRNAs destined to decay.

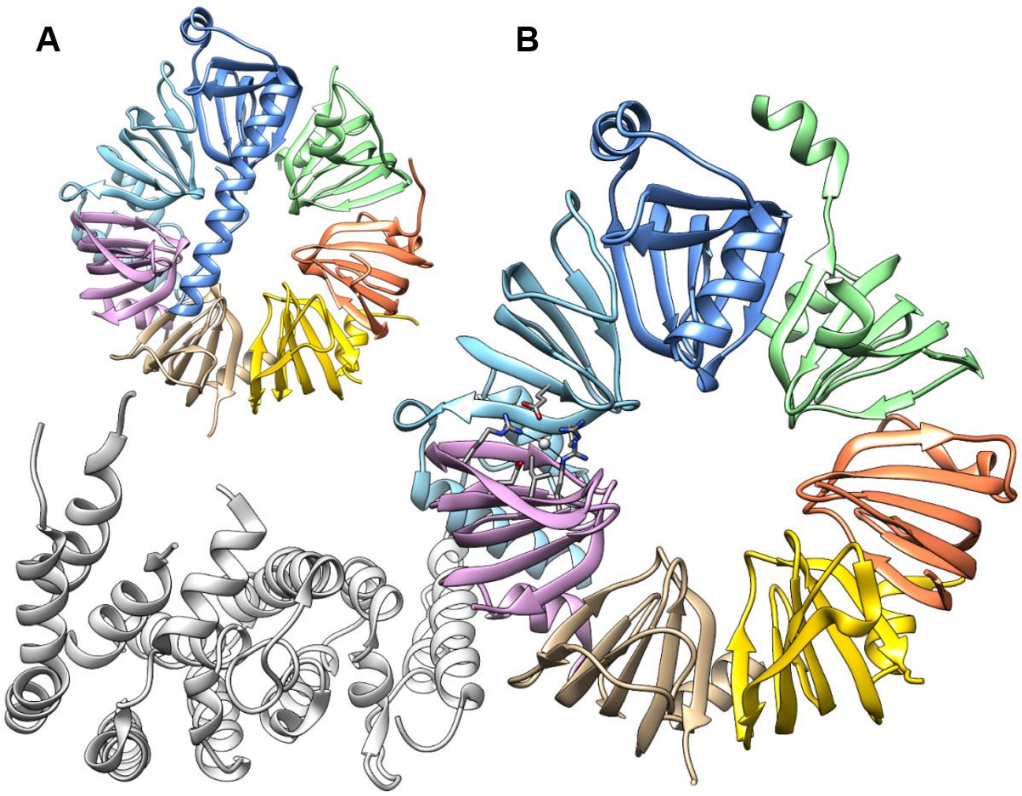



**Figure 1.12** – The Lsm1-7 complex. (A) Structure of the *S. cerevisiae* Lsm1-7 apo-complex (Sharif & Conti 2013). Lsm1 (blue) displays a long α-helix that has contacts with the opposite side of the heptameric ring. (B) Structure of the *S. cerevisiae* Lsm1-7-Pat1 complex (Sharif & Conti 2013). The regulatory protein Pat1 (red) interacts with the ring via Lsm2-3 (light blue and cyan, respectively).

### *1.6.2 The Dcp1-Dcp2 complex*

In eukaryotes, Dcp1-Dcp2 is the key heterodimeric complex for mRNA decapping. Dcp2, a 42.3 KDa enzyme, is the catalytically active subunit of the complex. It displays a Nudix hydrolase domain, that contains a 23 amino acids consensus Nudix motif ($GX_5EX_7REUXEEXGU$), which includes two glutamic acids that can coordinate metal ions to remove the cap structure (Z. Wang et al. 2002). This cytoplasmic enzyme has high specificity for capped RNAs, as it is incapable of binding the cap alone of unmethylated cap structures, in contrast with the scavenger decapping enzyme DcpS, reviewed elsewhere (Milac et al. 2014). A specific region of Dcp2 (residues 94-257) has an intrinsic ability to bind RNA, required for its decapping activity *in vitro* (Piccirillo et al. 2003). In addition, Dcp2 activity *in vitro* is greatly increased in presence of $Mn^{2+}$ ions, and it displays a positively charged BoxB region for RNA binding (Piccirillo et al. 2003).

In yeast, several conformations of the Dcp2p-Dcp1p complex have been detected, shading light on its mechanism of action (She et al. 2008). The structure of Dcp2 reveals a flexible hinge that allows a major conformational change of the N-terminal domain, suggesting a potential regulatory step for mRNA decapping (Figure 13). The 'open' conformation of the complex, while maintaining the RNA-binding ability, has a considerably lower catalytic activity compared to the 'closed' structure (She et al. 2008). The conformational shift is stimulated by the presence of the cap and assisted by Dcp1, that interacts with the Nudix domain of Dcp2 (She et al. 2008). The Dcp2 structure also reveals a positively charged channel, presumably responsible for RNA binding, that corresponds to the Box B previously described (She et al. 2008; Deshmukh et al. 2008). In yeast, Dcp1 does not increase the affinity of Dcp2 for mRNA, but it is required for decapping *in vivo* (Beelman et al. 1996; Sakuno et al. 2004) and stimulates Dcp2 activity *in vitro* (Steiger et al. 2003). Recently, additional structures of

the Dcp1-Dcp2 complex have introduced regulatory steps of decapping. The *S. pombe* complex has been co-crystallised with a disordered peptide of Ecd1 that represents the conserved decapping-activator motif (DAM) displayed at the C-terminus of Ecd1, able to bind the EVH1 domain of Dcp1 via hydrophobic interactions (Valkov et al. 2016). This trimeric complex indicates an additional regulation step, as Ecd1 triggers a rotation of 120° of the active site of Dcp2 and the complex undergoes a conformational change from the 'closed' into the 'scanning' form. A positively charged cleft is now exposed and the substrate is guided towards the active site of Dcp2 (Valkov et al. 2016) (Fig. 1.13). Another structure of the Dcp1-Dcp2 complex has been resolved, displaying a fragment of the decapping co-activator PNRC2 and a 'two-headed' cap-analog with a 20-fold increased affinity (Mugridge et al. 2016). The mRNA cap induces an additional conformational change in the complex; the Nudix domain of Dcp2 rotates by 25° away from Dcp1, and this converges the three aromatic residues (W43, Y220 and Y92) responsible for cap-binding (Mugridge et al. 2016). This 'cap-bound' structure also displays a region for RNA binding, represented by a tunnel-like cavity formed by 25-30 lysine amino acids (Mugridge et al. 2016).

The available structures of the Dcp1-Dcp2 complex helped identifying several steps of the decapping process in eukaryotes, which may be unified in this model. Dcp1-Dcp2 shifts from 'open' to 'close' conformation upon RNA recognition and co-factors binding, then it 'locks' (120° rotation) onto the mRNA substrate and 'scans' it until it finds the cap at the 5' end. When the cap is recognised by the conserved residues an additional 25° rotation is detected. A possible further conformational change may take place when the Dcp1-Dcp2 complex is ready to hydrolyse the cap structure. To this end, a structure including an mRNA substrate is necessary.

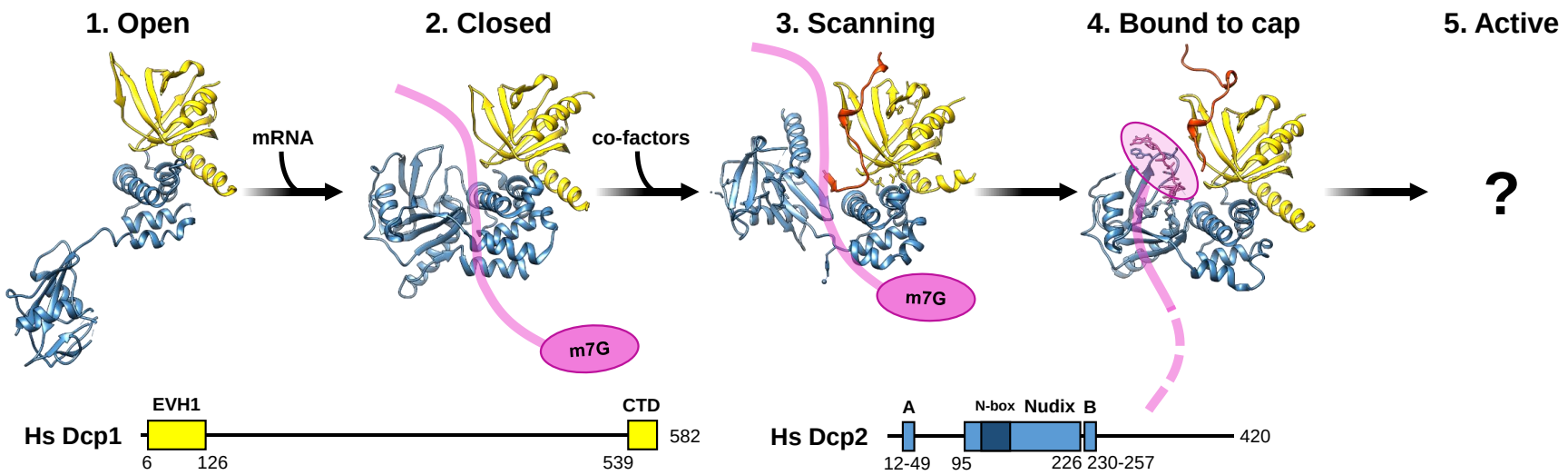



**Figure 1.13** – A proposed model for the step-wise regulation of the mRNA decapping mechanism. Several structures of the Dcp1-Dcp2 complex (yellow and blue, respectively) have been resolved in different conformational states. (1) The *S. pombe* complex 'open' conformation (PDB 2QKM) of the complex, that undergoes a major conformational change (2) upon mRNA binding (PDB 2QKM) (She et al. 2008). (3) When decapping co-activators bind the *S. pombe* complex (Edc-1, orange), the Nudix domain of Dcp2 rotates by 120° (PDB 5J3T) and the complex starts 'scanning' the substrate for the cap (Valkov et al. 2016). (4) Structure of the *S. pombe* complex bound to a cap analog (PDB 5KQ4). The Nudix domain has rotated by an additional 25° (Mugridge et al. 2016). (5) A crystal structure of the active state of the Dcp1-Dcp2 complex is missing.

## 1.7 Degradation

Deadenylation destines mRNAs to degradation for both the 5'-3' and the 3'-5' end pathways. However, the enzymes responsible for mRNA decay in the canonical eukaryotic pathways are also involved in the mRNA quality control mechanisms of Non stop decay (NSD) and No-go decay (NGD), discussed elsewhere. This suggests their central role in mRNA degradation.

### *1.7.1 XRN1*

In the 5'-3' the mRNA poly(A) tail shortening is directly linked to decapping. The Dcp1-Dcp2 complex generates 5'-monophosphorylated mRNA substrates that are promptly recognised and degraded by the XRN1 enzyme. However, this is not the only pathway where XRN1 is involved. Its function is required downstream nonsense and no-go decay pathways, where it degrades intermediates after endonucleolytic cleavage of mRNA molecules (Doma & Parker 2006; Eberle et al. 2009).

XRN1 is a highly conserved, 175 kDa exoribonuclease with 5'-3' processive activity (Stevens 2001), that requires divalent metal ions (Stevens 1980). It is specific for 5'-monophosphate RNA substrates (Pellegrini et al. 2008) and displays interactions with a number of protein complexes involved in mRNA regulation. For instance, its co-localisation with the Dcp1-Dcp2 and the Lsm1-7/Pat1 complexes is well established (Sheth 2003). The conserved catalytic site binds the RNA substrate and simultaneously performs hydrolysis and duplex unwinding (Jinek et al. 2011). Interestingly for crystallography purposes, a D207A mutation has been described as abolishing the exoribonuclease activity without affecting substrate binding (Shen et al. 1997).

The N-terminal of XRN1 displays a globular core including the catalytic domain (residues 1-674) (Fig. 1.14). The C-terminal is represented by three domains that stack over the catalytic domain. The first one is a β-barrel

(residues 675-799), followed by an α-helix domain that contains a winged helix (WH) motif. In a *D. melanogaster* XRN1 model (Jinek et al. 2011) this is positioned close to the active site, suggesting a potential role for regulating the accessibility of the catalytic pocket (Jinek et al. 2011). Finally, an SH3-like domain, that lacks essential residues for proline-rich motifs, is fundamental for the integrity of the complex, as its deletion results in XRN1 inactivation in yeast (Page et al. 1998). In the co-crystal of *D. melanogaster* XRN1 (D207A) with a DNA oligo, a substrate that XRN1 can degrade although at a slower rate (Stevens 2001), the first three DNA nucleotides are inserted in the active site, that can accommodate both purines and pyrimidines, while the rest of the oligo is disordered (Jinek et al. 2011). The active site also coordinates two $Mg^{2+}$ ions and interestingly is not able to accommodate 5'-groups with steric volume higher than a 5'-phosphate (Jinek et al. 2011), indicating the inability of XRN1 to bind, and therefore degrade, capped mRNAs or RNA duplexes. However, XRN1 can unwind short RNA duplexes when a sufficiently long single stranded RNA 5'-end is present (Jinek et al. 2011). The high processivity of XRN1 suggests the presence of a substrate translocation step in the catalytic cycle. A *D. melanogaster* XRN1 H41A mutant has been characterised as producing only partially degraded mRNAs, indicating the contribution of this residue, that engages stacking interactions with nucloetides, towards processivity (Jinek et al. 2011).

### *1.7.2 Exosome*

The exosome is a ribonuclease complex responsible for mRNA processing or degradation that is highly conserved across the species. The human complex exists in a few different isoforms according to its cellular localisation, and it is generally composed of 10-11 subunits. Two modules can be distinguished, different for protein composition and role.

The exosome core is the largest module and it includes nine subunits of about 20-50 kDa. The crystal structure of the core (Liu et al. 2006) reveals a lower ring represented by an hexamer of proteins homologous to bacterial Ribonuclease PH (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46 and Mtr3). The second structure of the core is the 'cap', composed by three proteins (Rrp4, Rrp40, Csl4) that display RNA binding motifs. Interestingly, these do not engage interactions with each other, but each tightly binds two subunits of the hexamer (Liu et al. 2006). The hexamer and the cap have no catalytic activity due to point mutations causing substitutions of the conserved active residues, but provide substrate selectivity by forming a central channel that can only sterically accommodate a single stranded RNA molecule (ssRNA).

The second module includes two large enzymes Dis3 and Rrp6 of about 100 kDa, which harbour the catalytic activity of the exosome complex. These associate with the core and give rise to three isoforms of the complex. Dis3 is a member of the RNR (ribonuclease R) superfamily (Zuo & Deutscher 2001) and the homology region displays a typical sequence of the domains. The RNR region contains the RNB domain, responsible for the processive exoribonuclease activity, found at the bottom of the channel where single strand RNA is threaded (Lorentzen et al. 2008; Frazão et al. 2006). Dis3 also contains a PIN domain, responsible for endonucleolytic activity (Lebreton et al. 2008) and for binding to the core via Rrp41 (Schneider et al. 2009; Schaeffer et al. 2009; Bonneau et al. 2009). Dis3 has an intrinsic capability to unwind RNA secondary structures, providing a few free nucleotides at the 3' end of the substrate (Lorentzen et al. 2008; Dziembowski et al. 2007). The second enzyme of this module is Rrp6, a member of the DEDD superfamily (Zuo & Deutscher 2001). Conversely to Dis3, Rrp6 harbours distributive activity, thanks to its more accessible active site, that coordinated two $Mg^{2+}$ (Januszyk et al. 2011; Midtgaard et al. 2006).

Moreover, its affinity for the exosome core is much higher than that of Dis3 (Dziembowski et al. 2007) and it shows a similar efficiency to uncoil RNA secondary structures (Januszyk et al. 2011).

As mentioned above, in human cells the exosome complex exists in three major isoforms, according to its localisation. Rrp6 is a generally nucleolar enzyme, therefore a core-Rrp6 complex is found in the nucleolus (i); Dis3 can be detected in both nucleus, forming a core-Rrp6-Dis3 complex (ii), and cytoplasm in a core-Dis3 isoform (iii). The stoichiometry of all the subunits appears to be the same in every cell compartment, thus the exosome is generally believed to be always assembled, with the Csl4 exception (Hernández et al. 2006).

In human cells, the exosome is a multi-protein conserved complex responsible for a broad range of RNA-related roles, from processing to surveillance. The core is the most important determinant for the complex specificity and mechanisms of action (Wasmuth & Lima 2012). In the cytoplasm, the exosome complex is the key component of at least two major processes. The Ski complex assists the exosome for its RNA surveillance role (Araki et al. 2001), whereas AU-rich elements (ARE)-binding proteins tether the multimeric exoribonuclease to mRNA targets for degradation (C. Y. Chen et al. 2001; Mukherjee et al. 2002).

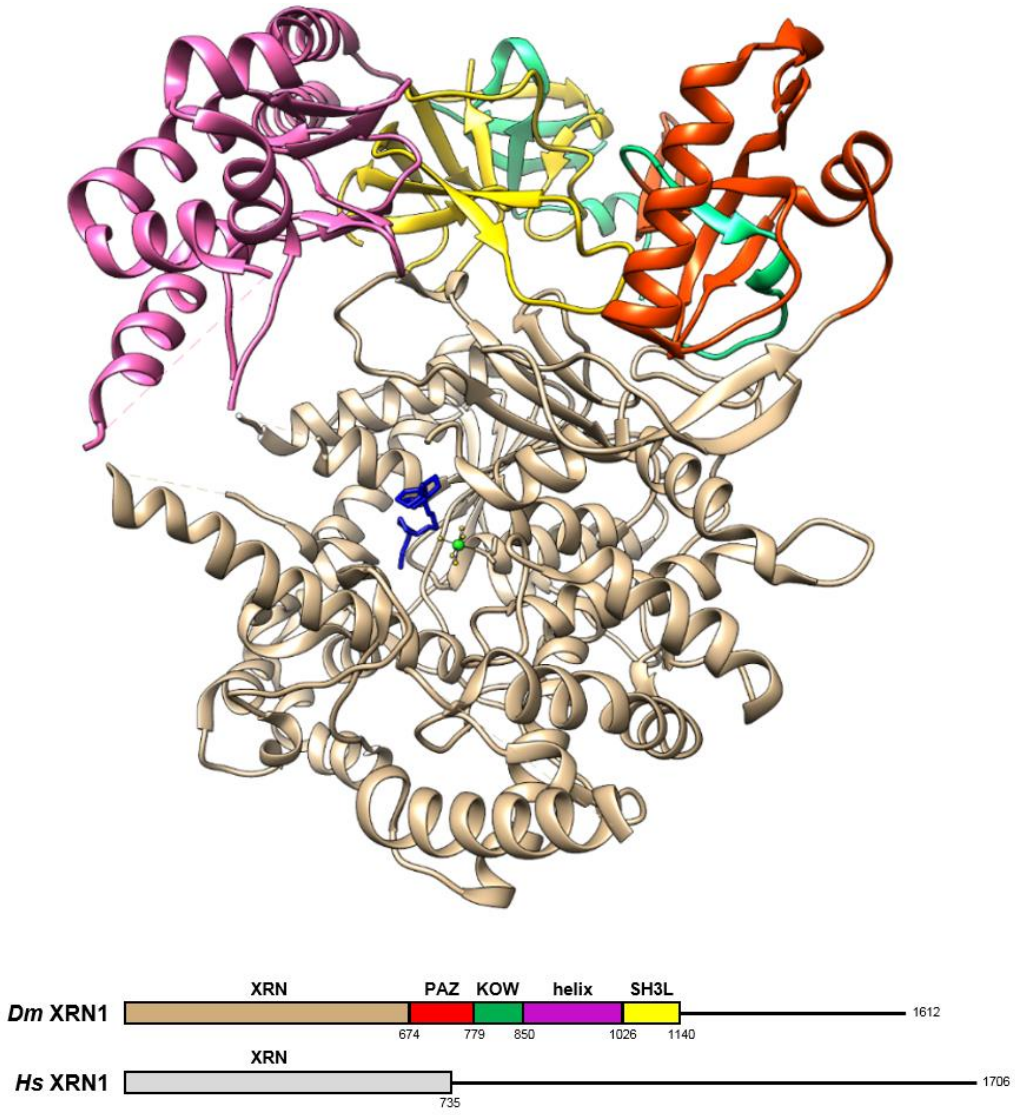



**Figure 1.14** – Structural representation of the *D. melanogaster* Xrn1 enzyme. The exoribonulcease domain (tan) contains the active site which binds a $Mg^{2+}$ ion (green) and the substrate (blue, shown in sticks) (Jinek et al. 2011). Human XRN1 and *D. melanogaster* Xrn1 are 37% identical and the XRN domains are 53% identical.

## 1.8 Aims of the study

Deadenylation, or shortening of the mRNA poly(A) tail, is the first and often rate-limiting step in mRNA decay in eukaryotes (Chen & Shyu 2011). In human cells, two complexes play a key role in this process, the Ccr4-Not and the PAN2-PAN3 complexes (Wahle & Winkler 2013). Since their discovery, they have been the focus of extensive studies from cellular, biochemical and structural prospectives. However, there are several aspects of their regulation and targeting that are not well understood. From a broader point of view, there is yet no explanation for the apparent redundancy of deadenylase enzymes both within a complex, in the case of Ccr4-Not, and between complexes (Winkler & Balacco 2013; Wahle & Winkler 2013).

Both PAN2-PAN3 and Ccr4-Not complexes are recruited by generic mechanisms as well as more target-specific interactions. The most important generic recruitment is the interaction with PABPC1 (Stupfler et al. 2016; Siddiqui et al. 2007), a protein that co-operatively binds poly(A) tails to promote mRNA re-circularisation (Wells et al. 1998; Amrani et al. 2008) and translation present on virtually all mRNAs.

Similarly, target-specific mechanisms, such as the ones triggered by AU-rich elements (ARE)-binding proteins (Fabian et al. 2013) or the miRNA-mediated mRNA decay pathway (Chen et al. 2014) have been extensively studied, but many of the regulatory interactions are still unclear. For instance, further studies will be required to shed light on the role of CNOT9, one of the canonical subunits of the Ccr4-Not complex that has no catalytic activity.

### *1.8.1 Objectives*

In order to further understand the role of these Ccr4-Not and PAN2-PAN3 deadenylation complexes in human cells, a dual approach was designed. On one hand, an investigation on the cellular role of the PAN2-PAN3 complex

and its regulation was conceived. On the other hand, in order to complete the characterisation of this complex, biochemical approaches were designed to further understand the deadenylation activity of the human PAN2-PAN3 complex in a system reconsituted *in vitro*.

Finally, a predominantly biochemical strategy was outlined for the Ccr4-Not complex. The main objective was to understand in more detail the role of non-catalytic subunits of the human Ccr4-Not complex in deadenylation that are proposed to enhance activity of the nuclease module (Stowell et al. 2016).

### *1.8.2 Specific aims*

The specific aims that will be addressed, are:

1. To develop tools for the investigation of the cellular role of the PAN2-PAN3 complex by using the CRISPR/Cas9 genome engineering technology. This will allow the creation of several cell lines carrying modified genomic DNA, by multiplexed gene targeting and introducing single point modifications. In addition, tools will be developed to characterise the the biochemical activity of the human PAN2-PAN3 complex. To this end, bacterial expression strategies and protein purification approaches will be designed, in order to isolate sufficient amounts of recombinant proteins to use in *in vitro* assays.

2. To develop protein expression and purification strategies to reconstitute Ccr4-Not (sub)-complexes that will allow to determine the contribution of non-catalytic subunits to the activity of the nuclease module. Specifically, a strategy will be developed that allows the investigation of the role of the CNOT9 subunit.

3. To investigate the contribution of non-catalytic subunits on deadenylation of various substrates, including model short and long poly(A) tails. The regulatory effects of PABPC1 will be investigated by coating poly(A) substrates with the full length PABPC1 protein and a fragment comprising the RRM1 and RRM2 domains only. In addition, the role of non-catalytic subunits on substrate selectivity will be investigated.

4. To apply the strategies and biochemical approaches outlined above to investigate the CNOT9 P131L variant. This variant is caused by a C>T mutation and is found to be recurrent in melanoma suggesting an important role for Ccr4-Not in melanoma biology.

### *1.8.3 Experimental approaches*

In order to achieve the aim of further characterising the cellular role of the PAN2-PAN3 complex, several HEK293 cell lines will be generated using the CRISPR/Cas9 technology. This will be the starting platform to perform the initial experiments, to then move to more complex systems such as mouse embryonic stem cells (mESCs) or zebrafish.

For the biochemical characterisation, recombinant protein complexes will be expressed in bacteria and isolated by various purification strategies (Maryati et al. 2015). The human PAN2-PAN3 recombinant complex and the human Ccr4-Not core complex will be compared and their deadenylation activity will be profiled by several *in vitro* deandylation assays (Maryati et al. 2014).

# Chapter 2

## Materials and methods

# Chapter 2. Materials and methods

## 2.1 Bacterial growth and transformation

### *2.1.1 Reagents, stock solutions and buffers*

Lysogeny Broth (LB) media and agar: 10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl, pH 7.2, for LB agar include 15 g/l bacteriological agar. The medium was sterilised by heat using a Prestige medical 2100 classic bench-top autoclave and stored at room temperature.

Ampicillin 1000x stock solution: 100 mg/ml in 50% ethanol/dd$H_2O$. Sterilised by filtration (0.22 μm pore size). Stored at -20°C.

Chloramphenicol 1000x stock solution: 34 mg/ml in 100% ethanol. Sterilised by filtration (0.22 μm pore size). Stored at -20°C.

IPTG 1 M: 2.38 g of Isopropyl β-D-1-thiogalactopyranoside in dd$H_2O$ up to 10 ml, sterilised by filtration (0.22 μm pore size). Stored in aliquots at -20°C.

Glycerol 50% solution: 50 ml of glycerol mixed with 50 ml dd$H_2O$, autoclaved with a Prestige medical 2100 classic bench-top autoclave and stored at 4°C.

### *2.1.2 Culture of* Escherichia coli

The *E. coli* DH5α strain was used for all the manipulations of DNA. *E. coli* BL21 DE3 and ArcticExpress DE3 strains were used for protein expression. To obtain single colonies, bacteria from glycerol stocks were streaked onto LB agar plates (with appropriate antibiotics) and grown at 37°C for 14-16h. Once colonies had formed the plates were stored at 4°C for up to 2-3 weeks. Liquid cultures (2-5 ml) were inoculated with single colonies and grown in into universal screw-cap tubes containing LB medium with appropriate antibiotics and incubated at 37°C, 200 rpm for 12-16 h. For larger cultures glass conical flasks were used.

### *2.1.3 Preparation of* Escherichia coli *competent cells*

*E. coli* DH5α strain were made competent for transformation following a standard calcium chloride protocol. A single *E. coli* colony from an LB plate was used to inoculate 5 ml of LB medium and grown at 37°C, 200 rpm for 14-16 h. The following morning the culture was diluted 1:50 in LB medium and grown at 37°C to an optical density ($OD_{600}$) of 0.3-0.4. The culture was incubated on ice for 10 min in pre-chilled 50 ml centrifuge tubes and then harvested by centrifugation at 3500 rpm, 4°C for 5 min. Cell pellets were re-suspended gently with 25 ml of ice-cold $CaCl_2$ 0.1 M and centrifuged again at 3500 rpm, 4°C for 5 min. Each cell pellet was then gently re-suspended in 4 ml of cold $CaCl_2$ 0.1 M. A solution of 50% glycerol was added to a final concentration of 25%. The cells were stored at 4°C, ready for transformation, or 100 µl aliquots were snap frozen in liquid nitrogen and stored at -80°C.

### *2.1.4 Transformation of competent cells*

Between 25 and 50 µl of competent cells were used for each transformation. When transforming plasmids, 1-10 ng of DNA were used; when transforming ligations or mutagenesis reactions, 5-10 µl of ligation reaction mix was used to transform 50-100 µl of competent cells. Competent *E. coli* cells and DNA were mixed in pre-chilled 1.5 ml microcentrifuge tubes and incubated on ice for 15-30 min. The transformation mix was heat shocked by incubation at 42°C for 50-60 s, followed by incubation on ice for 3-5 min. One ml of LB medium was added to the cells and they were incubated on a shaker at 37°C for 1 h. For plasmid transformation 50 µl of the cell suspension was spread onto LB agar plates and grown for 14-16 h at 37°C. For ligations or mutagenesis transformations the cells suspension was centrifuged at 8000 g for 30 s, supernatant was decanted and the cells re-suspended in 50 µl of LB. All the cells were then spread onto LB agar plates with the appropriate antibiotic and grown for 14-16 h at 37°C.

## 2.2 Molecular biology

2.2.1 Reagents, stock solutions and buffers

5x TBE: 40 mM Tris base, 40 mM boric acid, 1 mM EDTA, pH 7.9, stored at room temperature.

6x Ficoll loading dye: 16% Ficoll, 0.2% orange G in dd$H_2O$, stored at room temperature.

TE Buffer: 10 mM Tris-HCl pH 7.9, 1 mM EDTA.

Oligonucleotides: dry primers were solubilised in TE to a 100 µM final concentration and stored at -20$^0$C.

Plasmid elution buffer: 5 mM Tris-HCl pH 7.9

*2.2.2 Small scale DNA plasmid preparation*

Typically, Sigma-Aldrich GenElute Plasmid Miniprep Kit (Cat# PLN350) was used for the small scale preparation of plasmid DNA. After the transformation, a single *E. coli* colony was picked to inoculate 5 ml of LB (with the appropriate antibiotic) and grown for 14-16 h in an orbital shaker at 37°C, 200 rpm. The overnight culture was pelleted by centrifugation at 4000 rpm for 5 min at 4°C. The plasmid DNA was then isolated according to the manufacturer's protocol. Small scale plasmid isolations were typically eluted in 35 µl of elution buffer and then stored at -20°C.

*2.2.3 Large scale plasmid DNA preparation*

To isolate large quantities of high quality plasmid DNA for mammalian transfections, the Sigma GenElute HP Plasmid Maxiprep Kit was used (Cat# NA0310). After the transformation, a single *E. coli* colony was picked to inoculate 5 ml of LB with the appropriate antibiotics. This culture was grown for 4-6 h on a shaker at 37°C and 200 rpm and then used to inoculate 500 ml of LB with the appropriate antibiotic, grown for 14-16 h on a shaker

at 37°C and 200 rpm. The culture was then pelleted by centrifugation at 4000 rpm for 20 min at 4°C. The plasmid DNA was isolated following the manufacturer's protocol. Large scale plasmid isolations were typically eluted in 500 µl of elution buffer and stored at -20°C.

### *2.2.4 Determination of DNA/RNA concentration*

All DNA quantifications were performed by spectrophotometry using a NanoDrop ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies). The A260/A280 ratio values were checked to estimate sample purity; values between 1.8 and 2.0 were considered acceptable for DNA samples.

### *2.2.5 Agarose gel electrophoresis*

DNA samples were analysed with gel electrophoresis by using 0.8%-2% agarose gels (w/v) according to the size of the DNA fragments being resolved. Ethidium Bromide (0.5 µg/ml) was added to 1x TBE buffer containing agarose. DNA samples were loaded in a 1x Ficoll loading dye. DNA gels were run with 1x TBE buffer at 80 V for 1-2 h depending on the sample. The DNA gels were visualised by transillumination with UV light and imaged using a Gel-Doc 2000 (Bio-Rad) and Quantity One software (Bio-Rad).

### *2.2.6 DNA extraction and purification from agarose gel*

To purify DNA fragments from agarose gels, DNA bands were excised from the gel using a surgical scalpel. Purification was performed with a Monarch® DNA Gel Extraction Kit (Cat# T1020L), following the manufacturer's protocol. The DNA was typically eluted in 12 µl of elution buffer; 6 µl of elution buffer were used for DNA fragments smaller than 500 bp to increase the concentration.

### *2.2.7 Annealing of DNA oligo*

To anneal complementary DNA oligonucleotides 1 µl of each oligo (100 µM stock in TE buffer) was incubated with 3 µl of 10x T4 DNA ligase buffer and 15 µl of dd$H_2O$. Reactions were incubated in a thermocycler at 95°C and allowed to cool at 1°C/min until the temperature reached 25°C. The DNA duplexes were used in ligation reactions.

### *2.2.8 Restriction enzyme digestion of DNA*

Plasmid DNA or PCR products were digested with restriction endonucleases to generate compatible protruding ends for cloning or to verify the insertion of a DNA fragment into a vector. All restriction endonucleases and corresponding buffers were supplied by NEB and used according to the manufacturer's instructions. To verify the insertion of DNA fragments after ligation reactions, 100 ng of plasmid DNA were digested in 10 µl total reaction volume. Larger quantities of DNA were digested for cloning and ligations; typically, 5-10 µg of DNA were digested in a 50 µl reaction volume. Digestions were performed in a thermocycler with heated lid, to avoid buffer evaporation.

### *2.2.9 Removal of 5' phosphate group from digested plasmid DNA*

The removal of 5' phosphate groups from restriction enzyme digested plasmid was conducted to prevent DNA from self-ligating during ligation reactions. The 5' phosphate group was removed by adding Antarctic Phosphatase (NEB) to the digested plasmid DNA. Reactions were performed following the manufacturer's protocol.

### *2.2.10 Ligation of DNA fragments*

Recombinant plasmids were created by ligating DNA fragments using T4 DNA ligase (NEB). Ligation reactions were set up by mixing 50 ng of digested vector with a 3-fold molar excess of insert (digested with

compatible enzymes) and 1 µl of T4 DNA ligase, in a 1x T4 DNA ligase buffer final reaction of 10-15 µl. Reactions were incubated for 45-60 min at room temperature. Competent *E. coli* cells were immediately transformed with ligation reactions.

### *2.2.11 Polymerase Chain Reaction (PCR)*

PCR reactions were performed using Peqlab Primus 96 advanced thermocycler. PCRs were used to generate DNA fragments for cloning or to screen recombinant plasmids. For screening newly generated plasmids (colony PCR), single colonies obtained from a ligation reaction were used to inoculate 500 µl of LB with the appropriate antibiotic and grown for 1 h in an orbital shaker at 37°C and 200 rpm. Taq Polymerase (NEB) PCR reactions were set up following the manufacturer's standard protocol in a 25 µl final volume. The primers used are shown in Table 2.1. The standard protocol for a colony PCR is the following: initial denaturation step at 95°C for 3 min followed by 25 cycles each consisting of a denaturing step of 95°C for 20 s, annealing at 50-55°C for 20 s, then elongation at 68°C for 1 min/kb. A final elongation step at 68°C for 5 min was included. To check the size of the PCR product, 1 µl of the reaction was analysed by agarose gel electrophoresis.

In order to synthesise DNA fragments for cloning, Phusion DNA polymerase (NEB) was used. PCR reactions were set up following the manufacturer's standard protocol in a 25 µl final volume. The primers used are shown in Table 2.1. The standard protocol for a PCR with Phusion polymerase is the following: initial denaturation step at 98°C for 1 min followed by 30 cycles consisting of a denaturing step of 98°C for 10 s, annealing at 50-55°C for 20 s, then elongation step at 72°C for 30 s/kb. A final elongation step for 5 min at 72°C was included. To check the size of the PCR product, 1 µl of the reaction was analysed by agarose gel electrophoresis. For specific amplification requirements, a touch-down PCR protocol was established by

using Phusion polymerase (NEB). It includes a first set of 15 cycles with a denaturation step at 98°C for 1 min, an annealing step ramping down by 1°C each cycle from 75 to 60°C for 20 s, and the elongation step at 72°C for 30 s/kb. The second set of 30 cycles was as described above.

*2.2.12 Site-directed mutagenesis reaction*

For site-directed mutagenesis reactions, overlapping primers with the desired mutations were designed (Table 2.1) using the PrimerX programme (http://www.bioinformatics.org/primerx) using the primer design protocol optimised for the Stratagene Quikchange protocol. Reactions were set up following the manufacturer's protocol in a 25 µl final volume, using the Phusion polymerase (NEB). The annealing temperatures varied between 50 and 60°C to optimise the reaction efficiency. To remove template DNA, 1 µl of DpnI enzyme (NEB) was added to each mutagenesis reaction and incubated at 37°C for 1 h. Five microliters were used to transform *E. coli*. Isolated plasmids were sequenced to confirm the desired mutations presence.

| **Name** | **Sequence** |
|---|---|
| PQE30L R | GTTCTGAGGTCATTACTGG |
| pACYCDuetUP1 | GGATCTCGACGCTCTCCT |
| DuetDown1 | GATTATGCGGCCGTGTACAA |
| DuetUP2 | TTGTACACGGCCGCATAATC |
| T7 terminator | GCTAGTTATTGCTCAGCGG |
| CNOT7_D40A F | CTACGTTGCGATGGCGACCGAATTTCCGGG |
| CNOT7_D40A R | CCCGGAAATTCGGTCGCCATCGCAACGTAG |
| CNOT6L_E240A F | GCGGATATTATCAGTCTGCAGGCGGTGGAAAC CGAACAGTATTTTAC |

| | |
|---|---|
| CNOT6L_E240A R | GTAAAATACTGTTCGGTTTCCACCGCCTGCAGA CTGATAATATCCGC |
| EMX1 guide 1 | CACCGGTCACCTCCAATGACTAGGG |
| EMX1 guide 2 | AAACCCCTAGTCATTGGAGGTGACC |
| EMX1 F | AGTTTCTCATCTGTGCCCCT |
| EMX1 R | ACTACAGTGGTGCCTGGAAA |
| Human U6 Seq F | ACTATCATATGCTTACCGTAAC |
| PAN2_F | CTTGCCCCTCCACTCTAAGT |
| PAN2_R | AACAACCTCAGACCCCACAT |
| PAN2_F2 | GACCATGTCCCTGACCTTCTAT |
| PAN2_R2 | CACAGACCAAGGGAACAGAAAG |
| D1087A1 top | CACCGTGACCCGGAAGTCCTTCTGC |
| D1087A1 bottom | AAACGCAGAAGGACTTCCGGGTCAC |
| D1087A2 top | CACCGAAGAGTCTTCTTGTGAGAGT |
| D1087A2 bottom | AAACACTCTCACAAGAAGACTCTTC |
| D978A1 top | CACCGCCCACCAGGTCCCCAATCTG |
| D978A1 bottom | AAACCAGATTGGGGACCTGGTGGGC |
| D978A2 top | CACCGAATGAGGTAACCAAGACCAA |
| D978A2 bottom | AAACTTGGTCTTGGTTACCTCATTC |
| Pan2_BamHI_F | AAAAAAGGATCCATGAATTTTG |
| Pan2_SalI_R | AAAAAAGTCGACAAATGAATTTTG |
| Pan3_BamHI_F | AAAAAAGGATCCATGAATTCG |
| Pan3_SalI_R | AAAAAAGTCGACAAATGAATTCG |

| | |
|---|---|
| pBD_MCS_top | AATTCGGGGATCCAAGGCCTAGGTACCG |
| pBD_MCS_bottom | TCGACGGTACCTAGGCCTTGGATCCCCG |
| D1087A Set3 Top | CACCGAGAAAACGAAGCTTTAAGT |
| D1087A Set3 Bottom | AAACACTTAAAGCTTCGTTTTCTC |
| D1087A Set4 Top | CACCGATGGCCTGCAGAAGGACTTC |
| D1087A Set4 Bottom | AAACGAAGTCCTTCTGCAGGCCATC |
| Pan2_gDNA/DEDD_F | CGCGCCTGGCTATGAGCTCCTTATTAAGTACC |
| Pan2_gDNA/DEDD_R | TCACCCACCACCCCAACTTTTAGAGTCCC |
| Pan3_L757E-A760E_F | GAAGATGCCGAACAAATGCGTAATGATG |
| Pan3_F753E_R | CTGGGTATATTCGCGGGCACCGATCATCG |
| Pan3_D856T_F | ACAGAAAAATCGGTGCTGGTCGTG |
| Pan3_R855T_R | TGTTGAGATCAGGGAAATTTTTTCCGGG |
| Pan3_200_BamHI_F | AAAAAAGGATCCAGCAAAACCGTATTCGGCTC |
| Pan3_445_BamHI_F | AAAAAAGGATCCAATGGCCGATGAACTGCG |
| pQE80L_NcoI_univ_F | AAAAAACCATGGGAGGATCGCATCACCATC |
| Cnot9_BamHI_F | AAAAAAGGATCCACACAGCCTGGCGACGGCTGC |
| ARM_BamHI_F | AAAAAAGGATCCAAGAGAAAAGATCTATCAGTGG |
| Cnot1M_BamHI_F | AAAAAAGGATCCAACCGAGCGTATCGTGGAGCCGC |
| Pan2_57_BamHI_F | AAAAAAGGATCCAATTATGGAGGGTGTGTATTCG |
| Pan2_1177_SalI_R | AAAAAAGTCGACGCGACCTTTTTCATACAGG |

| | |
|---|---|
| Gibson Cnot7 F | CCACTGCTGGGCCTGGACAGCACCCTCGAGATG CCAGCGGCAACTGTAGA |
| Gibson Cnot7 R | TCTAGAGGTACCACGCGTGAATTCTCGACTCAT GACTGCTTGTTGGCTTC |
| PABPC1_BamHI_F | AAAAAAGGATCCCAGGAACCTTTGACTGCTTCC |
| PABPC1_SalI_R | AAAAAAGTCGACTTAAACAGTTGGAACACCGG |
| Pan2_Top1_Benchling | CACCGCTGTGGGTGGGGAGCCACG |
| Pan2_Bott1_Benchling | AAACCGTGGCTCCCCACCCACAGC |
| Pan2_Top2_Benchling | CACCGTGCAAGTCAAAGTGGGAGA |
| Pan2_Bott2_Benchling | AAACTCTCCCACTTTGACTTGCAC |
| Pan2_Exon2_F | TTGGGTCAGGCATCTAAAGG |
| Pan2_Exon2_R | ATGCAGGAGGACCATTTGAG |
| BTG2_ATG_F | ATGAGCCACGGGAAGGGAAC |
| pQE80L_NoHis_R | TCCTCTCATAGTTAATTTCTCC |
| Pan2_350_BamHI_F | AAAAAAGGATCCAGTGAAGGTTGCGTTCATC |
| Pan2_480_BamHI_F | AAAAAAGGATCCGAAGAACCGCATCTGCAC |
| Pan2_940_BamHI_F | AAAAAAGGATCCGCGAGCGTCCTGCTGGCTG |
| Pan2_1177_SalI_TAA_R | AAAAAGTCGACTTACCAATCCATTTTGCGACC |
| Pan2_501_BamHI_F | AAAAGGATCCCTGGGTCTGGAAGATTTCG |
| Pan2_1152_SalI_R | AAAAGTCGACTTATTCCAGGTATTTGCGATAC |
| Pan2_1171_SalI_R | AAAAGTCGACTTAATACAGGCCTTTCAGCAC |
| Pan2_501_BamHI_GB_F | GGATCGCATCACCATCACCATCACGGATCCCTG GGTCTGGAAGATTTCG |

| | |
|---|---|
| Pan2_1152_SalI_GB_R | AGCTCAGCTAATTAAGCTTGGCTGCAGGTCGAC TTATTCCAGGTATTTGC |
| Pan2_1171_SalI_GB_R | CAGCTAATTAAGCTTGGCTGCAGGTCGACTTA ATACAGGCCTTTCAG |
| Pan2_Loop1_M834_F | ATGCAGTGGGGTCCGGCGCG |
| Pan2_Loop1_N778_R | GTTTTTGCTGATTTCACCGCC |
| Pan2_Loop2_S1006_F | TCCGTCGCACGTATCACGTGC |
| Pan2_Loop2_E988_R | TTCTTCTTCGTTCAGGGTGAC |
| pQE80L_TEV_BamHI_PABP_GB_F | GGATCAGAAAACCTGTATTTTCAGGGCGGATC CATGAACCCCAGTGCCCCCAGC |
| pQE80L_SalI_PABP_GB_R | AGCTCAGCTAATTAAGCTTGGCTGCAGGTCGAC TTAAACAGTTGGAACACCGG |

**Table 2.1** – List of primers.

## 2.3 Yeast two-hybrid analysis

### *2.3.1 Reagents, stock solutions and buffers for use in yeast culture*

YPD medium: 20 g/l bacteriological peptone, 10 g/l yeast extract, 20 g/l glucose, $ddH_2O$. For YPD agar, 20 g/l bacteriological agar was added. The medium was sterilised by heat using a Prestige medical 2100 classic bench-top autoclave and stored at room temperature.

Yeast selective complete (YSC) medium: 6.7 g/l yeast nutrient broth, 20 g/l glucose, 1.4 g/l amino acid drop out mix (without tryptophan and leucine). For selective agar plates, 20 g/l bacteriological agar was added. The medium was sterilised by heat using a Prestige medical 2100 classic bench-top autoclave and stored at room temperature.

Single-stranded DNA carrier: Herring Sperm single-stranded DNA (2 mg/ml, Abcam, ab46666) was heated to 98°C for 5 min and quickly cooled on ice before use. Stored in aliquots at -20°C.

Lithium acetate solutions: LiAc solutions (1 M and 0.1 M; Sigma) were sterilised by filtration (0.22 μm pore size) and stored at room temperature.

PEG-3350 solution: a solution containing 50% (w/v) polyethylene glycol (PEG-3350; Sigma) was sterilised by filtration (0.22 μm pore size). The solution was stored at room temperature.

SDS loading buffer (4x): 40% glycerol, 1% β-mercaptoethanol, 1.25% Bromophenolblue in $ddH_2O$. Aliquots were stored at -20°C.

### *2.3.2 Culture of* Saccharomyces cerevisiae *strain YRG2*

The yeast strain *S. cerevisiae* YRG2 (Stratagene; stored at $-80^0$C in YPD containing 50% glycerol) was used for yeast two–hybrid experiments. The yeast was streaked onto YPD agar plates and grown at 30°C to obtain single colonies. Once colonies had formed (2-4 days) the plates were stored at 4°C

for 1-2 weeks and then re-streaked. Large single colonies were used to inoculate liquid cultures of 10 ml YPD or YSC medium and incubated at 30°C, 200 rpm for 14-16 h. For larger cultures glass flasks were used.

### *2.3.3 Transformation of YRG2 cells*

*S. cerevisiae* YRG2 cells were made competent with a heat shock treatment. A new batch of competent cells was prepared every time a yeast two-hybrid experiment was performed. One large single yeast colony was used to inoculate 10 ml of YPD medium and grown overnight at 30°C and 200 rpm. This culture was then used to inoculate 50 ml of YPD medium to give an $OD_{600}$ of 0.2 and then incubated at 30°C and 200 rpm until the $OD_{600}$ had reached 0.8 (5-6 h). The culture was centrifuged at 3000 g for 5 min. The supernatant was removed and the cell pellets were gently re-suspended in 25 ml of sterile water followed by another centrifuge step. The supernatant was removed and the yeast pellets re-suspended in 1 ml of 0.1 M lithium acetate, transferred to a 1.5 ml microfuge tube and centrifuged at 4000 g for 15 s. The supernatant was removed and the pellets re-suspended to a final volume of 500 µl using 0.1 M lithium acetate. One 50 ml culture provided sufficient cells for 10-12 transformations. The cell suspension aliquots were distributed in 1.5 ml microfuge tubes and the supernatant removed after a centrifuge step. At first, the 50% PEG solution was gently pipetted onto the yeast cell pellet, followed by a mix of the other transformation components (Table 2.2). After the addition of all components, the transformation mix was vortexed until the pellet was completely re-suspended. The yeast transformation mix was then incubated for 30 min at 30°C and 200 rpm followed by 20 min incubation at 42°C in a water bath. The samples were then centrifuged at 4000 g for 15 s and the yeast pellets re-suspended in 500 µl of $ddH_2O$. From this cell suspension, 100 µl were plated on YSC selective plates and incubated at 30°C for 3-5 days until colonies formed.

| Reagent | Volume (µl) |
|---|---|
| 50% polyethylene glycol (PEG-3350) | 240 |
| 500 ng plasmid DNA diluted in $ddH_2O$ | 50 |
| 1 M lithium acetate | 36 |
| ssDNA (2 mg/ml) | 25 |

**Table 2.2** - Mixture for yeast transformation.

### *2.3.4 Yeast two-hybrid β-galactosidase assay*

Derivatives of plasmids pBD-GAL4 Cam and pAD-GAL4-2.1 (Stratagene) were used for yeast two-hybrid assays. For each transformation, three independent yeast clones were tested. Yeast colonies from transformations were streaked into small areas (approximately 1 $cm^2$) on YSC agar plates. After 24 h the yeast streaks were used to inoculate 2 ml cultures of YSC medium. After incubation at 30°C and 200 rpm for 14-16 h, culture samples were diluted 10-fold and used to determine the $OD_{600}$. From the remaining culture, 25 µl were mixed with 25 µl of Beta-Glo reagent (Promega) in a white 96 well flat bottom plate and incubated in the dark for 30 min at room temperature. Luminescence was measured for 1 s using Promega GloMax microplate luminometer and the manufacturer's software. The luminescence values were normalised using the $OD_{600}$ readings of the cultures.

## 2.4 Cell culture

### *2.4.1 Reagents, stock solutions and buffers for tissue culture*

Phosphate buffered saline: 137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4$, 1.47 mM $KH_2PO_4$ , pH 7.4 (1x PBS without $Ca^{2+}$, $Mg^{2+}$; PAA, H15-002)

HEPES-buffered saline 2x: 50 mM HEPES, 280 mM NaCl, 1 mM $NA_2HPO_4$, pH 7.05. The solution was filter-sterilised (0.22 µm filter) and stored at 4°C.

DMEM complete: DMEM (Sigma, D6546), 10% fetal bovine serum (FBS; Biosera S1900-500), 100 units/ml penicillin and 100 units/ml streptomycin (PAA, P11-010), 2 mM glutamine (200 mM; Lonza, BE17-60SE)

Trypsin/EDTA solutions: 10x stock of PBS-containing 2.5% trypsin (Sigma-Aldrich CAT# 59427C) stored at -20°C. 1x solutions were stored at 4°C

1 M $CaCl_2$ solution: 14.7 g of $CaCl_2$ were dissolved in dd$H_2$0 to a 100 ml final volume. The solution was filter-sterilised (0.22 µm filter)

### *2.4.2 Retrieving cells from liquid nitrogen storage*

To start cultures from liquid nitrogen storage, HEK293 cells were thawed quickly by incubation in a 37°C water bath and then added to 9 ml of complete DMEM medium (pre-warmed at 37°C). To remove the DMSO, cells were centrifuged at 1000 rpm for 3 min and the supernatant was removed. The cell pellet was re-suspended in 10 ml of complete DMEM medium and cells were transferred to a 75 $cm^2$ flask and placed in a cell-culture incubator (5% $CO_2$ and 37°C). After 24 h, the medium was replaced with fresh complete DMEM. After 48 h the cells were passaged.

### *2.4.3 Maintenance of HEK293 cell line*

HEK293 cell lines were maintained with complete DMEM in a humidified incubator at 37°C and 5% $CO_2$. When cells were 70-80% confluent, they

were washed with phosphate buffered saline (PBS; pre-warmed at 37°C) and incubated with 3 ml of trypsin/EDTA for 3 min at 37°C. Trypsin was deactivated by adding 10 ml complete DMEM and cells were gently re-suspended. The cell suspension was centrifuged at 1000 rpm for 3 min to pellet the cells, the supernatant removed and the pellet re-suspended in complete DMEM for culturing or plating.

### *2.4.4 Freezing cells for long term storage*

To prepare HEK293 cell lines for liquid nitrogen storage, 60-70% confluent cells from a 75 $cm^2$ flask were washed with PBS and incubated for 3 min at 37°C with 3 ml of trypsin/EDTA. Then cells were re-suspended with 10 ml of complete DMEM. The cell suspension was centrifuged at 1000 rpm for 3 min and the pellet was re-suspended in 2 ml of complete DMEM containing 10% dimethylsulfoxide (DMSO). Two aliquots of the cell suspension were transferred into cryovials, then placed at -80°C in a Mr Frosty cryocontainer (NALGENE Labware), which allows slow freezing of -1°C /min. After 14-16 h, the cryovials were placed in liquid nitrogen for long-term storage.

### *2.4.5 Transfection with Calcium Phosphate*

HEK293 cells were transfected with the appropriate vector (Table 2.3) using the calcium phosphate precipitation method. Cells were seeded at 60% confluency into 6-wells plates (TPP, CAT# 92406) 24 h prior to transfection. After 14-16 h, the cells were washed with PBS before the addition of fresh complete DMEM medium. 24 h after seeding a 0.25 M $CaCl_2$ (pH 7.7) solution (100 µl) containing 1 µg of plasmid DNA was prepared and added to 100 µl of 2x HEPES-Buffered Saline solution while gently agitating. The mixture was left at room temperature to precipitate for 20 min before being added drop-wise to the cells. After 14-16 h cells were washed with PBS and fresh complete DMEM medium was added.

### *2.4.6 Transfection with GeneJuice*

HEK293 cells were transfected with appropriate vectors (Table 2.8) using the GeneJuice Transfection Reagent (Novagen, CAT# 70967-5). Cells were seeded at 60% confluency into 6-wells plates (TPP, CAT# 92406) 24 h prior transfection. After 12-16 h, cells were washed with PBS before addition of fresh DMEM medium. For each well, 100 µl of serum-free medium were placed into a sterile tube and 3 µl of GeneJuice reagent were added; the solution was mixed by vortexing and incubated at room temperature for 5 min. A total of 1-6 µg of DNA were added to each tube, the solution was mixed by gentle pipetting and incubated at room temperature for 5-15 min. The mixture was then added drop-wise to the cells over the entire surface of the well, and the plate was gently rocked. After 12 h cells were washed with PBS and fresh complete medium was added.

## 2.5 Genome editing with CRISPR/Cas9 technology

### *2.5.1 gRNAs design and cloning*

DNA oligonucleotides coding for guide RNAs (gRNAs) were designed using a specific tool included in the Benchling software. The gRNAs with the highest scores and lowest off-target effects were selected for subsequent steps. Oligonucleotide pairs were phosphorylated and annealed with T4 PNK (NEB) with the following steps: 30 min at 37°C, 5 min at 95°C and then ramp down to 25°C at 5°C/min. The annealed oligonucleotides reaction was diluted 1:250. One microgram of the relevant plasmid (Table 2.3) was digested with BbsI (NEB) in a 20 µl reaction for 1 h and then gel-purified (Sigma-Aldrich Cat# NA1111-1KT). Ligation reactions were set up in a 20 µl final volume with 50 ng of digested vector and 1 µl of diluted oligonucleotide dimers. *E.coli* DH5α competent cells were transformed with 1-3 µl of the ligation reaction and the presence of the insert was confirmed by sequencing.

| Plasmid | Addgene | Cas9 encoded | Selection marker |
|---|---|---|---|
| pSpCas9(BB)-2A-Puro (PX459) | 48139 | Wild type | Puromycin |
| pSpCas9n(BB)-2A-GFP (PX461) | 48140 | D10A nickase | GFP |
| pSpCas9n(BB)-2A-Puro (PX462) V2.0 | 62987 | D10A nickase | Puromycin |

**Table 2.3** - List of plasmids for genome editing.

### *2.5.2 HEK293 transfection and generation of stable cell lines*

HEK293 cells were transfected in a 6-wells plate following the standard GeneJuice protocol (Novagen) or the Calcium-Phosphate method described. 24 h after transfection the cells were diluted 1:50 and cultured in 10 cm

dishes with complete DMEM supplemented with 3 μg/mL puromycin. After 1-3 weeks, single clones were picked and grown for further experiments.

*2.5.3 gRNAs screening for efficiency*

The genomic DNA was extracted (Sigma-Aldrich GenElute Mammalian genomic DNA miniprep kit, Cat# G1N10) from populations of transfected cells. Touch-down PCRs were performed with Taq polymerase (NEB) using primers spanning 250-500 bases on either sides of the genomic target region. The band of interest was extracted from the agarose gel. 200 ng of gel-extracted DNA were added to NEB buffer 2 and dd$H_2O$ to a total of 19 uL. In a thermocycler, hybridisation reactions were set up with the following parameters: 5 min at 95°C, ramp down to 85°C at 2°C/min, ramp down to 25°C at 0.1°C/min. Finally, 1 uL of T7 endonuclease I (NEB) was added and reactions were incubated for 30-60 min. Reactions were loaded on agarose gels to screen for multiple bands. The efficiency of CRISPR/Cas9 with the designed gRNAs can be assessed by the intensity of the released fragment after the endonuclease digestion, only present when hetero-duplexes are formed as a consequence of the CRISPR/Cas9 cleavage.

*2.5.4 Genome editing testing*

Single clones transfected with CRISPR/Cas9 vectors were expanded in 6-wells plates and the genomic DNA was extracted. Touch-down PCRs were performed with Phusion polymerase (NEB) using primers spanning 250-500 bases on either sides of the genomic target region. The amplified fragments were extracted from agarose gels and insertions/deletions were confirmed by sequencing.

## 2.6 Protein expression and purification

### *2.6.1 Reagents, stock solutions and buffers*

Lysogeny Broth (LB) media and agar: 10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl, pH 7.2 (NaOH)

Ampicillin 1000× stock solution: 100 mg/ml in 50% ethanol/dd$H_2O$, stored at -20°C.

Chloramphenicol 1000× stock solution: 34 mg/ml in 100% ethanol, stored at -20°C.

IPTG (Isopropyl β-2-1-thiogalactopyranoside) 1 M stock solution: dissolve 2.38 g of IPTG in 10 ml of dd$H_2O$ and sterilise by filtration

NaCl (Sodium chloride), Imidazole, β-mercapthoethanol

### *2.6.2 Expression of His-tagged proteins*

*E. coli* BL21 DE3 cells were transformed with the relevant plasmid as described above. A single colony was used to inoculate 1 ml of LB (with appropriate antibiotics) for 6-8 hours at 37°C and 200 rpm. This starter culture was used to inoculate 50 ml of LB (with appropriate antibiotics), incubated at 37°C and 180 rpm for 14-16 h. One litre of pre-warmed LB was added and the culture was grown to an $OD_{600}$ of 0.6-0.8. Protein expression was then induced by adding IPTG to a final concentration of 0.2 mM. Cultures were grown for 14-16 h at 18°C and 150 rpm or 3 h at 30°C and 150 rpm. The cultures were harvested by centrifugation (4200 rpm) using a Sorvall SLC-6000 SUPER-LITE rotor at 4°C for 45 min. The supernatant was discarded and the bacterial pellet was resuspended in 5 ml of ice-cold buffer HA (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 20 mM Imidazole, 5% glycerol, 1 mM β-mercaptoethanol). Cell paste was stored at -80°C until further use. The bacterial suspension was thawed and lysed on ice by

sonication (Qsonica XL2000), with five cycles (30 s on and 30 s off) and 60% amplitude. The lysate was transferred to 2 ml tubes and centrifuged at 4°C and 14000 rpm for 45 min to separate the soluble lysate from the insoluble material. The supernatant was collected and directly used for the purification steps.

### *2.6.3 Expression of StrepII-tagged proteins*

*E. coli* BL21 DE3 cells were transformed with the relevant plasmid as described above. A single colony was used to inoculate 1 ml of LB (with appropriate antibiotics) for 6-8 hours at 37°C and 200 rpm. This starter culture was used to inoculate 50 ml of LB (with appropriate antibiotics), incubated at 37°C and 180 rpm for 14-16 h. One litre of pre-warmed LB was added and the culture was grown to an $OD_{600}$ of 0.6-0.8. Protein expression was induced by adding IPTG to a final concentration of 0.2 mM. Cultures were grown for 14-16 h at 18°C and 150 rpm. The cultures were harvested by centrifugation (4200 rpm) using a Sorvall SLC-6000 SUPER-LITE rotor at 4°C for 45 min. The supernatant was discarded and the bacterial pellet was resuspended in 3 ml of ice-cold buffer SA (50 mM Tris-HCl pH 7.9, 250 mM NaCl, 5% glycerol, 1 mM β-mercaptoethanol). Resuspended pellets were stored at -80°C until further use. The bacterial suspension was thawed and lysed on ice by sonication (Qsonica XL2000), with five cycles (30 s on and 30 s off) and 60% amplitude. The lysate was transferred to 2 ml tubes and centrifuged at 4°C and 14000 rpm for 45 min to separate the soluble lysate from the insoluble material. The supernatant was collected and directly used for the purification steps.

### *2.6.4 Purification of His-tagged proteins*

6His-tagged proteins were purified using HisTrap FF crude column (GE Healthcare, 1 ml volume). The column was washed with 5-10 column volumes of $ddH_2O$, then equilibrated with 5-10 column volumes of buffer

HA (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 20 mM Imidazole, 5% glycerol, 1 mM β-mercaptoethanol). The soluble lysate was then loaded on the column with a 0.5 ml/min flow rate. The column was washed with 20 column volumes of buffer A and proteins were eluted with buffer HB (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 250 mM Imidazole, 5% glycerol, 1 mM β-mercaptoethanol). Five fractions of 1 ml were collected. Elution fractions were analysed by SDS-PAGE and coomassie staining. Protein buffer was exchanged to the final buffer (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol, 1 mM β-mercaptoethanol) using PD-10 desalting columns (GE Healthcare, cat# 17085101). Protein concentrations were assessed using Bio-Rad Protein Assay Reagent. Fractions were flash-frozen with liquid nitrogen and stored at -80°C.

### *2.6.5 Purification of StrepII-tagged proteins*

StrepII-tagged proteins were purified in a single step using gravity flow columns (Bio-Rad, cat# 7311550) packed with 50% StrepTactin Sepharose solution (iba, cat# 2-1201-010). 1 ml of StrepTactin Sepharose solution was added to the column, which was then equilibrated with 5 column volumes (0.5 ml) of buffer SA (50 mM Tris-HCl pH 7.9, 250 mM NaCl, 5% glycerol, 1 mM β-mercaptoethanol). The soluble lysate was then loaded on the column by gravity for 2-3 h. The column was washed with 5 ml of buffer SA and proteins were eluted with buffer SB (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 10 mM D-desthiobiotin, 5% glycerol, 1 mM β-mercaptoethanol). Five fractions of 0.5 ml were collected. Elution fractions were analysed by SDS-PAGE and coomassie staining. Protein buffer was exchanged to the final buffer (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol, 1 mM β-mercaptoethanol) using PD-10 desalting columns (GE Healthcare, cat# 17085101). Protein concentrations were assessed using Bio-Rad Protein Assay Reagent. Fractions were flash-frozen with liquid nitrogen and stored at -80°C.

### *2.6.6 Purification by size-exclusion chromatography*

Proteins were purified by size-exclusion chromatography using a HiPrep 16/60 Sephacryl S-200 HRS200 column (GE Healthcare, cat# 17116601). The column was stored at 4°C in a 20% ethanol solution and before purification was washed with two column volumes of dd$H_2O$ and equilibrated with two column volumes of gel filtration buffer (50 mM Tris-HCl pH 7.9, 250 mM NaCl, 5% glycerol, 1 mM β-mercaptoethanol). Protein samples were then loaded on the column using a SuperLoop (1 ml volume) and 1.5 column volumes of gel filtration buffer were added with a 0.5 ml/min flow rate. Elution fractions were collected and analysed by SDS-PAGE and coomassie staining. Peak fractions were pooled and concentrated by centrifugation (Sartorius, cat# VS02V1). Protein concentrations were assessed using Bio-Rad Protein Assay Reagent. Fractions were flash-frozen with liquid nitrogen and stored at -80°C.

### *2.6.7 TEV cleavage*

6His-tags were cleaved from proteins displaying a TEV tag by incubation with the TEV protease. Test reactions were prepared with 0.5 µg of target protein with 0.01, 0.05, 0.1 and 0.2 equivalents of TEV protease, and incubated overnight at 4°C. Reactions were analysed by SDS-PAGE and coomassie staining. The lowest number of equivalent sufficient for cleaving the 6His-tag, resulting in a shift of the protein band, was used to perform the cleavage reaction.

## 2.7 Protein analysis

### *2.7.1 Reagents, stock solutions and buffers for use in protein analysis*

Upper buffer (4×): 0.5 M Tris base, 0.4% SDS, pH 6.8; stored at room temperature

Lower buffer (4×): 1.5 M Tris base, 0.4% SDS, pH 8.8; stored at room temperature

Running buffer (10×): 0.25 M Tris Base, 1.0% SDS, 1.92 M glycine, stored at room temperature

Transfer buffer (10×): 0.25 M Tris Base, 1.92 M glycine, stored at room temperature

SDS loading buffer (4×): 50% upper buffer, 40% glycerol, 1% β-merchaptoethanol, 1.25% Bromophenolblue (0.1%), in $ddH_2O$; aliquots stored at -20°C

Tris buffer saline supplemented with 0.1% Tween-20 (TBST 1x): 50 mM Tris-HCl pH 7.8, 150 m NaCl, 0.1% Tween-20. Stored at room temperature

10% APS: 10% ammonium persulphate (APS) in $ddH_2O$, stored at 4°C

### *2.7.2 Bradford assay to determine protein concentration*

Bradford assay was the preferred method to determine protein concentration. The Bio-Rad protein reagent (Bio-Rad) was diluted 1:5 in $ddH_2O$ before use. BSA dissolved in $ddH_2O$ was used for protein standard curve. Protein samples were diluted 1:100 with Bradford reagent, vortexed and incubated for 1-2 min at room temperature. Absorbance values were read at $OD_{600}$ and the standard curve was obtained by linear regression analysis using GraphPad Prism 7.0 and used to interpolate the protein concentration of the samples.

### *2.7.3 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis*

Protein samples were analysed using SDS-PAGE with the X-Cell SureLock Mini Cell system (Invitrogen). SDS-PAGE gels were prepared to the appropriate percentage of resolving gel based on the molecular mass of the protein of interest. The gels were cast according to the manufacturer's protocol. The wells were washed with 1x running buffer prior to use and protein samples were denatured by boiling for 5 minutes in 1x SDS loading buffer immediately before loading. Gels were run in 1x running buffer at 180 V for 1-1.5 h according to protein size and percentage of resolving gel.

### *2.7.4 Western Blot Analysis*

Protein separated on SDS-PAGE gels for western blotting were transferred to nitrocellulose (Whatman Protran 0.45µm) membranes using the X-Cell SureLock Mini Cell system (Invitrogen) for one hour at 25 V in 1x transfer buffer. The membranes were then blocked for 1 hour at room temperature while mixing on a roller shaker using a solution of 5% dried milk in 1x TBST. The membranes were transferred to a 50 ml tube containing 5 ml of the appropriate primary antibody dilution (Table 2.4) and incubated for 14-16 h at 4°C on a roller shaker. Membranes were then washed 3 times for 5 min with TBST, transferred to the secondary antibody solution (Table 2.4) and incubated at room temperature for one hour on a roller shaker. The membranes were then washed 3 times with TBST and placed in a plastic wallet before addition of 250 µl of Pierce ECL substrate (Thermo Fisher, cat# 32106). Membranes were incubated for one minute and signals were detected using a Fujifilm LAS-4000 digital imaging system. ImageJ software was used for image analysis.

| Antibody | Supplier | Cat # | Dilution |
|---|---|---|---|
| Anti-PAN2 (rabbit) | abcam | ab129797 | 1:500 |
| Anti-rabbit IgG (goat) | abcam | ab6721 | 1:5,000 |

**Table 2.4** - List of antibodies used for Western blotting.

## 2.8 Deadenylation assays

### *2.8.1 Reagents, stock solutions and buffers*

Fluorescent RNA substrate (5'- Flc-CCUUUCCAAAAAAAAA -3') (Eurogentec)

Fluorescent DNA probe (5'- TTTTTTTTTGGAAAGG-TAMRA -3') (Eurogentec)

5× reaction buffer: 100 mM Tris-HCl pH 7.8, 25% glycerol, 5 mM β-mercaptoethanol, 20 mM $MgCl_2$,

Purified deadenylase enzyme (1x reaction buffer supplemented with 500 mM NaCl)

RNA loading buffer: 95% formamide, 0.025% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol FF, 0.025% ethidium bromide, 0.5 mM EDTA

40% acrylamide:bisacrylamide (19:1) solution

10% ammonium persulphate (APS) in $ddH_2O$.

Urea, SDS (Sodium dodecyl sulphate)

TEMED (Tetramethylethylenediamine)

### *2.8.2 Denaturing polyacrylamide gel electrophoresis*

Deadenylation reactions were set up as described in the relative Chapters and stopped by addition of RNA loading buffer. Reaction samples were then heated for 3 min at 85°C and up to 3 µl of each sample were loaded in a 16% or 20% acrylamide:bisacrylamide (19:1) and 50 % urea gel (Table 2.5). The gel was pre-run in a X-Cell SureLock Mini Cell system (Invitrogen) at 300 V for 30 min, then wells were thoroughly washed with 1x TBE immediately before samples loading. The gel was run for 20-25 min at 300 V. A Fujifilm

LAS-4000 imager was used for imaging the gel with Epi-Blue Illuminator. ImageJ software was used to analyse the gels. Densitometry analysis were performed using the 'MultiPlot' function of the ROI manager. Data were exported to GraphPad Prism 7.0 for further analysis.

| **Reagent** | **Volume or weight** |
|---|---|
| 40% acrylamide:bisacrylamide (19:1) | 5 ml |
| 5x TBE | 2 ml |
| Urea | 5 g |
| TEMED | 10 µl |
| APS | 100 µl |

**Table 2.5** - The 20% denaturing gel composition for deadenylation assays. For a 16% gel, 4 ml of 40% acrylamide:bisacrylamide (19:1) were added.

*2.8.3 Fluorescence-based deadenylase assay*

Deadenylation reactions were set up in black Greiner 384-wells with flat bottom plates. Reactions were incubated for up to 60 min at 30°C and stopped by addition of 10 µl DNA probe mix (1 µl of SDS 10%, 6 µl of TE and 3 µl of DNA probe 10 µM). After 14-16 h incubation in the dark, the fluorescence signal was measured using a Biotek Synergy HT plate reader. The temperature set point for the measurement was 25°C, the detection method used was fluorescence and the read type was 'endpoint'. The filter sets were 485/20 for excitation and 528/20 for emission; the position of the optics was 'top' and the sensitivity was 90.

*2.8.4 Coating of poly(A)$_{20}$ and poly(A)$_{50}$ RNA oligos with PABPC1 proteins*

Electrophoretic mobility shift assays (EMSA) were set up to confirm the binding of PABPC1 (full length or RRM1-RRM2 domains) to poly(A) tails of

RNA oligonucleotides and to identify the lowest excess of PABPC1 required to bind the majority of the RNA substrate in assay conditions. Binding reactions were set up in a 10 µl final volume and 200 nM RNA substrate. After 15 min incubation at room temperature with PABP proteins excess, 2 µl native loading dye were added (0.05% bromophenol blue, 0.05% xylene cyanol, 25% glycerol and 1X TBE in dd$H_2O$) and 6 µl were loaded in a 8% RNA native gel, not containing urea. No pre-run was required and samples were run at 120 V for 50-60 min. A Fujifilm LAS-4000 was used for imaging the gel with Epi-Blue Illuminator. ImageJ software was used to analyse the gels.

## 2.9 Differential scanning fluorimetry (DSF)

In a 20 µl volume reaction, protein samples (2.5-10 µM) were mixed with Sypro Orange 5x final concentration (ThermoFisher, cat# S6650) and reaction buffers containing different NaCl final concentrations. Reactions were added in triplicate to a PCR plate and analysed with a qPCR machine (Stratagene, Mx3005P). The scanning was carried out at a temperature gradient from 25°C to 95°C (1°C/min) with the following filter settings: excitation 492 nm and emission 610 nm. Data analysis was performed using custom calculation software based on Macros in Microsoft Excel, GraphPad Prism 7.0 and DSF analysis manual 3.0 as described by Niesen et al. (ftp://ftp.sgc.ox.ac.uk/pub/biophysics) (Niesen et al. 2007).

# Chapter 3

## Cellular and biochemical tools to characterise the human PAN2-PAN3 complex

# Chapter 3. Cellular and biochemical tools to characterise the human PAN2-PAN3 complex

## 3.1 Introduction

The PAN2-PAN3 complex is a highly conserved multisubunit machinery (Wahle & Winkler 2013). It has been extensively studied, both *in vitro* and *in vivo* (Boeck et al. 1996; Uchida et al. 2004). Structural investigations have recently unravelled the architecture of the trimeric PAN2-$(PAN3)_2$ complex in model organisms such as *N. crassa* (Christie et al. 2013; Jonas et al. 2014), *D. melanogaster* (Christie et al. 2013), *C. thermophilum* (Wolf et al. 2014) and *S. cerevisiae* (Schäfer et al. 2014). However, no crystal structure of the human PAN2-PAN3 complex is available to date.

PAN2 is a $Mg^{2+}$-dependent exoribonuclease that belongs to the DEDD/RNAseD superfamily of exonucleases (Uchida et al. 2004). It has no ability to directly bind RNA and in absence of PAN3 has low catalytic activity (Wolf et al. 2014; Schäfer et al. 2014). Upon binding of its complex partner, it shows increased deadenylation activity and substrate specificity (Wolf et al. 2014).

PAN3 forms an asymmetric homodimer when assembled with PAN2 (Schäfer et al. 2014). Upon formation of the complex, specific residues of PAN3 in the coiled coil have been identified to engage conserved interactions with PAN2 (Christie et al. 2013; Jonas et al. 2014; Wolf et al. 2014; Schäfer et al. 2014).

The recently discovered asymmetric conformation of the PAN2-PAN3 complex gives a few insights on its role in mRNA deadenylation in eukaryotes. Interestingly, PAN2-PAN3 displays initial high deadenylation rates, which drop once the poly(A) tail is 20-30 nucleotides long. PAN3 contains a zinc-finger domain that contributes to specificity for poly(A) tails, while its PAM-2 motif binds PABP proteins. The pseudokinase domain

can provide low-specificity RNA binding (Wolf et al. 2014) and the C-terminus, by recruiting PAN2 WD40 domain, guides PAN2 exonuclease domain to the poly(A) tail. Consistently, PAN3 can bind PAN2 and PABP simultaneously (Uchida et al. 2004; Siddiqui et al. 2007; Huang et al. 2013).

The purpose of this study was to characterise the PAN2-PAN3 complex with cellular and biochemical approaches. On one hand, by engineering cell lines containing mutations present in PAN2 and PAN3 genes, found in tumour samples collected from patients. Understanding the effect of these mutations at the protein level can help shed light on cancer mechanisms and regulation, potentially leading to the discovery of novel drugs. On the other hand, reconstituting a human PAN2-PAN3 deadenylation system *in vitro* could help characterising the deadenylation activity of this complex.

A multipronged approach was designed to this purpose. Firstly, we envisaged to create CRISPR/Cas9 genome-engineered cell lines for selected point mutations, to then characterise them phenotypically using cellular assays. Along with patient mutations, cell lines will be engineered with defined genotypes, such as PAN2 and/or PAN3 knockout alleles, and PAN2 D1087A encoding a catalytically inactive version of PAN2. Secondly, we envisaged the use of yeast-two hybrid experiments to help characterising the impact of residue changes on the interactions within the PAN2-PAN3 complex. Finally, we designed experiments to express and purify a recombinant PAN2-PAN3 complex to assess whether the amino acid substitutions affect the deadenylation activity of the complex *in vitro*.

## 3.2 Bioinformatic analysis of PAN2 and PAN3 variants associated with cancer and conserved residues of the PAN2-PAN3 complex

Nonsense and missense mutations in both PAN2 and PAN3 genes were collected from the Catalogue Of Somatic Mutations In Cancer (cancer.sanger.ac.uk/cosmic v72). A total of 147 mutations were found in

PAN2 (Table 3.1 – 15 nonsense and 132 missense) and 98 in PAN3 (Table 3.2 – 18 nonsense and 80 missense).

In order to characterise the effect of these mutations resulting in amino acid substitutions, the residues involved were visualised in the structure of the complex. However, the structure of the human PAN2-PAN3 complex has not been resolved to date, therefore the amino acids were modelled on the structure of the *S. cerevisiae* Pan2-Pan3 complex (PDB ID: 4XR7) using USCF Chimera, after alignment of the human and yeast protein sequences and identification of the conserved residues. Sixty-one mutations in PAN2 and 37 in PAN3 could be modelled on the yeast Pan2-Pan3 structure, and they do not appear to be clustered in specific regions of either proteins (Fig. 3.1 and 3.2).

Further analysis of the conserved amino acids in the PAN2-PAN3 complex was performed using the ConSurf server (Glaser et al. 2003; Landau et al. 2005; Ashkenazy et al. 2010; Celniker et al. 2013; Ashkenazy et al. 2016), which estimates the evolutionary conservation of residues by multiple sequence alignment. The conservation of specific residues is often strongly associated with their structural or catalytical function.

The ConSurf analysis for PAN2 showed that, as expected, most of the residues that are highly conserved across the species are located in the RNase domain, particularly in the region of the catalytic site. Interestingly, the amino acid stretch that helps PAN2 to anchor to the PAN3 homodimer shows a number of conserved residues. The residues on the surface regions and at the interface with PAN3 are overall less conserved, with few exceptions (Fig. 3.3A).

| AA change | CDS mutation | Primary tissue | Histology | Somatic status | Zygosity |
|---|---|---|---|---|---|
| p.E42G | c.125A>G | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.M58I | c.174G>A | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.E85* | c.253G>T | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.G91R | c.271G>A | Haematopoietic and lymphoid | Lymphoid neoplasm | Confirmed Somatic | Heterozygous |
| p.G94E | c.281G>A | Kidney | Carcinoma | Variant of unknown origin | Unknown |
| p.A97V | c.290C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.D118N | c.352G>A | Cervix | Carcinoma | Confirmed Somatic | Unknown |
| p.E126* | c.376G>T | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.E126K | c.376G>A | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.L130F | c.388C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.K138N | c.414G>T | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.R142C | c.424C>T | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.R142C | c.424C>T | Lung | Carcinoma | Previously Reported | Unknown |
| p.L161F | c.481C>T | Skin | Carcinoma | Confirmed Somatic | Heterozygous |
| p.L161F | c.481C>T | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.D165N | c.493G>A | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.G172D | c.515G>A | Endometrium | Carcinoma | Confirmed Somatic | Unknown |
| p.T185I | c.554C>T | Kidney | Carcinoma | Variant of unknown origin | Unknown |
| p.E195D | c.585G>C | Breast | Carcinoma | Variant of unknown origin | Heterozygous |
| p.T196M | c.587C>T | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.M202K | c.605T>A | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.R207C | c.619C>T | Upper aerodigestive tract | Carcinoma | Variant of unknown origin | Unknown |
| p.V218A | c.653T>C | Large intestine | Carcinoma | Variant of unknown origin | Unknown |
| p.L220P | c.659T>C | Prostate | Carcinoma | Confirmed Somatic | Unknown |
| p.D233H | c.697G>C | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.D243N | c.727G>A | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.D243N | c.727G>A | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.S256T | c.767G>C | Upper aerodigestive tract | Carcinoma | Variant of unknown origin | Unknown |
| p.A277S | c.829G>T | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.T297A | c.889A>G | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.T297A | c.889A>G | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.R299C | c.895C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.Q305* | c.913C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.C309Y | c.926G>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.Q310H | c.930A>C | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.P320S | c.958C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.D322N | c.964G>A | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.D322N | c.964G>A | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.P331H | c.992C>A | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.F336Y | c.1007T>A | Upper aerodigestive tract | Carcinoma | Confirmed Somatic | Unknown |
| p.A346T | c.1036G>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.E351K | c.1051G>A | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.W357C | c.1071G>T | Breast | Carcinoma | Variant of unknown origin | Heterozygous |
| p.P361T | c.1081C>A | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.E362K | c.1084G>A | Skin | Carcinoma | Confirmed Somatic | Heterozygous |
| p.E371V | c.1112A>T | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.E371V | c.1112A>T | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.V380M | c.1138G>A | Breast | Carcinoma | Confirmed Somatic | Unknown |
| p.D387N | c.1159G>A | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.D410E | c.1230T>G | Liver | Carcinoma | Variant of unknown origin | Unknown |
| p.D410Y | c.1228G>T | Cervix | Carcinoma | Confirmed Somatic | Unknown |
| p.A414G | c.1241C>G | Ovary | Other | Confirmed Somatic | Unknown |
| p.R422Q | c.1265G>A | Autonomic ganglia | Neuroblastoma | Variant of unknown origin | Unknown |
| p.P424A | c.1270C>G | Kidney | Carcinoma | Confirmed Somatic | Unknown |
| p.L431V | c.1291C>G | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.R432C | c.1294C>T | Stomach | Adenocarcinoma | Variant of unknown origin | Unknown |
| p.T433P | c.1297A>C | Breast | Carcinoma | Confirmed Somatic | Unknown |
| p.R451C | c.1351C>T | Large intestine | Carcinoma | Confirmed Somatic | Heterozygous |
| p.P455S | c.1363C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.D462N | c.1384G>A | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.D462G | c.1385A>G | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.V471A | c.1412T>C | Prostate | Carcinoma | Confirmed Somatic | Unknown |
| p.R478* | c.1432C>T | Large intestine | Carcinoma | Previously Reported | Heterozygous |
| p.R478* | c.1432C>T | Large intestine | Carcinoma | Confirmed Somatic | Heterozygous |
| p.R478* | c.1432C>T | Large intestine | Carcinoma | Previously Reported | Heterozygous |
| p.R478* | c.1432C>T | Large intestine | Carcinoma | Previously Reported | Heterozygous |
| p.R478Q | c.1433G>A | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.R478L | c.1433G>T | Stomach | Carcinoma | Confirmed Somatic | Heterozygous |
| p.L484I | c.1450C>A | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.R492C | c.1474C>T | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.Y498C | c.1493A>G | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.N512D | c.1534A>G | Large intestine | Carcinoma | Confirmed Somatic | Heterozygous |
| p.N512D | c.1534A>G | Large intestine | Carcinoma | Previously Reported | Heterozygous |
| p.A517V | c.1550C>T | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.E520Q | c.1558G>C | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.N529S | c.1586A>G | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.N529S | c.1586A>G | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.P540L | c.1619C>T | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.G561D | c.1682G>A | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.H565Y | c.1693C>T | Cervix | Carcinoma | Confirmed Somatic | Unknown |
| p.Q576H | c.1728G>T | Large intestine | Carcinoma | Variant of unknown origin | Unknown |
| p.Q576R | c.1727A>G | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.L581I | c.1741C>A | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.R582W | c.1744C>T | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.Q614H | c.1842G>C | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.W616* | c.1848G>A | Oesophagus | Carcinoma | Confirmed Somatic | Unknown |
| p.R618C | c.1852C>T | Skin | Malignant melanoma | Confirmed Somatic | Heterozygous |
| p.R638* | c.1912C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.A640V | c.1919C>T | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.G641* | c.1921G>T | Breast | Carcinoma | Variant of unknown origin | Heterozygous |
| p.G654V | c.1961G>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.V674M | c.2020G>A | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.Y695C | c.2084A>G | Lung | Carcinoma | Confirmed Somatic | Unknown |
| p.R703Q | c.2108G>A | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.R703Q | c.2108G>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.C718S | c.2152T>A | Central nervous system | Glioma | Confirmed Somatic | Unknown |
| p.P723L | c.2168C>T | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.I730T | c.2189T>C | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.V738L | c.2212G>C | Breast | Carcinoma | Confirmed Somatic | Unknown |
| p.R753T | c.2258G>C | NS | NS | Confirmed Somatic | Unknown |

| | | | | | |
|---|---|---|---|---|---|
| p.H767Q | c.2301C>A | Ovary | Carcinoma | Confirmed Somatic | Heterozygous |
| p.E770* | c.2308G>T | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.S772P | c.2314T>C | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.E789K | c.2365G>A | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.V803I | c.2407G>A | Prostate | Carcinoma | Confirmed Somatic | Unknown |
| p.M811I | c.2433G>A | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.D828H | c.2482G>C | Ovary | Carcinoma | Variant of unknown origin | Heterozygous |
| p.E839D | c.2517G>T | Upper aerodigestive tract | Carcinoma | Confirmed Somatic | Unknown |
| p.E840D | c.2520G>T | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.G842S | c.2524G>A | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.H854L | c.2561A>T | Pancreas | Carcinoma | Confirmed Somatic | Unknown |
| p.R859H | c.2576G>A | Oesophagus | Carcinoma | Confirmed Somatic | Unknown |
| p.R859C | c.2575C>T | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.R859C | c.2575C>T | Oesophagus | Carcinoma | Previously Reported | Unknown |
| p.L864M | c.2590C>A | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.K869N | c.2607A>C | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.Y874F | c.2621A>T | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.R877H | c.2630G>A | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.D898N | c.2692G>A | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.A902D | c.2705C>A | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.L915F | c.2743C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.K948E | c.2842A>G | Stomach | Carcinoma | Confirmed Somatic | Heterozygous |
| p.T952A | c.2854A>G | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.P958L | c.2873C>T | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.M960L | c.2878A>T | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.E982K | c.2944G>A | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.R988C | c.2962C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.T992I | c.2975C>T | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.K997I | c.2990A>T | Ovary | Carcinoma | Variant of unknown origin | Heterozygous |
| p.R1107* | c.3319C>T | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.R1107* | c.3319C>T | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.R1010W | c.3028C>T | Kidney | Carcinoma | Confirmed Somatic | Unknown |
| p.E1016D | c.3048G>T | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.P1019S | c.3055C>T | Skin | Carcinoma | Confirmed Somatic | Heterozygous |
| p.V1031L | c.3091G>C | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.G1040V | c.3119G>T | Kidney | Carcinoma | Confirmed Somatic | Unknown |
| p.D1045G | c.3134A>G | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.R1066C | c.3196C>T | Skin | Carcinoma | Confirmed Somatic | Unknown |
| p.R1066H | c.3197G>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.G1079D | c.3236G>A | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.R1085W | c.3253C>T | Small intestine | Carcinoid-endocrine tumour | Confirmed Somatic | Unknown |
| p.L1097F | c.3289C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.T1099I | c.3296C>T | Endometrium | Carcinoma | Confirmed Somatic | Unknown |
| p.R1137H | c.3410G>A | Lung | Carcinoma | Previously Reported | Unknown |
| p.R1137H | c.3410G>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.R1137C | c.3409C>T | Pancreas | Carcinoma | Variant of unknown origin | Heterozygous |
| p.R1137C | c.3409C>T | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.Q1141* | c.3421C>T | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.R1144* | c.3430C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.R1144Q | c.3431G>A | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.S1158F | c.3473C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.G1165V | c.3494G>T | Adrenal gland | Adrenal cortical carcinoma | Confirmed Somatic | Unknown |
| p.W1175R | c.3523T>A | Lung | Carcinoma | Confirmed Somatic | Unknown |
| p.K1176T | c.3527A>C | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.V1191I | c.3571G>A | Prostate | Carcinoma | Confirmed Somatic | Unknown |
| p.F1192L | c.3576C>A | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.A1197V | c.3590C>T | Large intestine | Carcinoma | Confirmed Somatic | Heterozygous |

**Table 3.1** - PAN2 mutations collected from the COSMIC database.

| AA Mutation | CDS Mutation | Primary Tissue | Histology | Somatic Status | Zygosity |
|---|---|---|---|---|---|
| p.T32K | c.95C>A | Pancreas | Carcinoma | Confirmed Somatic | Unknown |
| p.Y34F | c.101A>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.M37I | c.111G>T | Prostate | Carcinoma | Variant of unknown origin | Heterozygous |
| p.S46F | c.137C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.L50V | c.148C>G | Kidney | Carcinoma | Confirmed Somatic | Unknown |
| p.S53C | c.157A>T | Kidney | Carcinoma | Variant of unknown origin | Unknown |
| p.F71V | c.211T>G | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.S91R | c.271A>C | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.V106M | c.316G>A | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.S109* | c.326C>A | Lung | Carcinoma | Confirmed Somatic | Unknown |
| p.S109* | c.326C>G | Cervix | Carcinoma | Confirmed Somatic | Unknown |
| p.S112F | c.335C>T | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.H120Y | c.358C>T | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.M123T | c.368T>C | Central nervous system | Glioma | Confirmed Somatic | Unknown |
| p.M123I | c.369G>T | Skin | Carcinoma | Confirmed Somatic | Heterozygous |
| p.T128A | c.382A>G | Central nervous system | Glioma | Confirmed Somatic | Unknown |
| p.G130* | c.388G>T | Breast | Carcinoma | Confirmed Somatic | Unknown |
| p.L131S | c.392T>C | Kidney | Carcinoma | Variant of unknown origin | Unknown |
| p.A132V | c.395C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.A132V | c.395C>T | Central nervous system | Glioma | Confirmed Somatic | Unknown |
| p.A132V | c.395C>T | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.P133A | c.397C>G | Small intestine | Carcinoid-endocrine tumour | Confirmed Somatic | Unknown |
| p.K148R | c.443A>G | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.K148N | c.444G>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.E196* | c.586G>T | Thyroid | Carcinoma | Confirmed Somatic | Unknown |
| p.T201M | c.602C>T | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.P210R | c.629C>G | Ovary | Other | Confirmed Somatic | Unknown |
| p.H222Y | c.664C>T | Urinary tract | Carcinoma | Variant of unknown origin | Unknown |
| p.P225S | c.673C>T | Skin | Carcinoma | Confirmed Somatic | Heterozygous |
| p.H230R | c.689A>G | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.P236R | c.707C>G | Ovary | Carcinoma | Variant of unknown origin | Heterozygous |
| p.A240T | c.718G>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.A240T | c.718G>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.A240T | c.718G>A | Large intestine | NS | Confirmed Somatic | Unknown |
| p.A240T | c.718G>A | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.R250* | c.748C>T | Oesophagus | Carcinoma | Previously Reported | Unknown |
| p.R250* | c.748C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.L258* | c.773T>G | Large intestine | Carcinoma | Confirmed Somatic | Heterozygous |
| p.L258* | c.773T>G | Large intestine | Carcinoma | Previously Reported | Heterozygous |
| p.E275* | c.823G>T | Prostate | Carcinoma | Confirmed Somatic | Unknown |
| p.Y279C | c.836A>G | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.K313N | c.939A>T | Central nervous system | Glioma | Confirmed Somatic | Unknown |
| p.D314Y | c.940G>T | Kidney | Carcinoma | Confirmed Somatic | Unknown |
| p.P317S | c.949C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.C319Y | c.956G>A | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.L320F | c.958C>T | Skin | Carcinoma | Confirmed Somatic | Heterozygous |
| p.R321W | c.961C>T | Central nervous system | Glioma | Confirmed Somatic | Unknown |
| p.R327C | c.979C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.D338N | c.1012G>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.Q344E | c.1030C>G | Lung | Carcinoma | Confirmed Somatic | Unknown |
| p.R352H | c.1055G>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.G374A | c.1121G>C | Lung | Carcinoma | Confirmed Somatic | Unknown |
| p.S380N | c.1139G>A | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.S380N | c.1139G>A | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.A390S | c.1168G>T | Prostate | Carcinoma | Confirmed Somatic | Unknown |
| p.Y391S | c.1172A>C | Central nervous system | Glioma | Confirmed Somatic | Unknown |
| p.R406K | c.1217G>A | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.P413Q | c.1238C>A | Kidney | Carcinoma | Confirmed Somatic | Unknown |
| p.S415Y | c.1244C>A | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.S415F | c.1244C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.S415Y | c.1244C>A | Large intestine | Carcinoma | Variant of unknown origin | Unknown |
| p.L416F | c.1246C>T | Skin | Malignant melanoma | Confirmed Somatic | Heterozygous |
| p.A419V | c.1256C>T | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.S426C | c.1277C>G | Upper aerodigestive tract | Carcinoma | Variant of unknown origin | Unknown |
| p.L428S | c.1283T>C | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.R439* | c.1315C>T | Lung | Carcinoma | Confirmed Somatic | Unknown |
| p.G450D | c.1349G>A | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.C458Y | c.1373G>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.F462V | c.1384T>G | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.V464G | c.1391T>G | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.Q471* | c.1411C>T | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.Q471* | c.1411C>T | Urinary tract | Carcinoma | Confirmed Somatic | Unknown |
| p.Q506R | c.1517A>G | Thyroid | Carcinoma | Confirmed Somatic | Unknown |
| p.R507L | c.1520G>T | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.R507G | c.1519C>G | Ovary | Carcinoma | Confirmed Somatic | Heterozygous |
| p.E508D | c.1524G>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.L510V | c.1528T>G | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.L510S | c.1529T>C | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.S523C | c.1568C>G | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.T534S | c.1601C>G | Lung | Carcinoma | Variant of unknown origin | Unknown |
| p.R538S | c.1614G>C | Liver | Carcinoma | Confirmed Somatic | Unknown |
| p.R540* | c.1618C>T | Endometrium | Carcinoma | Previously Reported | Heterozygous |
| p.R540* | c.1618C>T | Endometrium | Carcinoma | Previously Reported | Heterozygous |
| p.R540* | c.1618C>T | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.R540* | c.1618C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.R540Q | c.1619G>A | Cervix | Carcinoma | Confirmed Somatic | Unknown |
| p.E569D | c.1707A>C | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.E574Q | c.1720G>C | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.E574Q | c.1720G>C | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.E574Q | c.1720G>C | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.F581I | c.1741T>A | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.P594S | c.1780C>T | Endometrium | Carcinoma | Confirmed Somatic | Unknown |
| p.P594L | c.1781C>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.L611S | c.1832T>C | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.L611F | c.1833G>T | Breast | Carcinoma | Confirmed Somatic | Unknown |
| p.K612R | c.1835A>G | Endometrium | Carcinoma | Variant of unknown origin | Heterozygous |
| p.Q621* | c.1861C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.P628L | c.1883C>T | Skin | Malignant melanoma | Confirmed Somatic | Unknown |
| p.P628L | c.1883C>T | Skin | Malignant melanoma | Previously Reported | Heterozygous |
| p.S637A | c.1909T>G | Large intestine | Carcinoma | Variant of unknown origin | Heterozygous |
| p.E648G | c.1943A>G | Stomach | Carcinoma | Confirmed Somatic | Unknown |
| p.S654F | c.1961C>T | Cervix | Carcinoma | Confirmed Somatic | Unknown |
| p.R670H | c.2009G>A | Large intestine | Carcinoma | Confirmed Somatic | Unknown |
| p.E678* | c.2032G>T | Large intestine | Carcinoma | Confirmed Somatic | Unknown |

**Table 3.2** – PAN3 mutations collected from the COSMIC database.

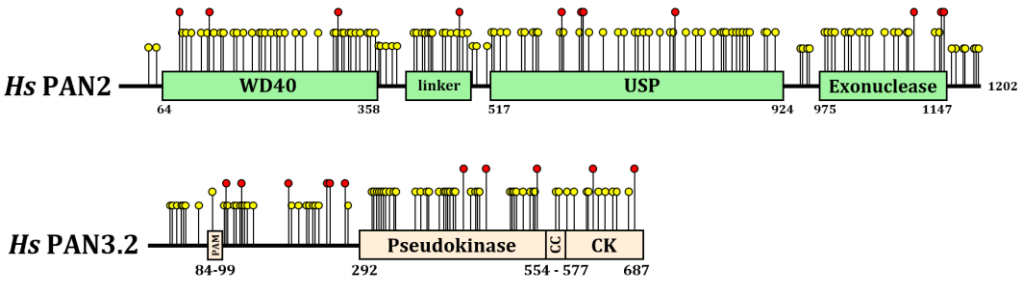



**Figure 3.1** – Schematic representation of the mutations distribution in the PAN2 and PAN3 genes. Missense mutations are labelled in yellow, nonsense mutations in red.

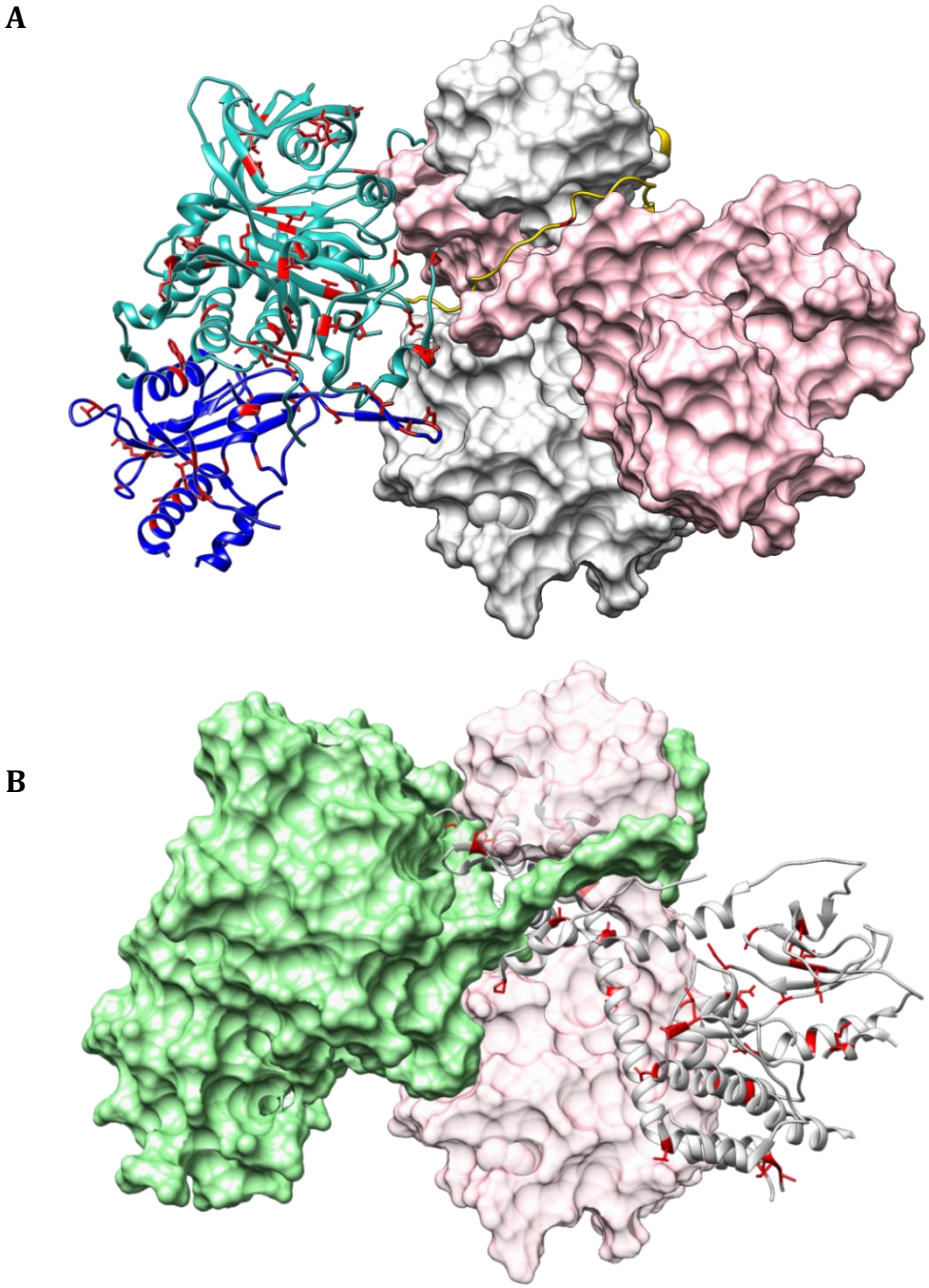



**Figure 3.2** – Structural representation of the PAN2-PAN3 complex. (A) *S. cerevisiae* PAN2-PAN3 complex displaying residue substitutions retrieved from COSMIC. PAN2 (ribbon) displays residue changes (red) in the RNase domain (blue), USP domain (cyan) and anchor (yellow). PAN3 forms a homodimer (pink and grey). (B) Amino acid substitutions (red) present in PAN3 (grey, ribbon) in complex with PAN2 (green) and PAN3 (pink).

The combination of the cancer-associated PAN2 variants modelled in the structures and the sequence conservation retrieved from ConSurf resulted in the identification of 14 variant residues that are highly conserved and altered in tumour samples (Fig. 3.4). The first variant residue (D410E) is positioned in the hinge, within a highly conserved stretch. Then, seven variant residues (R451C, N512D, N529S, G561D, C718S, L864M, A902D) are located in the core region, being either at the interface with the PAN3 homodimer or buried within the structure, found in α-helix or β-sheet structures. Finally, six variant residues (R1010W, P1019S, V1031L, D1045G, G1079D, T1099I) are located in the RNase domain.

The ConSurf analysis of PAN3 indicates that the most conserved residues are those forming the homodimer and those interacting with PAN2 (Fig. 3.3B). Similarly to PAN2, highly conserved residues that are altered in cancer-associated variants do not appear to cluster in specific domains of the proteins (Fig. 3.1). A total of 14 cancer-associated variants displayed amino acid substitutions at conserved positions. The first variant is located in the N-terminus of the pseudokinase domain (Y279C). Seven mutations are within the pseudokinase domain (D314Y, L320F, R321W, S380N, S426C, G450D, F462V). Two variants are located in the coiled coil region, responsible for the formation of the homodimer (E569D, E574Q). Finally, four variants are located within the C-terminal domain of PAN3 (F581I, L611S, E648G, S654F).

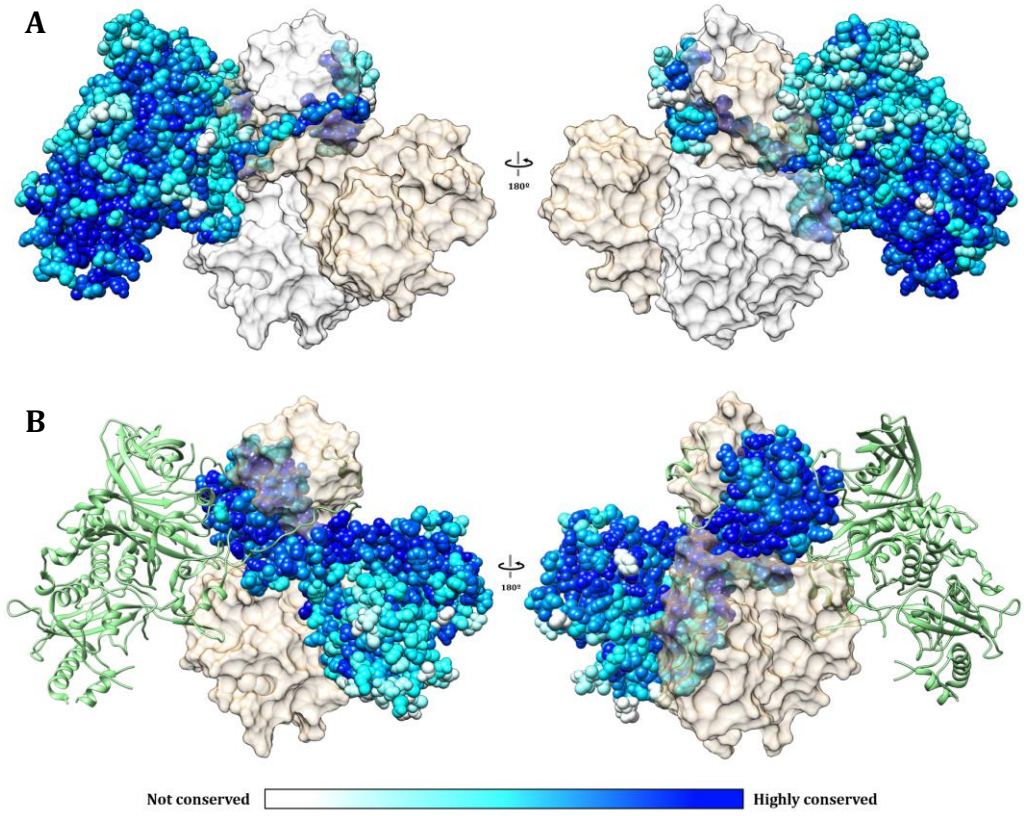



**Figure 3.3** – Conserved residues in the PAN2-PAN3 complex. Model showing the conserved residues for (A) PAN2 and (B) PAN3 as determined using the ConSurf server. A colour gradient shows residues from highly conserved (dark blue) to highly variable (white).

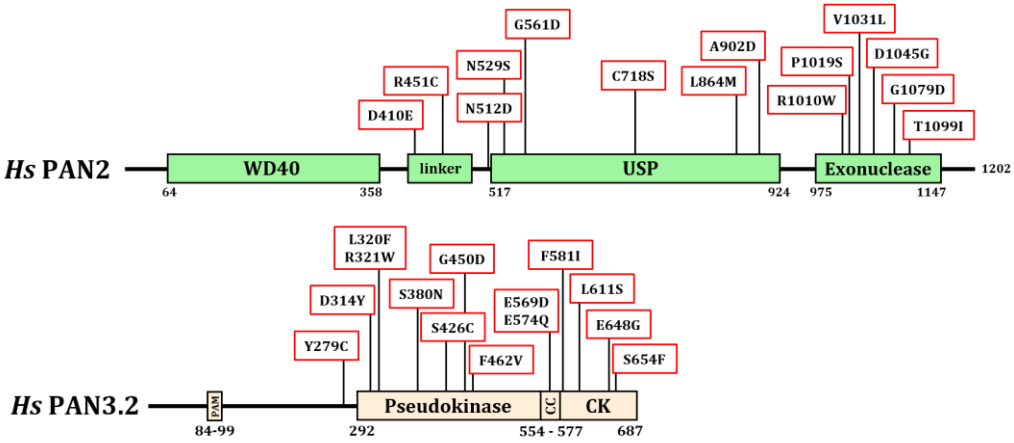



**Figure 3.4** –Schematic diagram indicating the distribution of cancer-associated variants at highly conserved positions in PAN2 and PAN3. The mutations appear to be spread throughout the domains in both proteins. The amino acid substitutions in the N-terminus of both proteins could not be modelled as residues 1-340 of PAN2 and residues 1-228 of PAN3 are not present in the *S. cerevisiae* structure.

## 3.3 Generation of PAN2 knockout HEK293 stable cell lines

### *3.3.1 Establishing a CRISPR/Cas9 protocol to target genes: a proof of principle*

CRISPR/Cas9 system was tested in Human Embryonic Kidney 293 cells (HEK293) to knock out the EMX1 gene using a previously reported guide RNA (Ran et al. 2013). To target the EMX1 gene, a vector pX459 was used, overexpressing Cas9 (Figure 3.5).

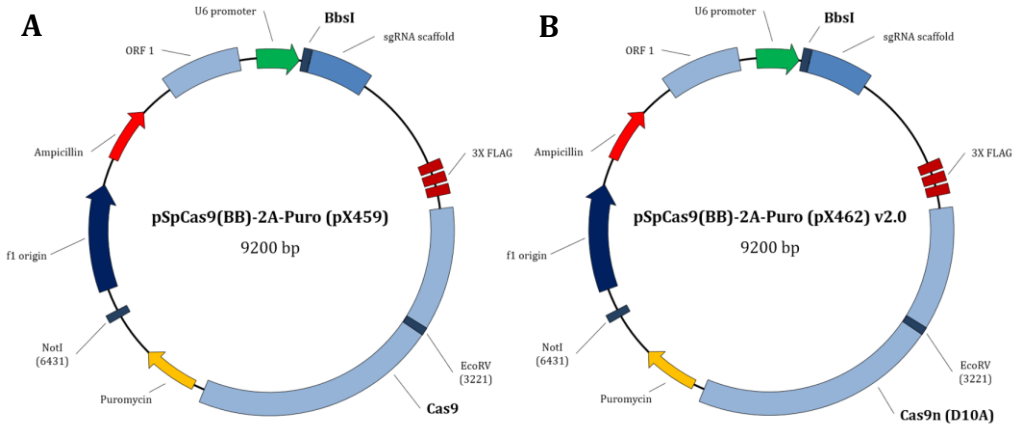
HEK293 cells were transfected with plasmids pX459 or pX459-EMX1. Positive transfected cells were selected with puromycin for 24 h and diluted 1:20 to obtain single clones. Ten clones per transfection were isolated and grown for two-three weeks and genomic DNA was extracted from approximately 6 million cells. The genomic locus of interest was amplified by PCR and the cell lines were genotyped by Sanger sequencing (Fig. 3.6A).

Two of the five clones analysed displayed an altered EMX1 gene sequence. Clone A presented a 31 bp deletion in Exon 3, while clone B showed a 54 bp insertion (Fig. 3.6A). Insertions or deletions are a natural consequence of the NHEJ repairing mechanisms that cells trigger when double-strand breaks are detected.

### *3.3.2 Targeting the PAN2 gene*

After confirming the efficacy of the CRISPR/Cas9 system in HEK293 cells, gRNAs targeting the N-terminal domain of PAN2 were designed to engineer knock-out cell lines. In order to reduce the number of off-target effects, a plasmid pX462 was used (Fig. 3.5B). This plasmid expresses a nickase Cas9, which induces a single strand break. Therefore, cells were transfected with two plasmids harbouring one gRNAs each. This pair of gRNAs was designed to target the DNA locus within 20-30 bp in order for the double nick to be sensed as a double-strand break by the cell.

The gRNA pair with the highest score was designed to target the second exon of PAN2. The Cas9n system has a lower efficiency, as determined by a screen on the population (Fig. 3.7A) therefore the gRNAs for this system were also cloned in vector pX459 (Fig. 3.7B). Twenty clones were analysed after transfection; ten clones generated by cells expressing Cas9 (1-10) and ten clones generated by expression of Cas9n (11-20). Only one clone (Clone 6) displayed an altered genomic DNA sequence in the locus of interest, having a 410 bp deletion (Fig. 3.6B).




| gRNA | Sequence |
|---|---|
| EMX1 | GTCACCTCCAATGACTAGGG |
| PAN2_1 | GCTGTGGGTGGGGAGCCACG |
| PAN2_2 | TCTCCCACTTTGACTTGCACG |

**Figure 3.5** – Plasmids used for CRISPR/Cas9 genome editing. (A) Plasmid pX459 (Addgene #48139) was used to target genes with Cas9 and select transfected cells with puromycin. (B) Plasmid pX462 V2.0 (Addgene #62987) is similar to pX459, but expresses a Cas9 mutant (Cas9n) that cleaves only one DNA strand. For both plasmids, gRNAs were inserted in the BbsI site.

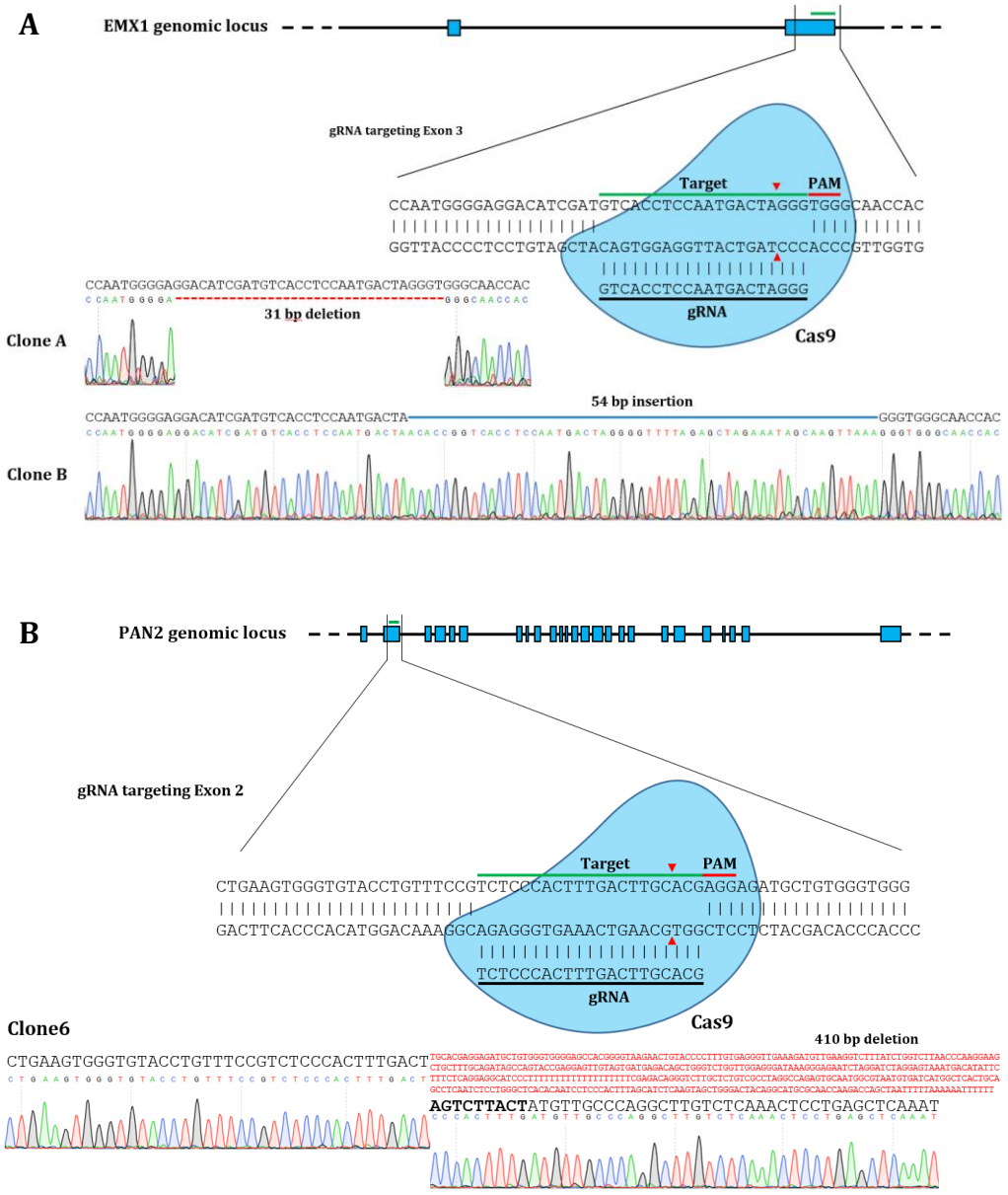



**Figure 3.6** – Analysis of cell lines edited by CRISPR/Cas9. (A) Disruption of EMX1 in HEK293 cells. The third exon of the EMX1 gene was targeted with a previously reported gRNA. Two of the screened clones displayed altered DNA sequences. (B) Disruption of PAN2 in HEK293 cells. In Clone 6, Cas9 induced a double-strand break that was repaired with a 410 bp deletion, resulting in an altered reading frame of PAN2.

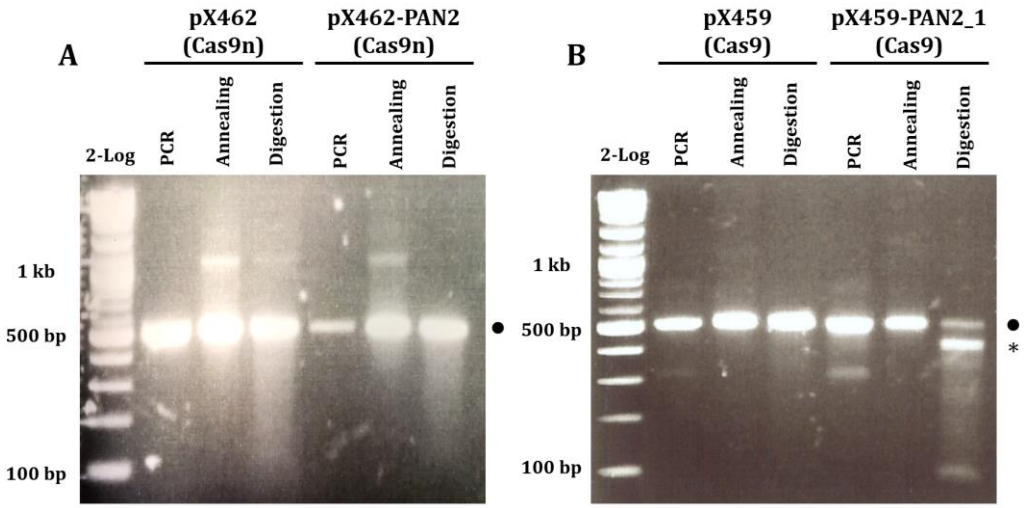



**Figure 3.7** – Analysis of the gRNAs efficiency in a population of cells. Genomic DNA was extracted from populations of cells transfected with (A) an empty vector pX462 (Cas9n) or a vectors pX462 carrying gRNAs targeting PAN2 or (B) an empty vector pX459 (Cas9) or a vector pX459 carrying the PAN2_1 gRNA. The targeted locus in the genome was amplified by PCR; PCR products (500 bp, indicated with a black dot) were then heated at 95°C and slowly cooled to room temperature, to allow for heteroduplexes to form. A digestions step with T7 endonuclease cleaves the double strand in presence of a mismatch caused by the Cas9 activity (B, indicated with a star).

## 3.4 Investigation of critical mutations in the PAN2-PAN3 complex

### *3.4.1 Potential amino acids mediating the PAN2-PAN3 interaction and the PAN3-PAN3 homodimer formation*

Yeast two-hybrid experiments are well established approaches to detect protein-protein interactions. The aim of these assays was to assess whether amino acid changes found in cancer-associated PAN2 and PAN3 variants impacted the interactions between PAN2 and PAN3. Previously reported mutations (Christie et al. 2013) that were confirmed to impair the interactions between PAN2 and PAN3 and to disrupt the PAN3 homodimer formation were used as positive controls. Specifically, two PAN3 mutants were prepared. The first (A) contained a triple amino acid substition F753E, L755A and A760E, which impairs formation of the PAN3 homodimer. The second (B) contained a double amino acid substitution R855T and D856T, which strongly affects the PAN3-PAN2 interaction.

### *3.4.2 Generation of plasmids for yeast two-hybrid experiments*

The PAN2 and PAN3 cDNAs were amplified by PCR to include an EcoRI restriction site at the 5' end and a SalI restriction site at the 3' end. Digested PCR products were cloned into pDB-Gal4 Cam and pAD-Gal4 Cam for yeast expression (Fig. 3.8). The desired groups of point mutations were introduced with a single step inverse PCR (Hemsley et al. 1989) and confirmed by Sanger sequencing.

### *3.4.3 Identification of residues involved in the PAN2-PAN3 interaction and the PAN3-PAN3 homodimer formation*

Yeast two-hybrid analysis was carryed out to assess the impact of specific residue changes on the interactions within the PAN2-PAN3 complex. The relevant plasmids were used to transform yeast competent cells (strain YRG-2), then grown on selective media (-Leu/-Trp). Three colonies for each transformation were analysed with the β-galactosidase assay (Promega).

Vectors expressing only the Gal4 binding domain and the Gal4 activation domain were used as negative controls, and confirmed a low background β-galactosidase activity (Figure 3.9). Similarly, either PAN2 or PAN3 alone did not induce any β-galactosidase expression (Figure 3.9). As expected, activity was detected upon expression of Gal4.AD-PAN2 and Gal4-DBD.PAN3 and Gal4.AD-PAN3 and Gal4.DBD-PAN3 fusion proteins, or the reciprocal fusions with Gal4.AD and Gal4.DBD. However, the activty of β-galactosidase was greatly reduced in presence of specific amino acid changes in PAN3. Expression of PAN3 mutant A (containing the triple amino acid substitution F753E, L575A and A760E) did not show β-galactosidase activity in presence of wild type PAN3, indicating that the PAN3 homodimer formation was impaired (Fig. 3.9B). By contrast mutant B (containing the double amino acid substitution R855T, D856T) impaired the interaction between PAN2 and PAN3 (Fig. 3.9A).

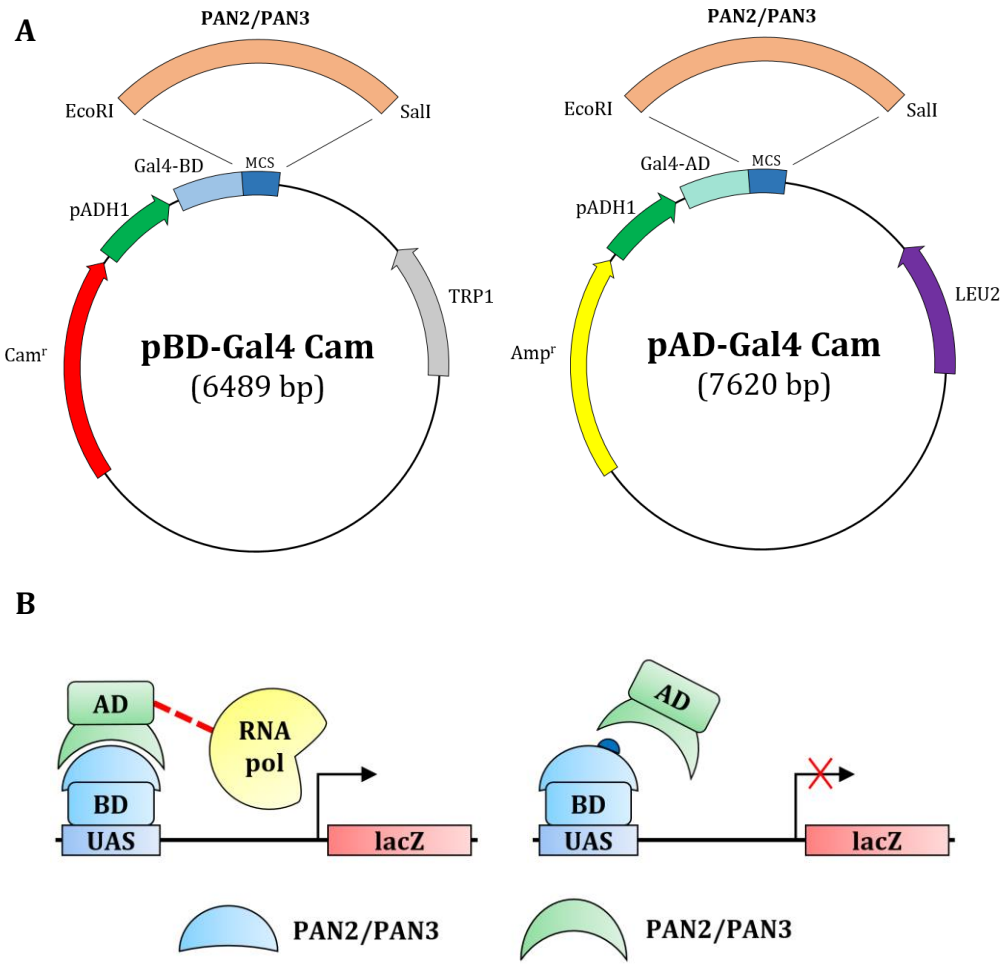



**Figure 3.8** – Overview on the yeast two-hybrid experiments. (A) Vectors generated for yeast two-hybrid analysis. Both PAN2 and PAN3 cDNAs were amplified by PCR using primers including a 5' EcoRI and 3' SalI restriction site. After digestions, PCR products were cloned into vectors linearised with the same restriction enzymes. (B) Representation of the Gal4 yeast two-hybrid system. Two proteins of interest are fused with either Gal4 DNA binding domain (BD) or Gal4 activation domain (AD). When interactions occur, the Gal4 domains are adjacent and induce lacZ transcription and therefore β-Galactosidase expression. If no interactions occur, the lacZ gene is not expressed.

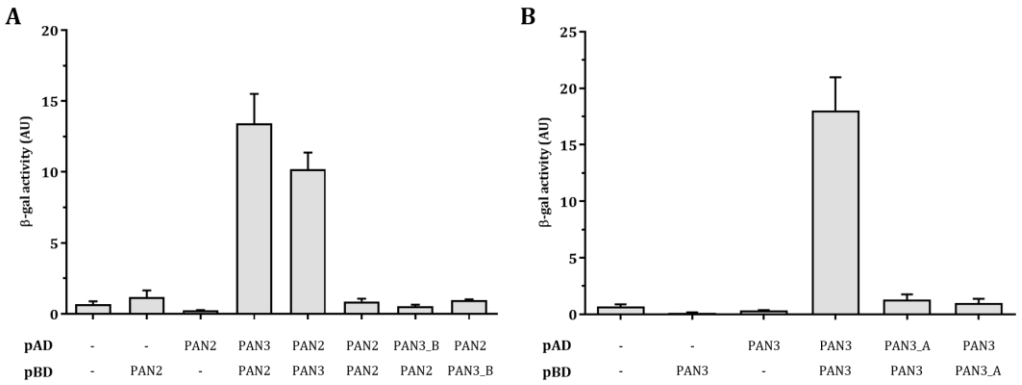



**Figure 3.9** – Analysis of the yeast two-hybrid assays. Yeast YRG-2 cells were transformed with the indicated vectors. β-Galactosidase activity was normalised against $OD_{600}$. Error bars represent the standard error of the mean (n=3).

## 3.5 Expression and purification of the human PAN2-PAN3 complex

### *3.5.1 Vectors for protein expression in* E. coli *BL21 DE3 strain*

PAN2 and PAN3 codon-optimised cDNA sequences for protein expression in bacteria were designed and synthesised (GenScript). The PAN2 cDNA sequence was split into two fragments (PAN2N and PAN2C). PAN2N was cloned into a pUC57 vector using BamHI and HindIII restriction sites at the 5' and 3' end, respectively. PAN2C was cloned in a pUC57 vector using HindIII and SalI sites at the 5' and 3' end, respectively. PAN3 sequence was cloned into a pUC57 vector, with BamHI and SalI restriction sites at the 5' and 3' end, respectively. As PAN2 and PAN3 form a trimeric complex, vector pACYC-Duet1 was selected, because it allows for the co-expression of two proteins, which may help protein folding and complex assembly. This plasmid contains two multiple cloning sites (MCS) downstream of T7 promoters. The first MCS is in frame with an N-terminal 6His tag and MCS 2 is in frame with a C-terminal S-tag.

Vectors pUC57 containing PAN2 or PAN3 cDNAs were digested with the appropriate restriction enzymes and cloned in two steps into linearised pACYC-Duet1 vectors. In the first step, pACYC-Duet1 was linearised with BamHI and SalI and either both fragments of PAN2 or PAN3 were cloned into MCS 1; in the second step, this vector was linearised with BglII and XhoI (isoschizomers of BamHI and SalI, respectively) and the second cDNA was cloned into MCS 2. Two pACYC-Duet1 vectors were prepared: the first containing PAN2 in MCS 1 and PAN3 in MCS 2; the second with PAN3 in MCS 1 and PAN2 in MCS 2. Only the 6His-tag at the N-terminal of MCS 1 was cloned in frame with the relative cDNA (Fig. 3.10).

### *3.5.2 PAN2-PAN3 complex expression in* E. coli *BL21 DE3*

A single colony of *E. coli* BL21 DE3 containing either vector pACYC-Duet1 L01 (MCS1-PAN2 / MCS2-PAN3) or pACYC-Duet1 L02 (MCS1-PAN3 /

MCS2-PAN2) was used to inoculate a 2 mL Lysogeny Broth (LB) culture with appropriate antibiotic for 4-6h. Twenty-five milliliters of LB were added to this culture and bacteria were incubated for 16h at 37°C. Then, 1 L of LB was added and cultures were grown at 37°C until the $OD_{600}$ was between 0.7-0.8. IPTG (final concentration of 0.2 mM) was added to induce protein expression. Cultures were incubated for either 3h at 30°C or 16h at 18°C and cells were harvested by centrifugation.

Both vectors (L01 and L02) used to transform *E. coli* BL21 DE3, then induced with IPTG for 3h at 30°C, yielded an insoluble PAN2-PAN3 complex (Fig. 3.11A). A similar result was obtained when proteins were expressed for 16h at 18°C (Fig. 3.11B). To confirm these findings, protein expressions were analysed by western blotting using antibodies against PAN2 (Fig. 3.12).

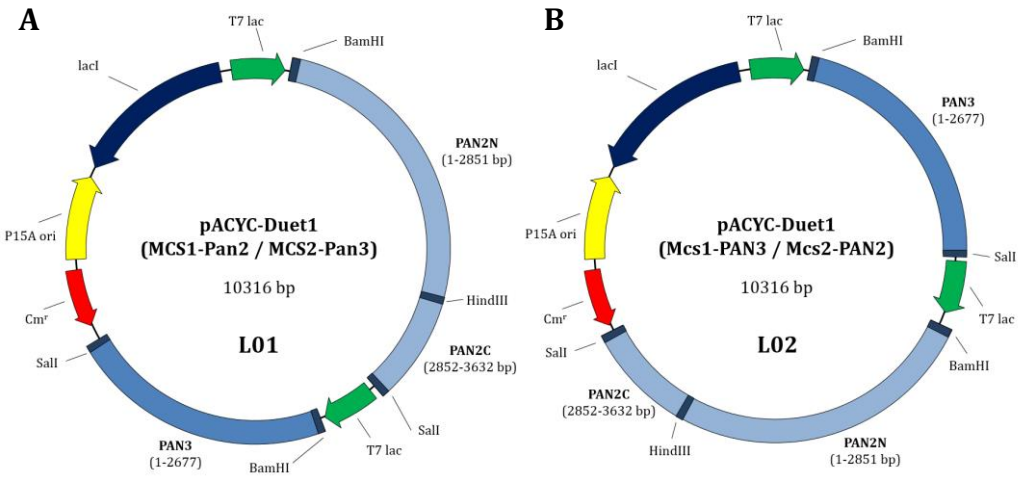



**Figure 3.10** - Vectors generated for the PAN2-PAN3 complex expression in *E. coli* BL21 DE3. Both PAN2 and PAN3 cDNAs were digested from the respective pUC57 plasmids using BamHI and HindIII (PAN2N), HindIII and SalI (PAN2C) and BamHI and SalI (PAN3). After digestions, the cDNAs were cloned in two steps into vectors linearized with BamHI and SalI, and then BglII and XhoI. (A) map of the L01 vector, expressing a 6His-tagged PAN2 and PAN3. (B) map of the L02 vector, expressing a 6His-tagged PAN3 and PAN2.

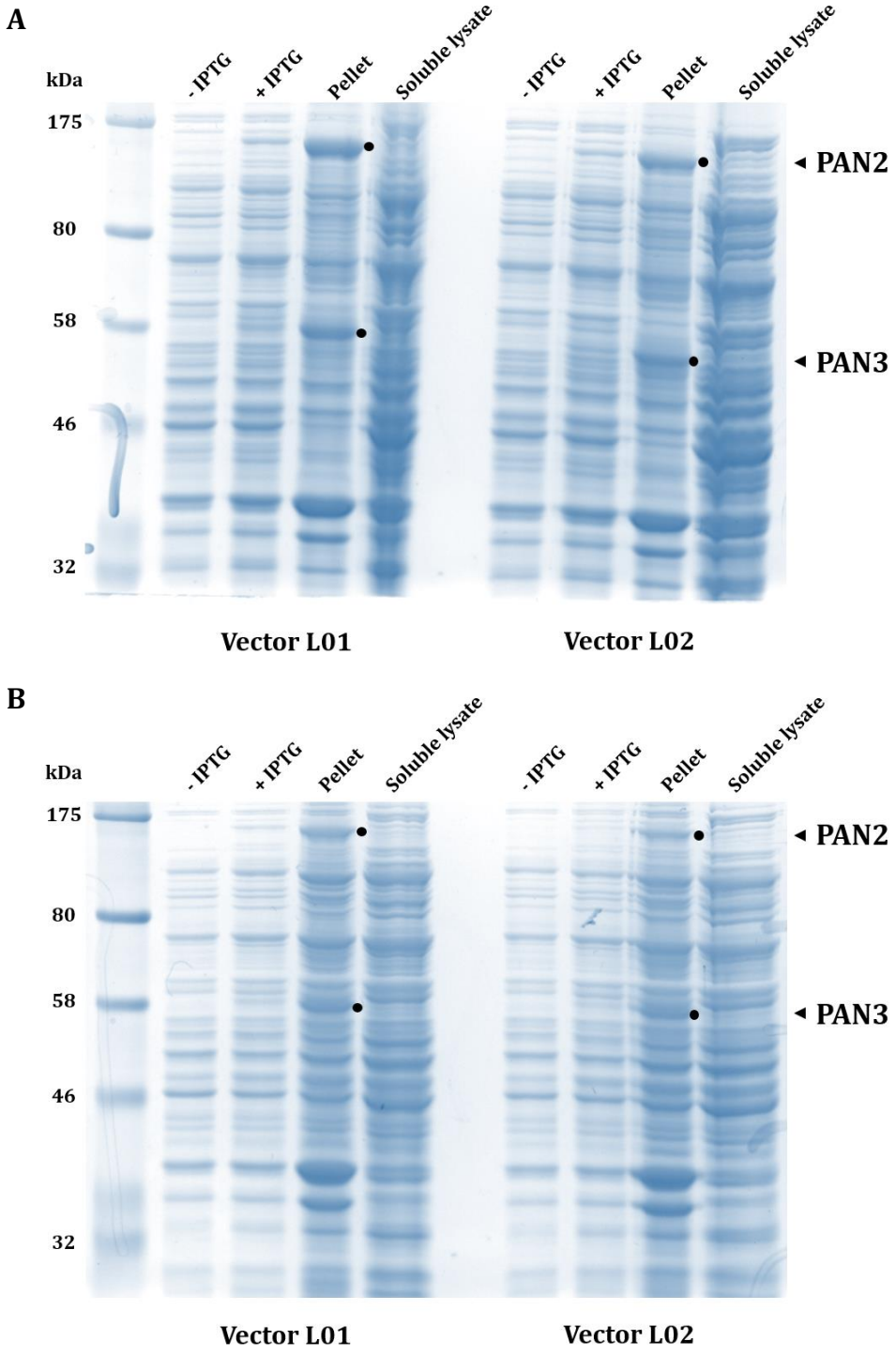



**Figure 3.11** – Expression of the PAN2-PAN3 complex in *E. coli* BL21 DE3. After IPTG addition (0.2 mM) cultures were incubated for 3h at 30°C (A) or for 16h at 18°C (B). Both incubation conditions yielded PAN2-PAN3 complex in the insoluble fraction for both vectors (L01 and L02).

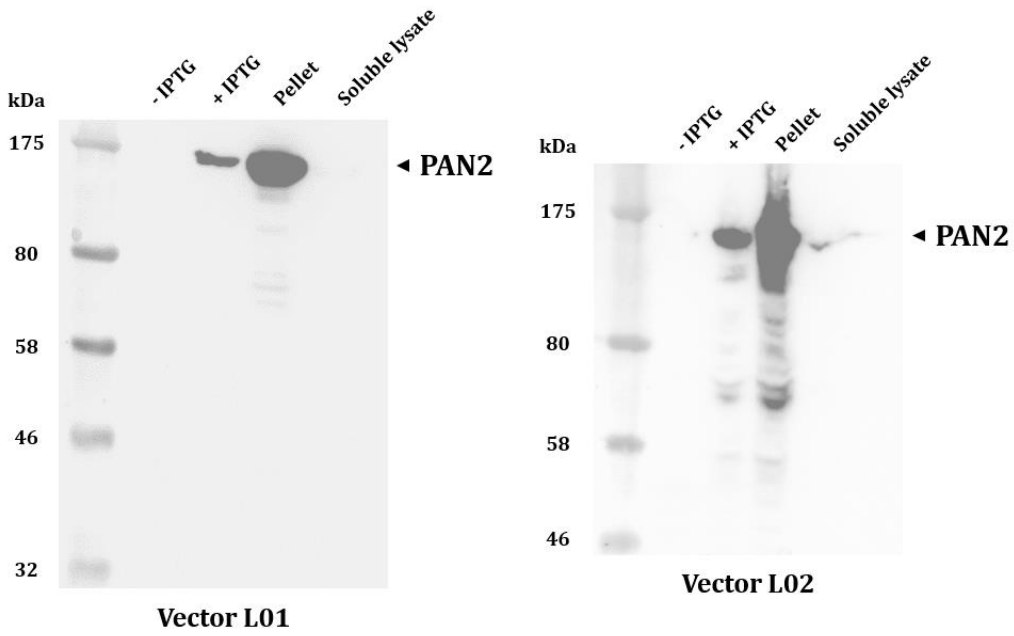



**Figure 3.12** – Western blots analysis of the protein expression cultures (induction for 3h at 30°C). Anti-PAN2 antibodies were used to detect PAN2. With both vectors (L01 and L02) PAN2 presence is detected upon IPTG induction. When vector L01 is used, PAN2 is not detected in the soluble fraction, whereas with vector L02 a modest fraction of PAN2 is detected in the soluble lysate.

### *3.5.3 Vectors for protein expression in* E. coli *ArcticExpress DE3*

Several bacterial strains were designed to improve the solubility of recombinant proteins. The *E. coli* ArcticExpress DE3 strain was used because of its characteristic to co-express cold-adapted chaperonines Cpn10 and Cpn60 at temperatures between 10-13°C, that can help protein folding. Cpn10 and Cpn60 are constitutively co-expressed from a pACYC-based vector carrying a gentamycin resistance gene. For this reason, the PAN2 and PAN3 cDNAs were cloned into the co-expression vector pET-Duet 1, which has a compatible origin of replication.

The expression cassettes present in either L01 or L02 vectors were amplified by PCR using ACYCDuetUp1 and T7 Terminator primers. PCR products were digested using the NcoI and AvrII restriction sites at the 5' and 3' end, respectively, and cloned into a pET-Duet1 vector linearised with the same restriction ezymes. As a result, vector L03 is a pET-duet1 containing PAN2 in MCS 1 and PAN3 in MCS 2; vector L04 is a pET-duet1 containing PAN3 in MCS 1 and PAN2 in MCS 2 (Fig. 3.13).

### *3.5.4 PAN2-PAN3 complex expression in* E. coli *ArcticExpress DE3*

A single colony of *E. coli* ArcticExpress DE3 containing either vector L03 or L04 was used to inoculate a 2 mL Lysogeny Broth (LB) culture with appropriate antibiotics for 4-6h. Twenty-five milliliters of LB were added to this culture and bacteria were incubated for 16h at 37°C. Then, 1 L of LB was added and cultures were grown at 30°C until the $OD_{600}$ was between 0.7-0.8. IPTG (final concentration of 0.2 mM) was added to induce the proteins expression. Cultures were incubated for 24h at 12°C before cells were harvested by centrfugation..

Both vectors (L03 and L04) were used to transform *E. coli* ArcticExpress DE3, and reulted in PAN2 and PAN3 expression. However, all PAN2 and PAN3 proteins were present in the insoluble fraction (Fig. 3.14).

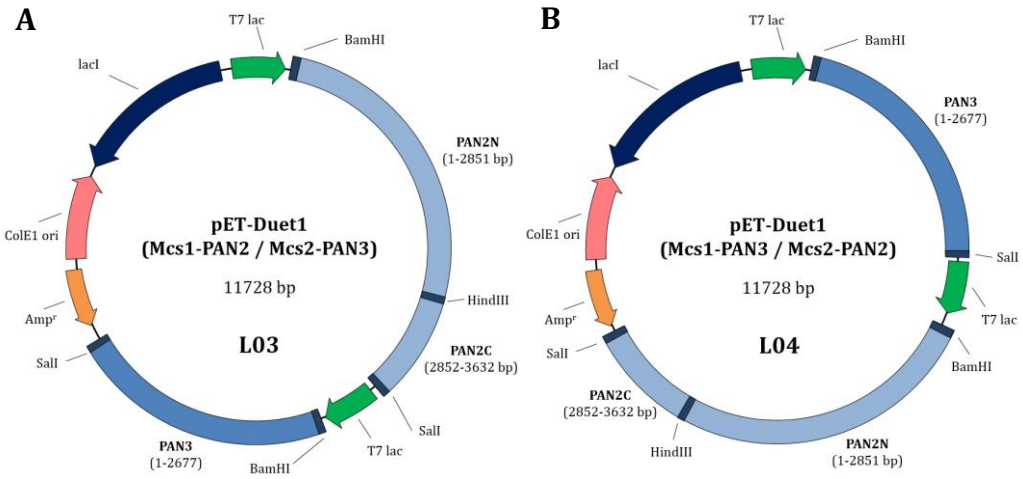



**Figure 3.13** - Vectors generated for the PAN2-PAN3 complex expression in *E. coli* ArcticExpress DE3. The expression cassettes were digested from vectors L01 and L02 with NcoI and AvrII restriction enzymes and cloned into vectors linearized with the same enzymes. (A) map of the L03 vector, expressing a 6His-tagged PAN2 and PAN3. (B) map of the L04 vector, expressing a 6His-tagged PAN3 and PAN2.

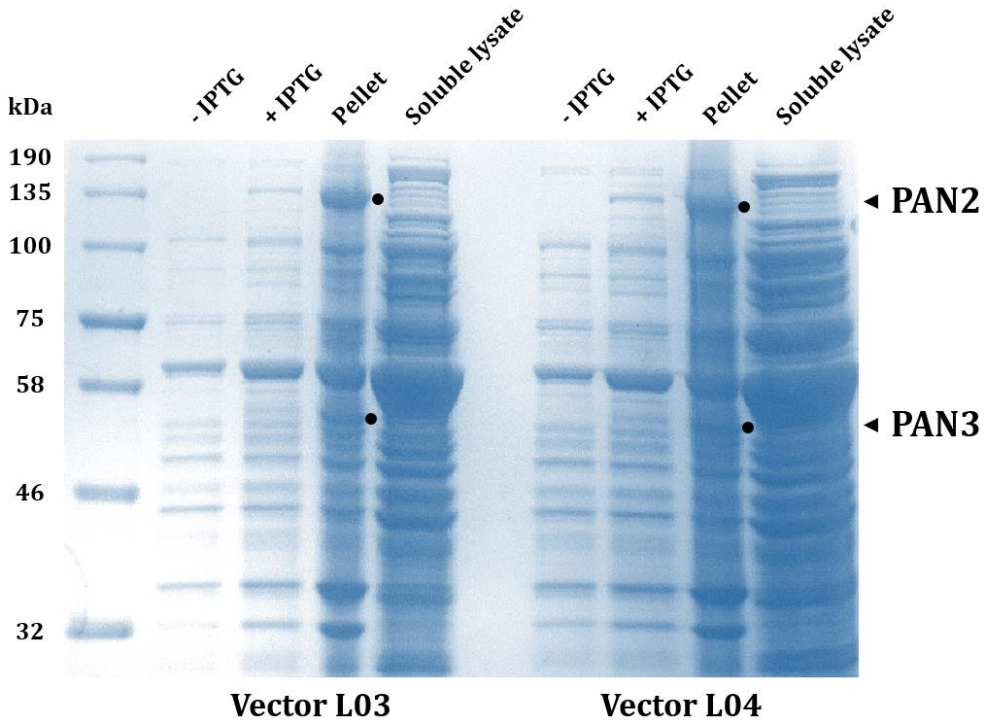



**Figure 3.14** - Expression of the PAN2-PAN3 complex in *E. coli* ArcticExpress DE3. After IPTG addition (0.2 mM) cultures were incubated for 24h at 12°C. Both incubation conditions yielded PAN2-PAN3 complex in the insoluble fraction (pellet) for both vectors (L03 and L04). The protein band at 60 kDa is Cpn60, constitutively expressed by the ArcticExpress strain.

### *3.5.5 Vectors for the expression of PAN2 and PAN3 fragments*

Removing disordered domains can increase the solubiity of recombinant proteins. I-TASSER (zhanglab.ccmb.med.umich.edu/I-TASSER) was used to predict structural characteristics of human PAN2 and PAN3. Two disordered regions are predicted in PAN2, at both N and C-terminal ends (Fig 3.15A). A PAN2 fragment (residues 57-1177) was amplified by PCR with primers including a BamHI site (5' end) and a SalI site (3' end) (Fig. 3.16). A long disordered N-terminal region of the PAN3 is predicted (Fig. 3.15B). Two PAN3 fragments were cloned (amino acids 200-887 and 445-887) using identical restriction sites (Fig. 3.16).

### *3.5.6 Expression of PAN2 and PAN3 fragments in* E. coli *BL21 DE3*

A single colony of *E. coli* BL21 DE3 containing vector L05 or L06 was used to inoculate a 2 mL LB culture for 4-6h. This culture was expanded (25 ml) and incubated for 16h at 37°C. Then, 1 L of LB was added and cultures were grown at 37°C until the $OD_{600}$ was 0.7-0.8. IPTG (final concentration of 0.2 mM) was added and cultures were incubated for 3h at 30°C before cells were harvested by centrifugation. Both vectors (L05 and L06) resulted in PAN2 and PAN3 fragments expression. However, these were in the insoluble fraction (Fig. 3.17).

### *3.5.7 PAN2 and PAN3 fragments expression in* E. coli *ArcticExpress DE3*

A single colony of *E. coli* ArcticExpress DE3 containing vector L05 or L06 was used to inoculate a 2 mL LB culture for 4-6h. Twenty-five milliliters of LB were added and bacteria were incubated for 16h at 37°C. Then, 1 L of LB was added and cultures were grown at 30°C until the $OD_{600}$ was 0.7-0.8. IPTG (final concentration of 0.2 mM) was added and cultures were incubated for 24h at 12°C before cells were harvested by centrifugation. Both vectors (L05 and L06) resulted in good expression of PAN2 and PAN3 fragments, that were however in the insoluble fraction (Fig. 3.17).

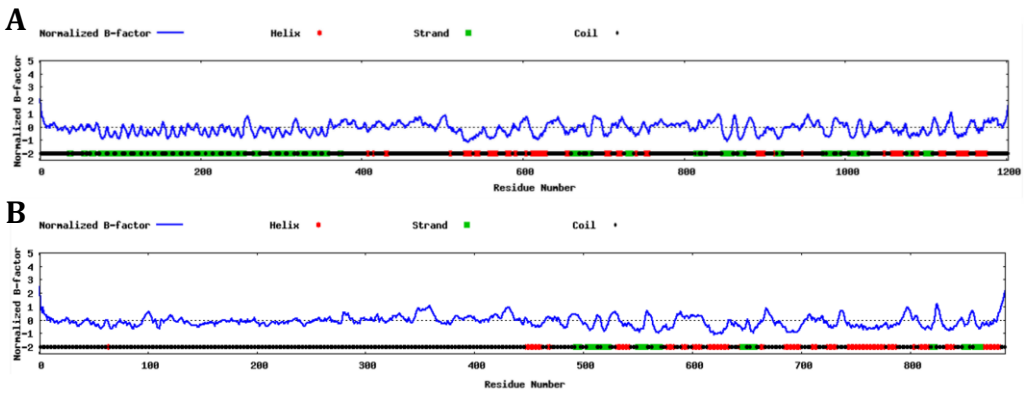



**Figure 3.15** –Prediction of the structured regions of both human PAN2 and PAN3 using the I-TASSER server. The server aligns a given sequence with available protein structures to predict structured regions of the protein of interest, such as α-helices (red) and β-strands (green). (A) PAN2 presents two coiled coils at both ends of the proteins, that could decrease protein solubility. (B) PAN3 is predicted to have an extended N-terminal coiled coil, that could be problematic to correctly fold.

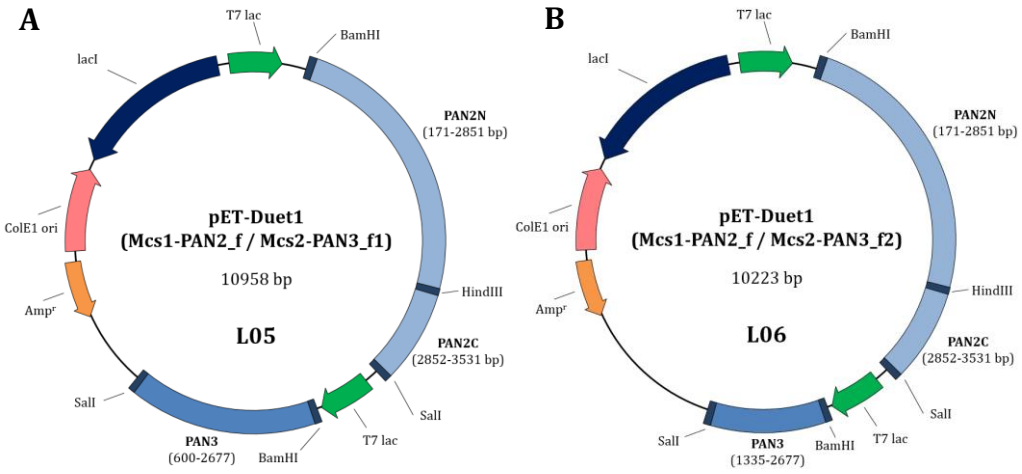



**Figure 3.16** – Vectors generated for the expression of a PAN2-PAN3 complex composed by fragments in *E. coli* BL21 DE3 and ArcticExpress DE3. Fragments of PAN2 and PAN3 cDNAs were amplified by PCR using primers introducing a BamHI restriction site at the 5' end and a SalI site at the 3' end. PCR products were digested with BamHI and SalI and the cDNAs were cloned in two steps into vectors linearized with BamHI and SalI for the MCS 1, and then BglII and XhoI for the MCS 2. (A) map of the L05 vector, expressing a 6His-tagged PAN2f 57-1177) and PAN3_f1 (200-887). (B) map of the L06 vector, expressing a 6His-tagged PAN2f (57-1177) and PAN3_f2 (445-887).

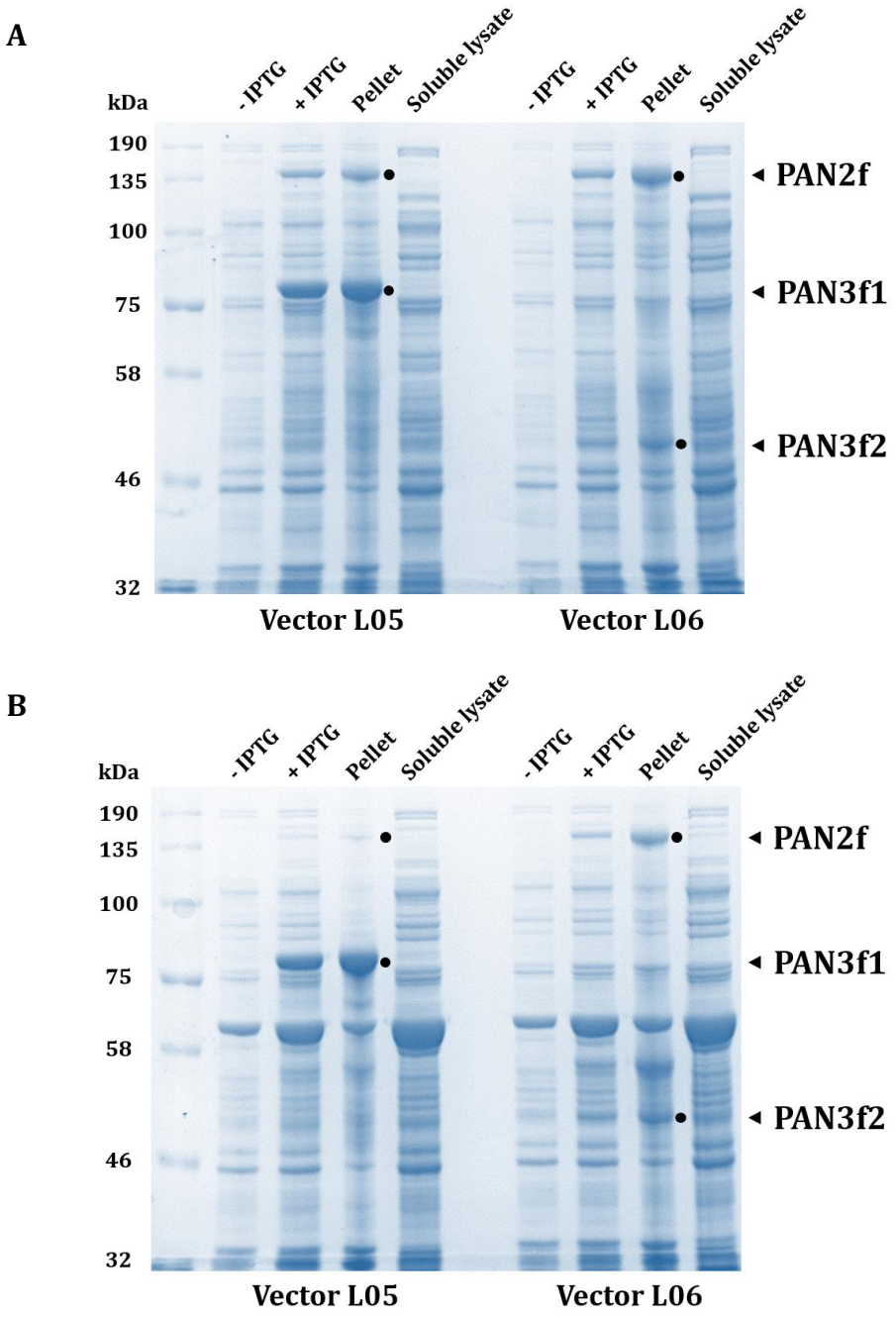



**Figure 3.17** - Expression of the PAN2-PAN3 complex in *E. coli* BL21 DE3 (A) and ArcticExpress DE3 (B). After IPTG induction (0.2 mM) cultures were incubated for 3h at 30°C (A) or for 24h at 12°C (B). Both incubation conditions yielded PAN2-PAN3 complex in the insoluble fraction (pellet) for both vectors (L05 and L06).

### *3.5.8 Vectors for co-expression of PAN2-PAN3 complex and PABPC1 in* E. coli *BL21 DE3*

Co-expression of binding partners can result in improved solubility of protein complexes. The PAN2-PAN3 complex expressed with vectors L01 and L02 previously described was co-expressed with a pQE80L vector including the CDS of PABPC1 (PolyA binding protein), that has been reported to bind PAN3. The full length PABPC1 CDS was amplified by PCR using primers including a BamHI site at the 5' and a SalI site at the 3' end. The CDS was then cloned into a pQE80L vector linearised with the corresponding restriction enzymes.

### *3.5.9 PAN2-PAN3 complex and PABP co-expression in* E. coli *BL21 DE3*

A single colony of *E. coli* BL21 DE3 containing either vector L01 or L02 (Fig. 3.9) and vector L07 (Fig. 3.17) was used to inoculate a 2 mL LB culture with appropriate antibiotic for 4-6h. Twenty-five milliliters of LB were added to this culture and bacteria were incubated for 16h at 37°C. Then, 1 L of LB was added and cultures were grown at 37°C until the $OD_{600}$ was between 0.7-0.8. IPTG (final concentration of 0.2 mM) was added, cultures were incubated for either 3h at 30°C before cells were harvested by centrfugation.

The co-expression of PABPC1 with the PAN2-PAN3 complex did not improve the complex solubility. All three expressed proteins were detected in the insoluble fraction (data not shown).

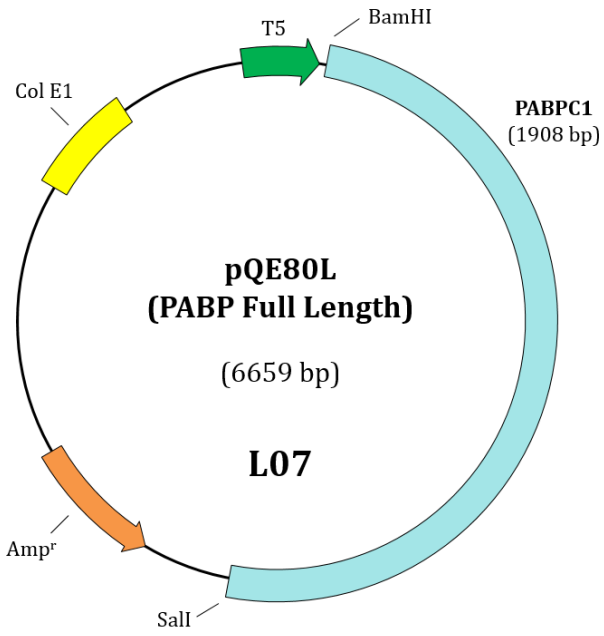



**Figure 3.18** – Vector L07 generated to express PABPC1 in *E. coli* BL21 DE3. PABPC1 (full length – 636 amino acids) CDS was amplified by PCR using primers introducing a BamHI restriction site at the 5' end and a SalI site at the 3' end. PCR products were digested with BamHI and SalI and cloned in pQE80L vector linearised with BamHI and SalI.

### *3.5.10 Vectors to express catalytically active PAN2 fragments*

A set of plasmid constructs were designed to express a catalytically active fragment of PAN2, in order to test its activity *in vitro*. Analysis of the structure of the *S. cerevisiae* PAN2 structure revealed extensive interactions between the exoribonuclease domain (amino acids 975-1147) and the USP domain (amino acids 517-924). This suggested that the presence of both domains could have a positive impact on the solubility of the recombinant fragment. Different PAN2 Deadenylase Fragments (DF) were designed starting from PAN2_f described above (Fig. 3.15) and cloned into vector pQE80L (Table 3.3). Vector L08 expresses a fragment of PAN2 without the N and C-terminal disordered loops. PAN2 DF in L09 lacks the WD40 domain. L10, L11 and L12 vectors express PAN2 fragments lacking different portions of the linker region and a shorter C-terminal domain. Structural analysis of the *N. crassa* PAN2 crystal revealed the presence of two disordered loops (Loop1 779-833 and Loop2 989-1005) that are not present in the human protein. The last two vectors (L13 and L14) were designed to express PAN2 DF (vector L12) but lacking either Loop1 or both loops, in order to try to improve the solubility of the recombinant fragment.

### *3.5.11 PAN2 fragments expression in* E. coli *BL21 DE3*

A single colony of *E. coli* BL21 DE3 containing a vector expressing a PAN2 DF (L08-L14, Table 3.3) was used to inoculate a 2 mL LB culture with appropriate antibiotic for 4-6h. Twenty-five milliliters of LB were added to this culture and bacteria were incubated for 16h at 37°C. Then, 1 L of LB was added and cultures were grown at 37°C until the $OD_{600}$ was between 0.7-0.8. IPTG (final concentration of 0.2 mM) was added, cultures were incubated for either 3h at 30°C before cells were harvested by centrfugation. The expression of different PAN2 deadenylase fragments did not result in soluble proteins. The PAN2 DF fragments were all detected in the insoluble fraction (data not shown).

| Vector | PAN2 fragment | Fragment length | Expression |
|---|---|---|---|
| L08 | 57-1177 | 1121 | Insoluble fraction |
| L09 | 350-1177 | 828 | Insoluble fraction |
| L10 | 480-1177 | 698 | Insoluble fraction |
| L11 | 501-1177 | 677 | Insoluble fraction |
| L12 | 501-1152 | 652 | Insoluble fraction |
| L13 | 501-1152 (ΔLoop1) | 598 | Insoluble fraction |
| L14 | 501-1152 (ΔLoop1-2) | 581 | Insoluble fraction |

**Table 3.3** - List of pQE80L vectors generated to express different PAN2 fragments. All PAN2 fragments were amplified by PCR using primers including a 5' end BamHI restriction site and a 3' end SalI restriction site. After PCR, cDNAs were digested and ligated into pQE80L vector linearised with BamHI and SalI.

## 3.6 Discussion

The PAN2-PAN3 complex is an RNA machinery conserved across several species (Wahle & Winkler 2013). Its involvement in RNA degradation has been characterised both *in vitro* and *in vivo* (Boeck et al. 1996; Uchida et al. 2004). Recent structural studies have discovered the architecture of this trimeric complex (PAN2-PAN3/PAN3) in model organisms (Jonas et al. 2014; Schäfer et al. 2014; Wolf et al. 2014), but no crystal structure of the human complex has been resolved yet.

Here, we aimed at characterising the human PAN2-PAN3 complex with a multipronged approach. PAN2-3 was previously linked to regulation of HIF1A (hypoxia-inducible factor 1α, which is involved in cancer progression (Bett et al. 2013). Here, we planned to assess the effect of a number of missense mutations in both PAN2 and PAN3 genes, collected from tumour samples, that could help understanding cancer mechanisms and the function of the PAN2-PAN3 complex. The collected mutations had FATHMM-MKL scores higher than 0.7, and therefore classified as 'pathogenic'; this algorithm predicts the phenotypic, functional and molecular consequences of selected missense variants. However, as this is only a prediction, it is possible that several of these are 'passenger' mutations only. Then, the impact of residue changes was to be characterised in four different ways. First, a computational analysis was carried out to screen the mutations and select those predicted to affect the stability of the complex or its function. Second, we planned to engineer HEK293 cell lines to introduce these mutations with CRISPR/Cas9, to study their impact in a more complex biological system. Third, we designed a yeast two-hybrid experiment to directly assess the impact of these mutations on the assembly of the human PAN2-PAN3 complex *in vitro*. Finally, we defined several approaches to purify a human PAN2-PAN3 protein complex in order to

determine whether the predicted mutations were having an impact on the deadenylase activity *in vitro*.

Nonsense and missense mutations in both PAN2 and PAN3 genes were collected from the Catalogue Of Somatic Mutations In Cancer. A total of 147 mutations of PAN2 (15 nonsense and 132 missense) were collected from the COSMIC database, and 98 in PAN3 (18 nonsense and 80 missense). The combination of further computational analysis performed using the ConSurf server and the modelling of the residue changes in the available crystal structures of the yeast PAN2-PAN3 complex allowed the selection of 14 mutation in PAN2 and 14 mutations in PAN3 as potentially impacting either the structure or the function of the complex. They are dispersed across the entire proteins, without any significant clustering. These mutations were to be characterised, along with a number of synthetic residue changes, either inactivating the activity of the complex or disrupting the formation of the human trimeric complex.

The CRISPR/Cas9 system used to engineer the genome of HEK293 cells needed extensive optimisation. As a proof of principle, we replicated previously reported results aiming at introducing small insertions or deletions in the EMX1 gene. Targeting the N-terminus of the PAN2 gene required the design of several gRNAs that were screened for efficiency within a population of cells. Eventually, a HEK293 cell line displaying an altered genome in the PAN2 locus was engineered. The impact of this alteration was tested via western blotting but led to inconclusive results, as the antibody to detect PAN2 was not showing sufficient specificity and affinity.

The designed yeast two-hybrid experiments were aimed at identifying the residue changes responsible for affecting the interaction between either PAN2 and PAN3 or within the PAN3 homodimer. Two clusters of residue

changes in PAN3 were previously reported to impact the interaction with PAN2 or PAN3, and were here used as controls. In this Gal4 yeast two-hybrid system, they strongly affected protein interaction and resulted in β-galactosidase activity levels comparable to the negative controls. However, these findings need to be analysed taking into account that western blotting was inconclusive, as the PAN2 and PAN3 antibodies were not showing enough affinity.

Lastly, we designed several expression and purification approaches to obtain a soluble human PAN2-PAN3 complex to characterise its deadenylase activity *in vitro* and the extent of the impact of the residue changes. All expression attempts resulted in an insoluble protein complex. At first, we tried to express the full complex in two *E. coli* DE3 strains (BL21 and ArcticExpress). Then, we expressed fragments of PAN2 and PAN3 that were not including disordered regions, to try to improve solubility. Finally, we tried to co-express PABPC1, a reported binding partner of PAN3, to try to stabilise the complex. Additionally, an expression of an exonuclease fragment of PAN2 was attempted, to try to characterise the catalytic activity of the active domain of the human PAN2 *in vitro*. All the protein expression approaches were unsuccessful and resulted in insoluble proteins, undermining the novelty addition of the other approaches. Future work could include different expression systems to obtain a recombinant human PAN2-PAN3 complex. For instance, baculovirus-mediated expression in insect cells or expression using human cell lines such as HEK293 could help overcoming the solubility issues.

# Chapter 4

## Expression and purification of a human Ccr4-Not core complex composed of BTG2, Caf1, Ccr4, CNOT1M and CNOT9

# Chapter 4. Expression and purification of a human Ccr4-Not core complex composed of BTG2, Caf1, Ccr4, CNOT1M and CNOT9

## 4.1 Introduction

The Ccr4-Not complex is an mRNA deadenylation machine conserved across the eukaryotes (Winkler & Balacco 2013). The human complex is composed of four modules. The N-terminus of CNOT1, the large scaffold protein, is bound by CNOT11, that engages interactions with CNOT10 (Bawankar et al. 2013). A central region of CNOT1 contains the DUF3819 domain of CNOT1 that recruits CNOT9, and the MIF4G domain that binds Caf1. In turn, Caf1 recruits Ccr4 and these two deadenylase enzymes form the nuclease module (Wahle & Winkler 2013). At the C-terminal end of CNOT1 there is the NOT module including CNOT2 and CNOT3 (Boland et al. 2013).

The mechanisms to target the Ccr4-Not complex to a specific mRNA have been extensively studied in the last decade. There are three main mechanisms: (i) generic recruitment via PABPC1; (ii) targeted recruitment mediated by RNA-binding proteins; and (iii) miRNA-mediated recruitment.

The purpose of this study was to reconstitute an *in vitro* deadenylation system to characterise the activity of the human Ccr4-Not complex and to understand the contributions of the individual modules. To this end, we assembled a core complex composed of five proteins. Ccr4, Caf`1 and BTG2 form the nuclease module, whereas CNOT1M and CNOT9 constitute the CNOT9 module. Two different assembly strategies were designed to reconstitute the complex. Firstly, a co-expression of the five selected proteins was attempted, using three plasmids carrying different origins of replication and genes providing resistance against antibiotics; both the StrepII and 6His affinity tags were evaluated to purify the core complex.

Secondly, a three-step purification approach was developed; in the first step, the single modules were purified separately with affinity purifications using the 6His-tag or the StrepII-tag. Then, the two modules were incubated to allow Caf1 to bind the MIF4G domain of CNOT1M and therefore reconstitute the core complex. Finally, three different purification strategies were designed to isolate a core complex with the correct stoichiometry; the first one aimed at isolating the core complex from the modules based on its larger molecular weight using size-exclusion chromatography, while the other two approaches utilised the two affinity tags (6His and StrepII) present.

The most efficient scheme involved the reconstitution of the core complex from the isolated nuclease and CNOT9 modules followed by isolation of a complex with equimolar stoichiometry by affinity purification using the StrepII-tag on Ccr4.

## 4.2 Expression and purification of the deadenylase enzyme Caf1 and the BTG2-Caf1 dimer

### *4.2.1 Vectors for protein expression in* E. coli *BL21 DE3*

To express Caf1, a vector was previously assembled (Maryati et al. 2014). Briefly, Caf1 was expressed transforming *E. coli* BL21 DE3 with a vector pQE80L P06 containing a human Caf1 codon-optimised cDNA for expression in *E. coli*, fused with a 6His-tag at the N-terminus (Fig. 4.1A). Caf1 cDNA was amplified by PCR from a vector pUC57 (GenScript) with primers containing a BamHI site at the 5' end and a SalI site at the 3' end. The amplification product was digested with BamHI and SalI and cloned into vector pQE80L linearised with identical restriction enzymes. The inactive mutant of Caf1 (D40A) was created by inserting a point mutation in the Caf1 cDNA using vector P06 as template. The presence of the mutation in vector P06M was confirmed by Sanger sequencing.

To co-express the dimer BTG2-Caf1, a vector pACYC-Duet1 was used. This vector allows for the co-expression of two proteins, which may help protein folding and complex assembly. This plasmid contains two multiple cloning sites (MCS) downstream of T7 promoters. The first MCS is in frame with an N-terminal 6His-tag and MCS 2 is in frame with a C-terminal S-tag. Briefly, Caf1 cDNA (either wild type or D40A) was amplified by PCR using vectors P06 or P06M. The PCR product was digested using BamHI and SalI restriction enzymes and cloned into a vector pACYC-Duet1 P10 linearised with identical restiction enzymes. Then, BTG2 cDNA was amplified by PCR using vector P02 as template (Fig. 4.2B), digested with BamHI and HindIII and cloned into the same pACYC-Duet1 P10. A similar cloning strategy was used to generate vector P10M, co-expressing BTG2 and Caf1 D40A. Sequences of both vectors P10 and P10M were confirmed by Sanger sequencing (Fig. 4.1A).

### *4.2.2 Caf1 and BTG2-Caf1 expression and purification*

A single colony of *E. coli* BL21 DE3 containing either vector P06 (or P06M) or P10 (or P10M) was used to inoculate a 2 mL LB culture with appropriate antibiotic for 4-6h. Twenty-five milliliters of LB were added to this culture and bacteria were incubated for 16 h at 37°C. Then, 1 L of LB was added and cultures were grown at 37°C until the $OD_{600}$ was between 0.7-0.8. IPTG (final concentration of 0.2 mM) was added to induce protein expression. Cultures were incubated for 16 h at 18°C and cells were harvested by centrifugation. Cells were resuspended in 5 mL Buffer HA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol, 5% glycerol and 20 mM imidazole) and lysed by sonication. The soluble lysate was isolated by centrifugation and then loaded on a HisTrap FF pre-packed column (GE) equilibrated with Buffer HA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol, 5% glycerol and 20 mM imidazole). After washing the unbound fraction, three elution fractions of 1 mL were collected using Buffer HB (as Buffer HA, but containing 250 mM imidazole). Typically, fraction number two, containing about 80% of the purified protein/complex, was used in subsequent steps. One microliter of sample was analysed by SDS-PAGE (Fig. 4.1C). The BTG2-Caf1 dimer was further purified by size-exclusion chromatography. Approximately 1 mL of sample was applied to a SEC column (GE Sephacryl S-100 HR) and 2 mL fractions were collected. Typically, 4-5 peak fractions were pooled and re-concentrated. One microlitre of sample was analysed by SDS-PAGE (Fig. 4.1C).

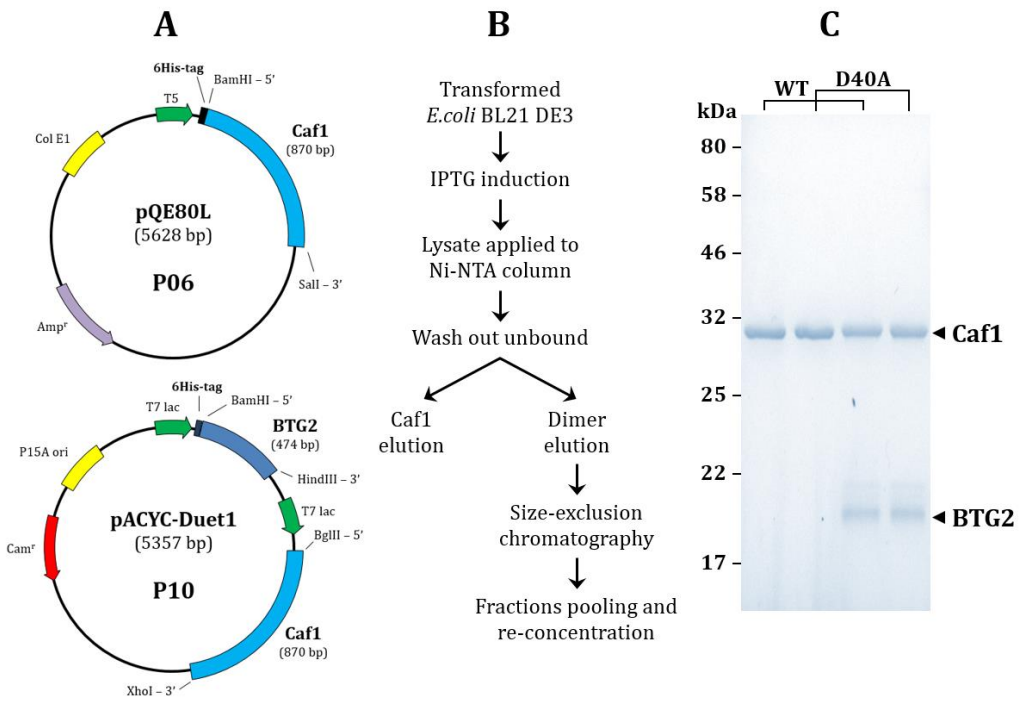



**Figure 4.1** – Caf1 expression in *E. coli* BL21 DE3. (A) map of vector P06, expressing 6His-tagged Caf1. (B) Outline of the purification strategy designed for Caf1 expression and purification. (C) Analysis of the purified deadenylase enzymes Caf1 (wild type and D40A) and BTG2-Caf1 (wild type and D40A) using 14% SDS-PAGE. Proteins were visualised by Coomassie staining.

## 4.3 Reconstitution of a nuclease module of the Ccr4-Not complex including Ccr4, Caf1 and BTG2

### *4.3.1 Vectors for protein co-expression in* E. coli *BL21 DE3*

To express and purify the nuclease module of the Ccr4-Not core complex, a modified scheme was used based on the procedure described before (Maryati et al. 2015). Briefly, three proteins were co-expressed using two vectors. First, vector pACYC-Duet1 P01 containing in MCS 1 a human Ccr4 codon-optimised cDNA for expression in *E. coli*, with a StrepII-tag at the N-terminus, and in MCS 2 a human Caf1 codon-optimised cDNA for expression in *E. coli* (Fig. 4.2A). Second, vector pQE80L P02 containing a human BTG2 cDNA cloned from HEK293 cells (Fig. 4.2B). In addition, a modified vector P01M was generated to express Ccr4 E240A and Caf1 D40A inactive mutants. Primers introducing single-point mutations were designed and two inverse PCR reactions were set up using vector P01 as template. The presence of both mutations in vector P01M was confirmed by Sanger sequencing.

### *4.3.2 Nuclease module expression and purification*

A single colony of *E. coli* BL21 DE3 co-transformed with vectors P01 and P02 was used to inoculate a 2 mL LB culture with antibiotics for 4-6 h. Fifty milliliters of LB were added to this culture and bacteria were incubated for 16 h at 37°C. Then, 2-4 L of LB were added and cultures were grown at 37°C until the $OD_{600}$ was between 0.7-0.8. IPTG (final concentration of 0.2 mM) was added to induce protein expression. Cultures were incubated for 14-16 h at 18°C and cells were harvested by centrifugation. Cells were resuspended in 5 mL Buffer SA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol and 5% glycerol) per litre of culture. Cells were lysed by sonication and the soluble lysate was isolated by centrifugation. The soluble lysate was then loaded on a gravity-flow column packed with Strep-Tactin beads (IBA) and equilibrated with Buffer SA. After washing the unbound

fraction, five elution fractions of 0.5 mL were collected using Buffer SB (Buffer SA supplemented with 10 mM desthiobiotin). Typically, fractions number two and three, containing about 90% of the purified complex, were pooled and re-concentrated. Approximately 1 mL of sample was applied to a column for size-exclusion chromatography (GE Sephacryl S-100 HR) and 2 mL fractions were collected. Typically, 4-5 fractions were pooled and re-concentrated. Five micrograms of sample were analysed by SDS-PAGE (Fig. 4.2C).

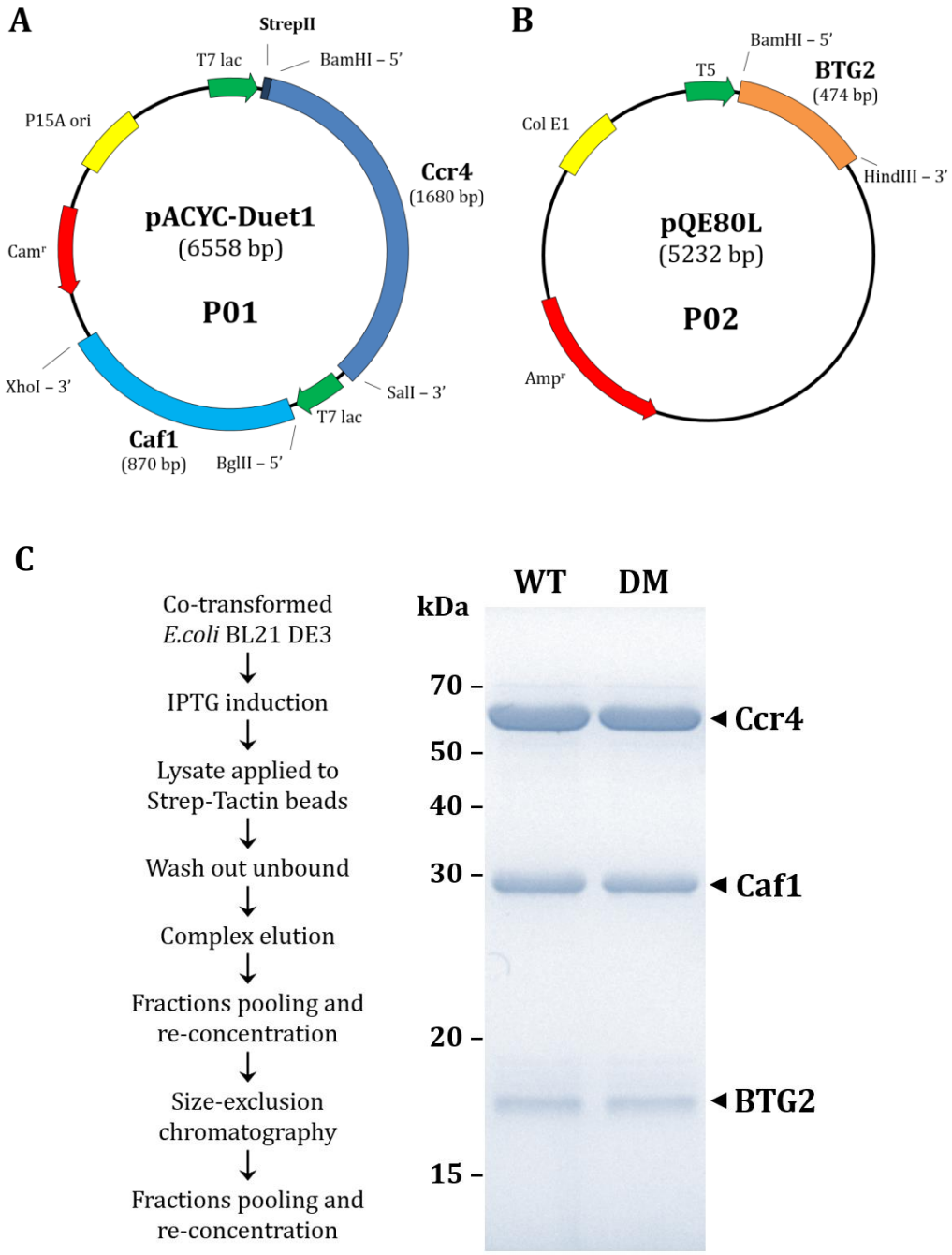



**Figure 4.2** – Vectors generated for the nuclease module expression in *E. coli* BL21 DE3. (A) map of vector P01, expressing StrepII-tagged Ccr4 and Caf1. (B) map of vector P02, expressing BTG2. (C) Outline of the purification strategy designed for the nuclease module and analysis of the purified nuclease module using 14% SDS-PAGE. Proteins were visualised by Coomassie staining.

## 4.4 Expression and purification of a CNOT1M/CNOT9 module

### *4.4.1 Vector for protein co-expression in* E. coli *BL21 DE3*

To reconstitute the CNOT9 module of the Ccr4-Not complex, a middle fragment of the scaffold protein CNOT1 (CNOT1M) was selected, including the MIF4G domain, tethering Caf1, and the DUF3819 domain, recognised by CNOT9. CNOT1M codon-optimised cDNA sequence for protein expression in bacteria was designed and synthesised (GenScript). The cDNA sequence was cloned into a pUC57 vector using BamHI and SalI restriction sites at the 5' and 3' end, respectively. The codon-optimised sequence of the ARM domain of CNOT9 (amino acids 33-283) was designed and synthesised for bacterial expression (GenScript). CNOT9 (ARM) sequence was cloned into a pUC57 vector, with BamHI and SalI restriction sites at the 5' and 3' end, respectively. Vector pACYC-Duet1 was selected for the co-expression of CNOT1M and CNOT9 (ARM).

Vectors pUC57 containing CNOT1M and CNOT9 (ARM) cDNAs were digested with the appropriate restriction enzymes and cloned in two steps into linearised vectors pACYC-Duet1. In the first step, pACYC-Duet1 was linearised with BamHI and SalI and CNOT1M cDNA was cloned into MCS 1 with the 6His-tag cloned in frame with CNOT1M (Fig. 4.3); in the second step, this vector was linearised with BglII and XhoI (isoschizomers of BamHI and SalI, respectively) and the CNOT9 cDNA was cloned into MCS 2.

### *4.4.2 CNOT9 module expression and purification*

A single colony of *E. coli* BL21 DE3 containing vector P03 was used to inoculate a 2 mL LB culture with appropriate antibiotic for 4-6 h. Twenty-five milliliters of LB were added to this culture and bacteria were incubated for 16 h at 37°C. Then, 1 L of LB was added and cultures were grown at 37°C until the $OD_{600}$ was between 0.7-0.8. IPTG (final concentration of 0.2 mM) was added to induce protein expression. Cultures were incubated for 16 h

at 18°C and cells were harvested by centrifugation. Cells were resuspended in 5 mL Buffer HA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol, 5% glycerol and 20 mM imidazole) per litre of culture, cells were lysed by sonication and the soluble lysate was isolated by centrifugation. The soluble lysate was then loaded on a HisTrap FF pre-packed column (GE) equilibrated with Buffer HA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol, 5% glycerol and 20 mM imidazole). After washing the unbound fraction, five elution fractions of 1 mL were collected using Buffer HB (as Buffer HA, but containing 250 mM imidazole). Typically, fraction number two, containing about 85% of the purified complex, was used in subsequent steps. One microgram of sample was analysed by SDS-PAGE (Fig. 4.3C)

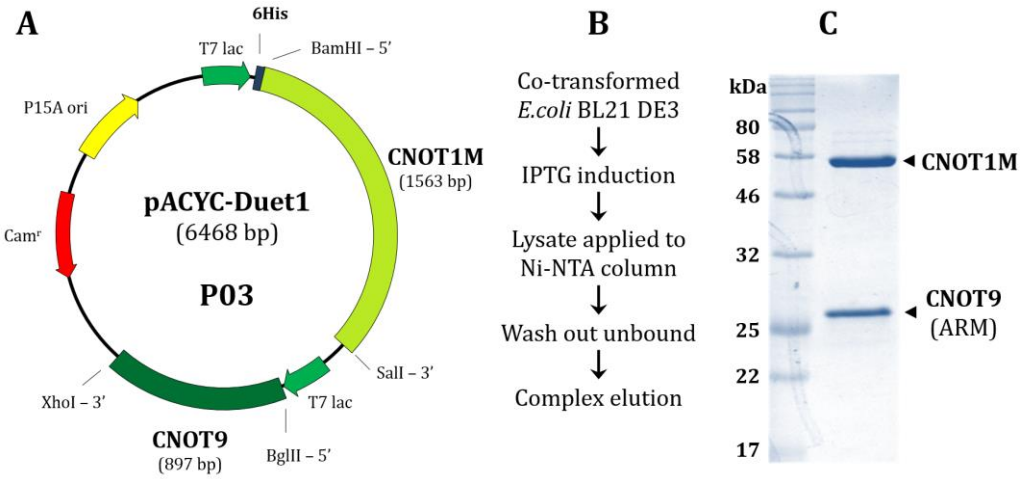



**Figure 4.3** – Vector generated for the CNOT9 module expression in *E. coli* BL21 DE3. (A) map of vector P03, expressing 6His-tagged CNOT1M fragment encompassing MIF4G and DUF3819 domains. (B) Outline of the purification strategy designed for the CNOT9 module and (C) analysis of the purified CNOT9 module using 14% SDS-PAGE. Proteins were visualised by Coomassie staining.

## 4.5 Co-expression and purification of BTG2, Caf1, Ccr4, CNOT1M and CNOT9

### *4.5.1 Vectors for protein co-expression in* E. coli *BL21 DE3*

The first approach to reconstitute the Ccr4-Not core complex was the simultaneous co-expression of the five proteins. To achieve this, cDNAs of the proteins of interest were cloned into vectors containing three different origins of replication. The StrepII-tagged Ccr4 and Caf1 cDNA were expressed using vector P01 described above (Fig. 4.2). BTG2 cDNA was amplified by PCR; then, the PCR product was digested with BamHI and HindIII restriction enzymes and cloned into MCS 1 of vector pCDF-Duet1 P04 linearised with identical enzymes (Fig. 4.4A). In order to express the CNOT9 module, the expression cassette of vector P03, containing both CNOT1M and CNOT9 cDNAs was amplified by PCR and digested with NcoI and AvrII restriction enzymes. The fragment was then cloned into a vector pET-Duet1 P05 linearised with identical restriction enzymes (Fig. 4.4B).

### *4.5.2 Core complex co-expression in* E. coli *BL21 DE3*

A single colony of *E. coli* BL21 DE3 containing vectors P01, P04 and P05 was used to inoculate a 2 mL LB culture with appropriate antibiotic for 4-6 h. Twenty-five milliliters of LB were added to this culture and bacteria were incubated for 16 h at 37°C. Then, 2 L of LB were added and cultures were grown at 37°C until the $OD_{600}$ was between 0.7-0.8. IPTG (final concentration of 0.2 mM) was added to induce protein expression. The cultures were incubated for 16 h at 18°C and cells were harvested by centrifugation.

### *4.5.3 Core complex purification with IMAC*

Cells were resuspended in 5 mL Buffer HA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol, 5% glycerol and 20 mM imidazole) per litre of culture, cells were lysed by sonication and the soluble lysate was

isolated by centrifugation. The soluble lysate was then loaded on a HisTrap FF pre-packed column (GE) equilibrated with Buffer HA. After washing the unbound fraction, five elution fractions of 1 mL were collected using Buffer HB (as Buffer HA, but containing 250 mM imidazole). One microgram of sample was analysed by SDS-PAGE (Fig. 4.4D).

### *4.5.4 Core complex purification with affinity chromatography using the StrepII tag*

Cells were resuspended in 5 mL Buffer SA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol and 5% glycerol) per litre of culture, cells were lysed by sonication and the soluble lysate was isolated by centrifugation. The soluble lysate was then loaded on a gravity-flow column packed with Strep-Tactin beads (IBA) and equilibrated with Buffer SA. After washing the unbound fraction, five elution fractions of 1 mL were collected using Buffer SB (Buffer SA supplemented with 10 mM desthiobiotin). One microgram of sample was analysed by SDS-PAGE (Fig. 4.4D).

### *4.5.5 Analysis of core complex obtained after co-expression and purification*

The core complexes purified by single step affinity purification using the 6His or StrepII-tag were analysed by SDS-PAGE (Fig. 4.4D). Unfortunately, both approaches to reconstitute the Ccr4-Not core complex were not successful. The affinity purification, using the StrepII-tag fused with Ccr4, resulted in non-specific protein binding to the gravity column (Fig. 4.4D). On the other hand, the IMAC purification, targeting the 6His-tag fused with CNOT1M, resulted in a purified CNOT1M, but none of the other components of the two modules are clearly identifiable (Fig. 4.4D).

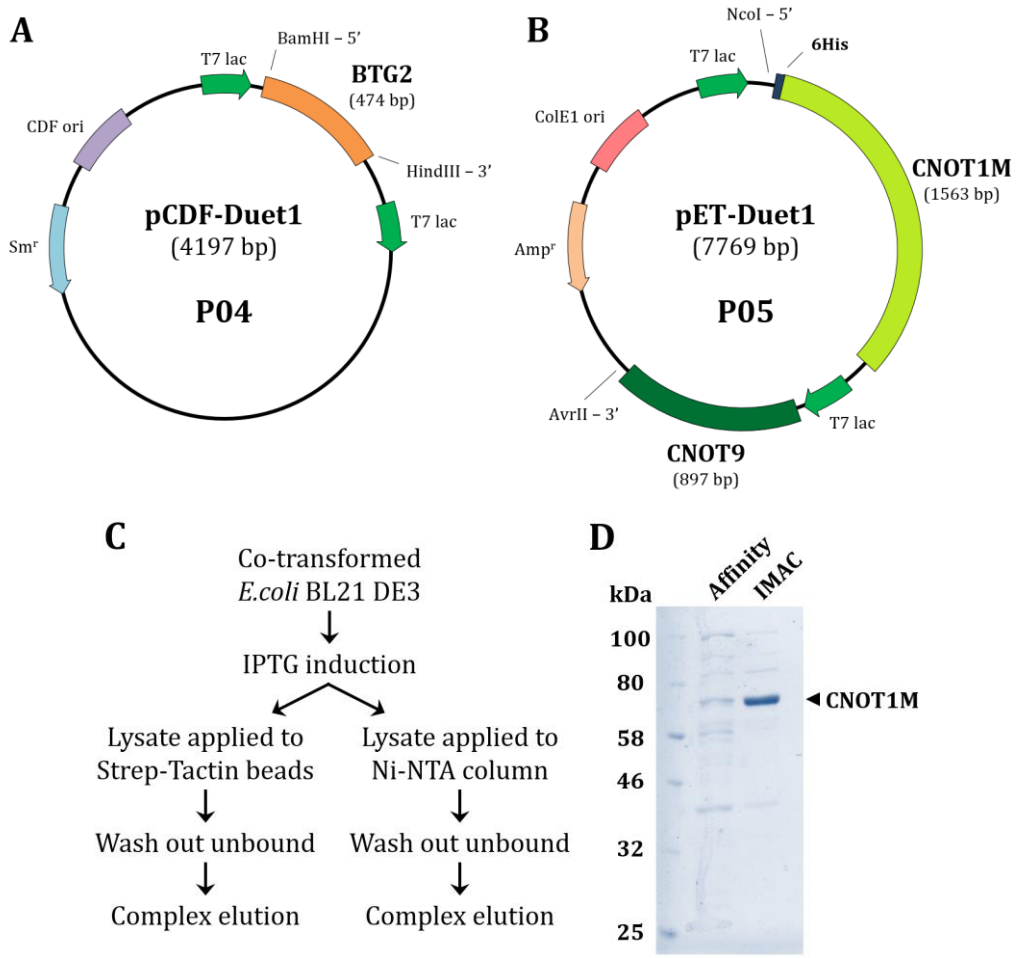



**Figure 4.4** – Vectors generated for the co- expression of the Ccr4-Not core complex in *E. coli* BL21 DE3. (A) map of vector P04, expressing non-tagged BTG2. (B) Map of vector P05, co-expressing 6His-tagged CNOT1M and CNOT9. These two vectors were used in conjuction with vector P01 to co-express five proteins. (C) Outline of the purification strategy designed for the Ccr4-Not core complex. (D) Analysis of the purified nuclease module using 12% SDS-PAGE. Proteins were visualised by Coomassie staining.

## 4.6 Assembly of the core complex by reconstitution using purified nuclease and CNOT9 modules

Since purification of the core complex was unsuccessful following co-expression of the five proteins, a different strategy was evaluated. In this approach, the nuclease and CNOT9 modules were purified separately as described above (Fig. 4.2 and 4.3). Subsequently, the purified modules were mixed and isolated using three different methods: (i) size-exclusion chromatography; (ii) affinity purification based on the 6His-tag fused to CNOT1M; and (iii) affinity purification based on the StrepII-tag fused to Ccr4 (Fig. 4.5A).

### *4.6.1 Core complex isolation by size-exclusion chromatography*

First, the nuclease and CNOT9 modules were mixed in a 1:1 ratio and incubated for 16 h at 4°C to allow binding. After protein re-concentration, approximately 1 mL of sample was applied to a column for size-exclusion chromatography (GE Sephacryl S-100 HR) equilibrated with Buffer SA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol and 5% glycerol) and 2 mL fractions were collected. Typically, 4-5 fractions were pooled and re-concentrated. Five micrograms of sample were analysed by SDS-PAGE (Fig. 4.5B). This approach yielded two purification peaks; the first one is a mixture of the CNOT9 module, more abundant, and the core complex (Fig 4.5B). The second peak, less significant, represents the core complex, with a correct stoichiometry. However, the yield was insufficient for down-stream applications, despite a re-concentration step (Fig. 4.5B).

### *4.6.2 Core complex purification with IMAC*

Next, we evaluated the isolation of a stoichiometric core complex using affinity purification based on the 6His-tag. An excess of the nuclease module over the CNOT9 module was used (1.2:1). The mixture was incubated for 16 h at 4°C to allow the binding between Caf1 and the MIF4G domain of

CNOT1M. After incubation, the buffer was exchanged with PD-10 columns (GE) to Buffer SA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol and 5% glycerol) in order to remove imidazole. After buffer exchange, the sample was loaded on a HisTrap FF pre-packed column (GE) equilibrated with Buffer HA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol, 5% glycerol and 20 mM imidazole). After washing the unbound fraction, five elution fractions of 1 mL were collected using Buffer HB (As Buffer HA, but containing 250 mM imidazole). One microgram of sample was analysed by SDS-PAGE. The single protein peak obtained corresponded to a mixture of the CNOT9 module with the subunits of the nuclease module present in sub-stoichiometric amounts (Fig. 4.5C).

### *4.6.3 Core complex purification with StrepII affinity chromatography*

Finally, we evaluated the use of affinity purification with the StrepII-tag fused to Ccr4. The CNOT9 module was incubated with a 1.2:1 excess compared to the nuclease module. Then, buffer was exchanged using a PD-10 column (GE) to Buffer SA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol and 5% glycerol) to remove D-desthiobiotin. After buffer exchange, the sample was loaded on a gravity-flow column packed with Strep-Tactin beads (IBA) and equilibrated with Buffer SA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol and 5% glycerol). After washing the unbound fraction, five elution fractions of 0.5 mL were collected using Buffer SB (Buffer SA supplemented with 10 mM D-desthiobiotin). One microgram of sample was analysed by SDS-PAGE (Fig. 4.5D). This strategy resulted in a core complex displaying all five subunits with the correct stoichiometry. Approximately 1.5 mg of core complex was isolated (1.8 mg/mL). This procedure was also used to isolate a core complex containing inactive nuclease enzymes Caf1 D40A and Ccr4 E240A (Fig. 4.5D).

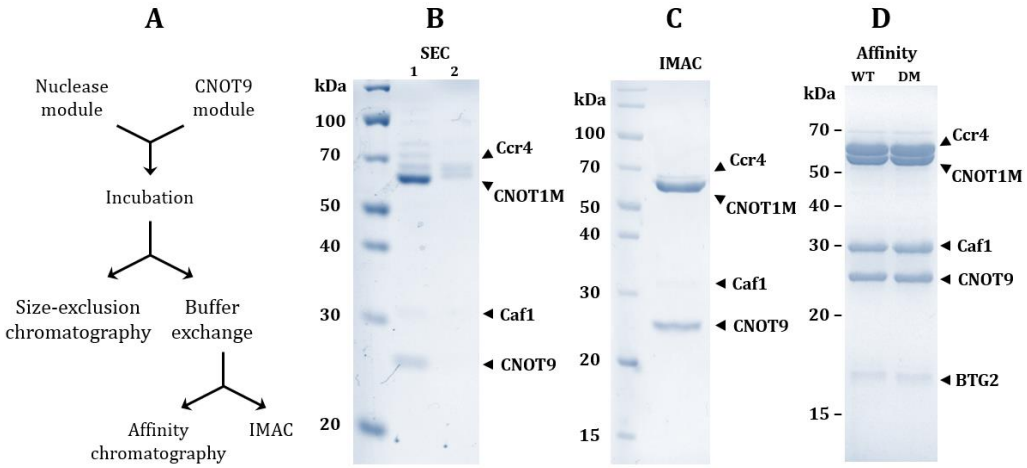



**Figure 4.5** – Outline of the purification strategy designed to isolate the core complex of Ccr4-Not (A). Analysis of the two peaks identified by size-exclusion chromatography using 12% SDS-PAGE. Proteins were visualised by Coomassie staining (B). Analysis of the elution fraction from IMAC using 14% SDS-PAGE. Proteins were visualised by Coomassie staining (C). Analysis of the core complexes (wild type and double mutant) isolated with affinity purification, using 14% SDS-PAGE. Proteins were visualised by Coomassie staining (D).

## 4.8 Discussion

The Ccr4-Not complex is a higly conserved machine involved in mRNA deadenylation. It contains at least eight subunits (Winkler & Balacco 2013). The human complex is composed of four 'canonical' modules: the N-terminal module, the nuclease module, the CNOT9 module and the NOT module. The Ccr4-Not complex contains two active deadenylase enzymes (Ccr4 and Caf1) that are associated with the poly(A) tail removal. The complex is targeted to mRNAs with several mechanisms, both generic and specific. The generic recruitment is based on the interaction between Poly(A) Binding Proteins (PABPC1) and BTG2, a co-factor of Caf1.

Significant progress has been made regarding the structure of the Ccr4-Not complex. Several structures illuminate the structure and activity of Caf1 (Horiuchi et al. 2009), the Ccr4 catalytic domain (Wang et al. 2010), the *S. cerevisiae* nuclease module (Basquin et al. 2012), the assembly of the human Caf1-CNOT1 MIF4G (Petit et al. 2012) and CNOT1 DUF3819-CNOT9 modules (Chen et al. 2014). Further insight has been obtained by biochemical analysis of the intact *S. pombe* Ccr4-Not complex (Ukleja et al. 2016). However, the analysis of partially reconstituted *Drosophila* and human nuclease modules has also led to unexpected insight (Maryati et al. 2015; Niinuma et al. 2016; Stupfler et al. 2016). The catalytic activity of the human module is abolished by single mutations in either active sites of Caf1 or Ccr4, suggesting an allosteric mechanism. Therefore, here it was attempted to reconstitute a core complex, based on the previously reported reconstitution of the human BTG2-Caf1-Ccr4 nuclease module and on the purification of the CNOT9 module with a CNOT1M fragment including MIF4G and DUF3819. A number of biochemical studies characterised the catalytic activity of the deadenylase enzymes of the Ccr4-Not complex *in vitro* (Maryati et al. 2015; Stupfler et al. 2016).

In this chapter, different strategies designed to assemble the human Ccr4-Not core complex were evaluated. First, five proteins were co-expressed, as this could improve complex formation, and two purification strategies were carried out to isolate the assembled core complex (Fig. 4.4D). However, neither resulted in a complex with the correct stoichiometry or sufficient yields. The second reconstitution approach consisted in three subsequent steps; first, the nuclease and CNOT9 modules were purified separately (Fig. 4.2 and 4.3), as the yields were satisfactory and complexes were stable. Second, according to the downstream steps, different ratios of the modules were incubated to allow them to bind. Third, different purification approaches were optimised to isolate the core complex. As showed above, only the affinity purification, using the StrepII-tag fused with Ccr4, yielded a correctly assembled core complex with sufficient concentration and amounts for *in vitro* assays. This reconstituted Ccr4-Not core complex is a step beyond the nuclease module platform for testing the deadenylase activity of Ccr4 and Caf1, which are tethered to their natural scaffold CNOT1, also containing the closest next module (CNOT9). Moreover, a more complete environment could allow additional conformational changes that are not triggered in presence of the single catalytic enzymes. In addition, this system allows the characterisation of the interaction between BTG2, a co-factor of the Ccr4-Not complex and more specifically of Caf1, and PABPC1, which regulates a recurrent deadenylation pathway in human cells (Ezzeddine et al. 2007; Mauxion et al. 2009; Stupfler et al. 2016).

Future studies could aim at including additional proteins to this system, such as tristetraprolin (TTP), to characterise the activity of the Ccr4-Not complex when it targets mRNA molecules with AU-rich elements in their 3'-UTR (Sandler et al. 2011; Fabian et al. 2013). The contribution of CNOT9 could be further defined by investigating the role of human Roquin, which

recruits Ccr4-Not onto mRNAs displaying 3'-UTR stem-loop structures (Sgromo, Raisch, Bawankar, et al. 2017).

# Chapter 5

## Biochemical activity of a human Ccr4-Not core complex

# Chapter 5. Biochemical activity of a human Ccr4-Not core complex

## 5.1 Introduction

The best characterised function of the Ccr4-Not complex is targeting and shortening poly(A) tails of mRNAs that are destined to degradation. The complex contains two different deadenylase subunits: Caf1 and Ccr4, that have been well characterised structurally and biochemically in yeast (Tucker et al. 2001; Stowell et al. 2016), mouse (Draper et al. 1995; Viswanathan et al. 2004) and human (Yamashita et al. 2005; Chen et al. 2009; Mauxion et al. 2009; Horiuchi et al. 2009; Wang et al. 2010).

The apparent redundancy of two enzymes sharing the same catalytic activity, although belonging to different protein families, remains unexplained, but indicates allosteric regulation in the nuclease module (Aslam 2010; Winkler & Balacco 2013; Maryati et al. 2015). Ultimately, while Caf1 and Ccr4 have a clear function within the Ccr4-Not complex but do not seem to be intrinsically selective towards specific mRNAs, the activity of the complex may be influenced by other subunits, such as BTG2, CNOT1 and CNOT9. The architecture of the nuclease module of the Ccr4-Not complex has been characterised in yeast (Basquin et al. 2012) but further studies are required to shed light on the role of the other subunits of the complex, such as CNOT1 (Ito, Takahashi, et al. 2011) and CNOT9 (Mathys et al. 2014; Chen et al. 2014), that appear to have a fundamental role in the Ccr4-Not complex recruitment process.

In this chapter, we designed experiments aiming at understanding whether some of the non-catalytic subunits of the human Ccr4-Not contribute towards its deadenylation activity. To define the contribution of single subunits or modules in this catalytic activity, we compared the Caf1-containing complexes that were introduced in the previous chapter: Caf1

and the Caf1-BTG2 dimer; the nuclease module of the Ccr4-Not complex, containing Caf1, BTG2 and Ccr4; the Ccr4-Not core complex, including the nuclease module and the CNOT9 module, composed of a central region of CNOT1 and CNOT9.

Several assays were designed to directly compare the activities of these complexes under different conditions. To this end, reactions with equimolar quantities of deadenylation enzymes and complexes were set up. Three poly(A)-containing RNA substrates were used with different poly(A) lengths: $(A)_9$, $(A)_{20}$ and $(A)_{50}$. In addition, the $(A)_{20}$ and $(A)_{50}$ substrates were coated with the full length poly(A)-binding protein PABPC1, or a fragment of PABPC1 containing RRM1 and RRM2 (amino acids 1-190). Both full length PABPC1 and the RRM1-RRM2 fragment can simultaneously bind poly(A) tails and BTG2 (Stupfler et al. 2016). Finally, the ribonucleases were compared in terms of selectivity against poly(U), poly(C) and poly(G) substrates.

## 5.2 Deadenylase activity degrading short poly(A)$_9$-tailed mRNAs

### *5.2.1 A fluorescence-based assay to characterise deadenylation activity*

To determine the effect of Ccr4 and non-catalytic subunits on the activity of Caf1, the activities of equimolar amounts of purified Caf1 and Caf1-containing protein complexes were compared (Fig. 5.1).

In the first instance, a fluorescence-based assay was used (Maryati et al. 2014). Briefly, a fluorescein-labelled RNA molecule containing a poly(A)$_9$ tail, was added to a mixture of deadenylase enzyme and buffer containing $Mg^{2+}$. Reactions were incubated for up to 60 minutes, aliquots were taken at specific time points and a mixture of TAMRA-labelled DNA probe and SDS was added. The DNA probe is complementary to the sequence of the RNA substrate, therefore intact RNA substrate molecules are annealed. This brings the TAMRA dye into close proximity of the fluorescein, quenching its fluorescence (Fig. 5.2A). Conversely, when the poly(A) tail is degraded, the DNA probe is not able to bind the RNA substrate, resulting in a high fluorescence (Fig. 5.2A).

### *5.2.2 Both the anti-proliferative protein BTG2 and the CNOT1-CNOT9 module enhance deadenylation*

The *in vitro* assay described above was used to compare the activities of the various complexes described in Chapter 4 (Fig. 5.1). Specifically, aliquots of the reactions were stopped with DNA probe/SDS mix at given times (0, 5, 10, 20, 30, 40 and 60 minutes) and fluorescence, a direct measure of the deadenylase activity, was detected to directly compare the deadenylation rates (Fig. 5.2A).

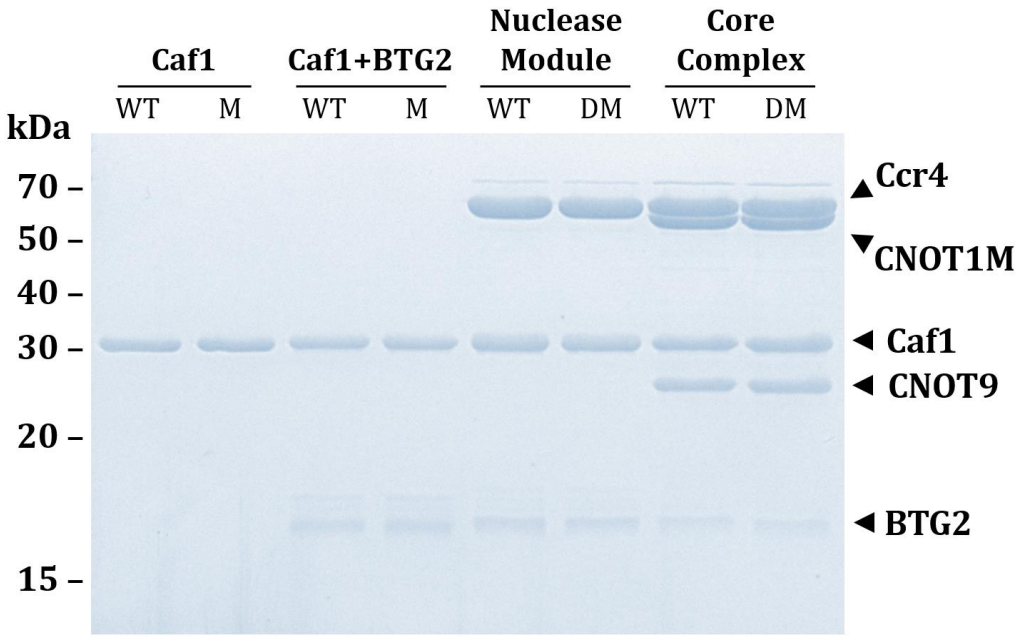



**Figure 5.1** – Analysis of the purified deadenylase Caf1; Caf1 in complex with BTG2; the nuclease module (Caf1, Ccr4 and BTG2); the Ccr4-Not core complex (Caf1, Ccr4, BTG2, CNOT1M and CNOT9). Protein samples were analysed by 14% SDS-PAGE and visualised by Coomassie staining. Identical molar concentrations were loaded for all protein complexes. M, Caf1 D40A mutant; DM, Caf1 D40A/Ccr4 E240A double mutant.

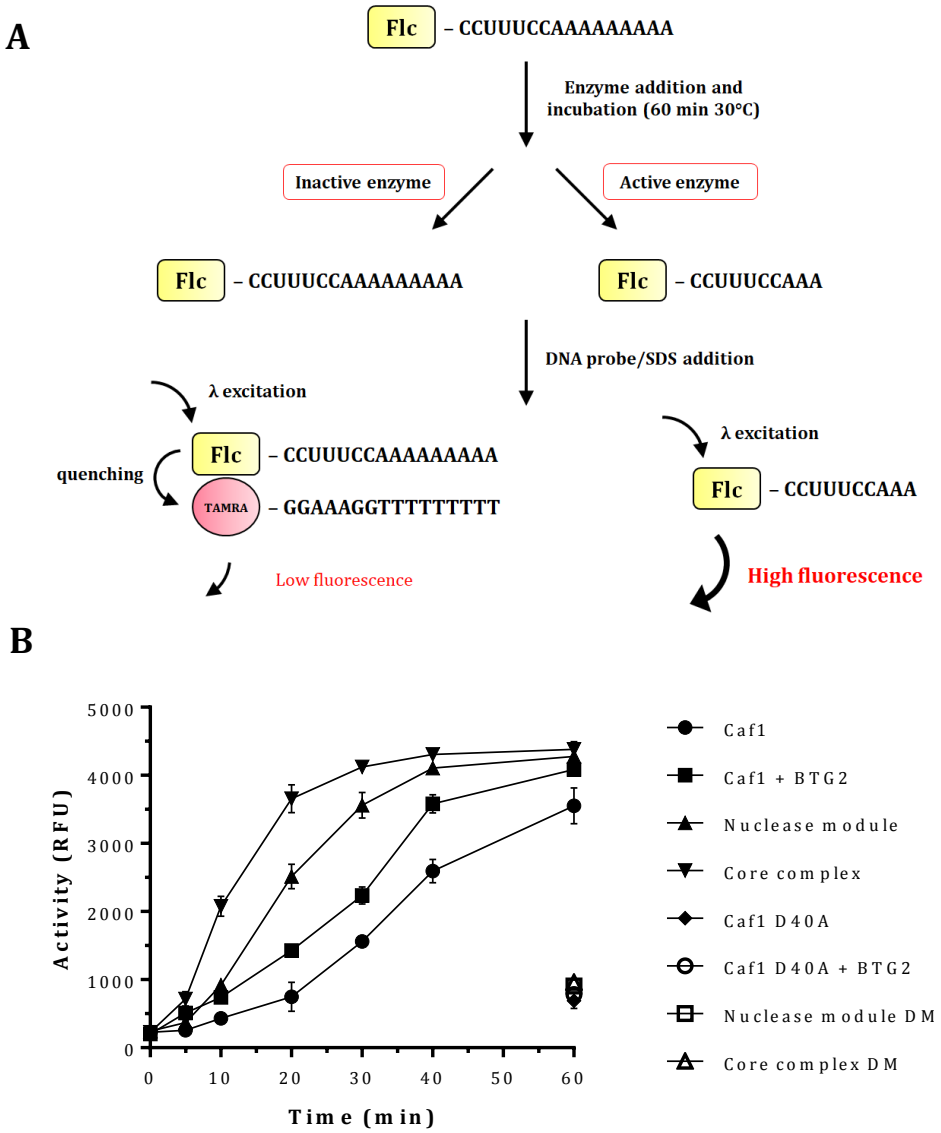



**Figure 5.2** – Fluorescence-based deadenylation assays. (A) Schematic of the fluorescence-based assay. RNA substrate is added to either active or inactive enzymes. Upon addition of a DNA probe/SDS mix, the intact RNA molecules anneal and TAMRA quenches the signal. DNA probe cannot bind degraded RNA substrates that, when excited with light at 485 nm wavelength, produce a high fluorescent signal. (B) Comparison of the deadenylation activities of the enzymes/complexes described above. The RNA substrate was a 16-mer oligo with a poly(A)$_9$ tail. Reactions (50 µl) were setup and 5 µL aliquots taken at given time points and immediately mixed with 5 µL of DNA probe/SDS mix. Reactions contained RNA substrate (1.0 µM) and enzyme (0.2 µM).

In this *in vitro* fluorescence-based assay, the catalytic activity of the deadenylase enzyme was tested with a substrate displaying a short poly(A) tail. Under equimolar conditions, BTG2 presence contributes to an increased deadenylation activity, when compared to Caf1 alone (Fig. 5.2B), which is consistent to results reported by Maryati et al. (Maryati et al. 2015). Similarly, the presence of an additional catalytically active enzyme, Ccr4, in the nuclease module, resulted in accelerated deadenylation rates (Fig 5.2B). Surprisingly, however, the presence of the CNOT1-CNOT9 module produced a considerable increase in the deadenylation activity, which was particularly noticeable within the first 20 minutes (Fig. 5.2B).

## 5.3 Deadenylase activity degrading poly(A)$_{20}$-tailed RNA

Next, the activity of the different deadenylation complexes was compared using a 20-mer composed of a fluorescein fluorophore conjugated with a poly(A)$_{20}$. PAGE analysis in denaturing conditions (50% UREA) was carried out to visualise and characterise the deadenylation patterns with identical time points described above.

### *5.3.1 A PAGE-based assay to characterise deadenylation activity*

In order to detect the ability of the different complexes to degrade the poly(A) tail of substrates, a poly-acrylamide gel electrophoresis-based assay was used (Fig. 5.3). Briefly, a fluorescein-labelled RNA molecule with poly(A)$_{20}$ tail (200 nM) was added to a mixture of deadenylase enzyme (100 nM and 50 mM NaCl) and buffer containing $Mg^{2+}$ (2 mM). Reactions were incubated for up to 60 minutes at 30°C. Aliquots were taken at specific time points and mixed with RNA denaturing dye and subsequently analysed with PAGE (20% poly-acrylamide and 50% UREA).

### *5.3.2 Enzymes and complexes catalytic activity with a poly(A)$_{20}$ substrate*

Under these conditions, the activity of Caf1 was hardly detectable after one hour (Fig. 5.4A). However, a noticeably increased deadenylation activity was observed when Caf1 was in complex with the anti-proliferative factor BTG2 (Fig. 5.4B). As expected, the presence of a second active enzyme (Ccr4) in the nuclease module caused a remarkable acceleration in the activity, with fully-deadenylated products appearing after 30 minutes (Fig. 5.4C). Interestingly, the module appears to slow down after removing several bases (5-10), and after that, poly(A) tail degradation is rapidly completed. Finally, the presence of the CNOT1-CNOT9 module has a significant impact on the deadenylation rates. After 30 minutes, the majority of the poly(A)$_{20}$ tails are fully deadenylated when mixed with the core complex (Fig. 5.4D).

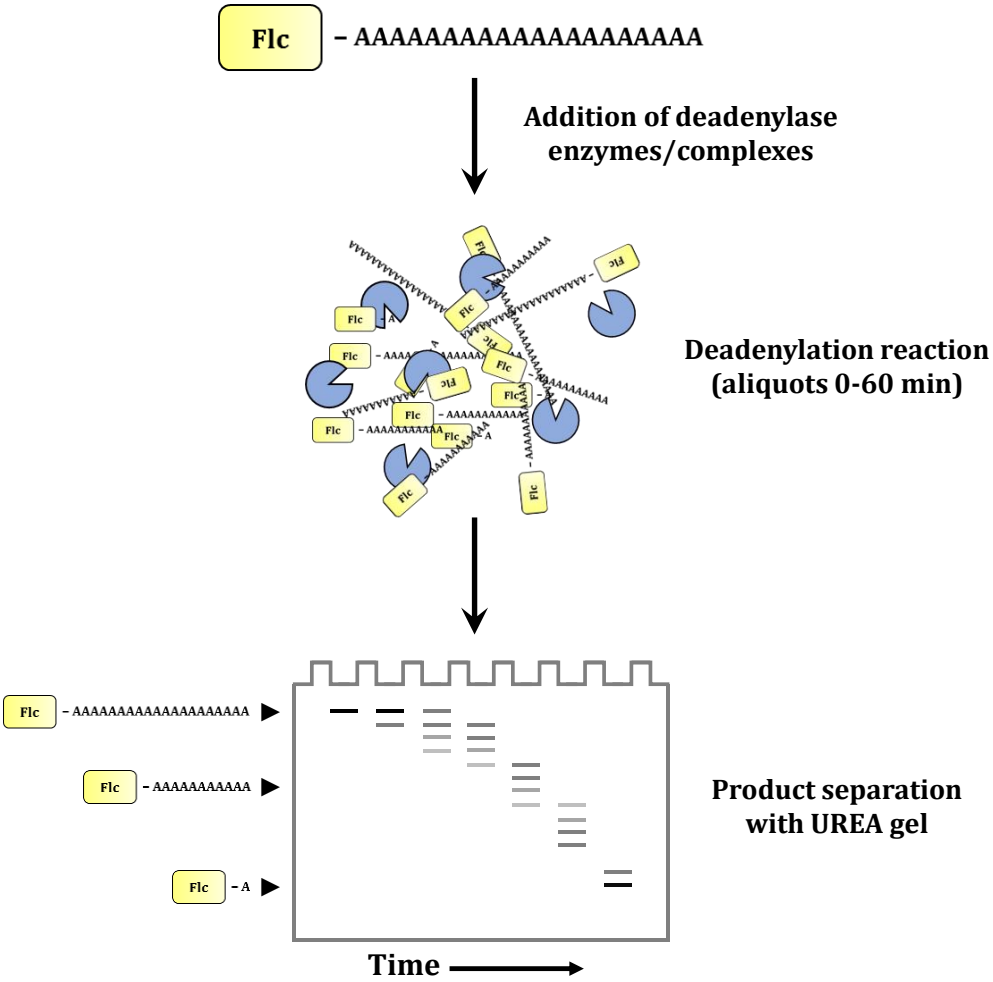



**Figure 5.3** – Outline of the *in vitro* deadenylation reactions designed with the poly(A)-tailed RNA substrate. Reactions (50 µL) were started upon addition of the deadenylase enzymes/complexes (100 nM) to RNA substrate (200 nM). Aliquots (5 µL) were taken at given time points and heated at 85°C for few minutes after addition of RNA loading dye. Samples were analysed by 50% UREA/20% PAGE.

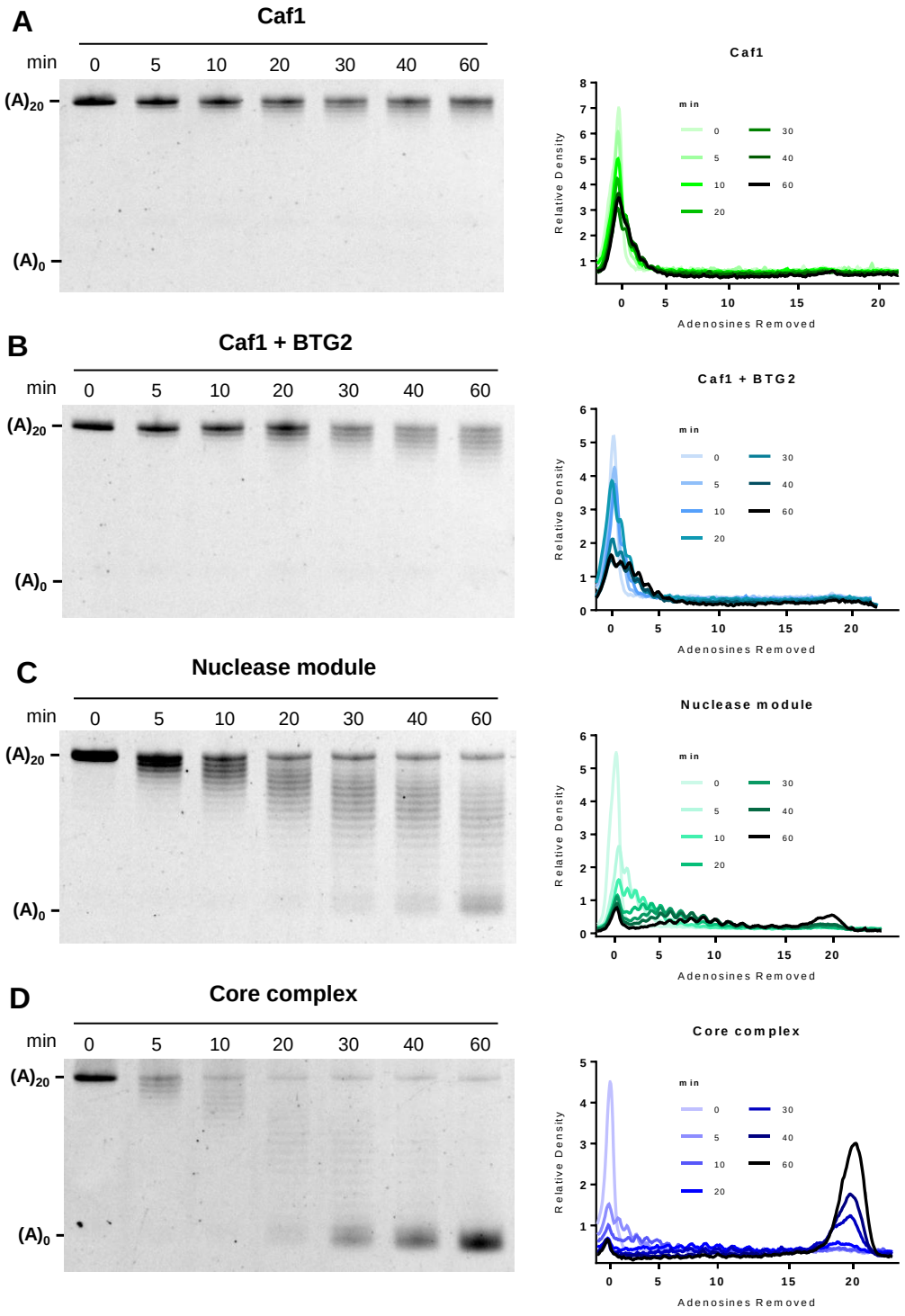



**Figure 5.4** – Deadenylation activities of the Caf1 enzyme and the Caf1-containing complexes. Reactions were started upon addition of the deadenylase enzymes/complexes (100 nM) to RNA substrate (200 nM). Aliquots from the

reactions were collected at specific time points and analysed by 50% UREA/20% PAGE (left panels). Densitometry analysis of the scanned gels is provided (right panels). The number of adenosines removed is estimated based on the densitograms.

## 5.4 Deadenylase activity degrading poly(A)$_{50}$-tailed RNA

The third *in vitro* assay used to compare the activity of the different deadenylation complexes contained a 57-base oligonucleotide composed of a fluorescein fluorophore conjugated to a RNA substrate containing a poly(A)$_{50}$ tail. PAGE analysis in denaturing conditions (50% UREA) was carried out to visualise and characterise the deadenylation patterns with identical time points described above.

### *5.4.1 A PAGE-based assay to characterise deadenylation activity*

In order to detect the ability of the different complexes to degrade the poly(A)$_{50}$ tail of the RNA substrates, a poly-acrylamide gel electrophoresis-based assay was used. Briefly, a fluorescein-labelled RNA molecule with poly(A)$_{50}$ tail (200 nM) was added to a mixture of deadenylase enzyme (100 nM and 50 mM NaCl) and buffer containing $Mg^{2+}$ (2 mM). Reactions were incubated for up to 60 minutes at 30°C. Aliquots were taken at specific time points and mixed with RNA denaturing dye and subsequently analysed with PAGE (16% poly-acrylamide and 50% UREA).

### *5.4.2 Enzymes and complexes catalytic activity with a poly(A)$_{50}$ substrate*

Under these conditions, the activity of Caf1 was below the detectable limit after one hour (Fig. 5.5A). BTG2 marginally enhanced the acvity of Caf1 (Fig. 5.5B). The nuclease showed faster deadenylation activity compared to the Caf1-BTG2 complex (Fig. 5.5C). Finally, the core complex displayed the highest deadenylation activity. After 40 minutes, fully deadenylated products were observed (Fig. 5.5D).

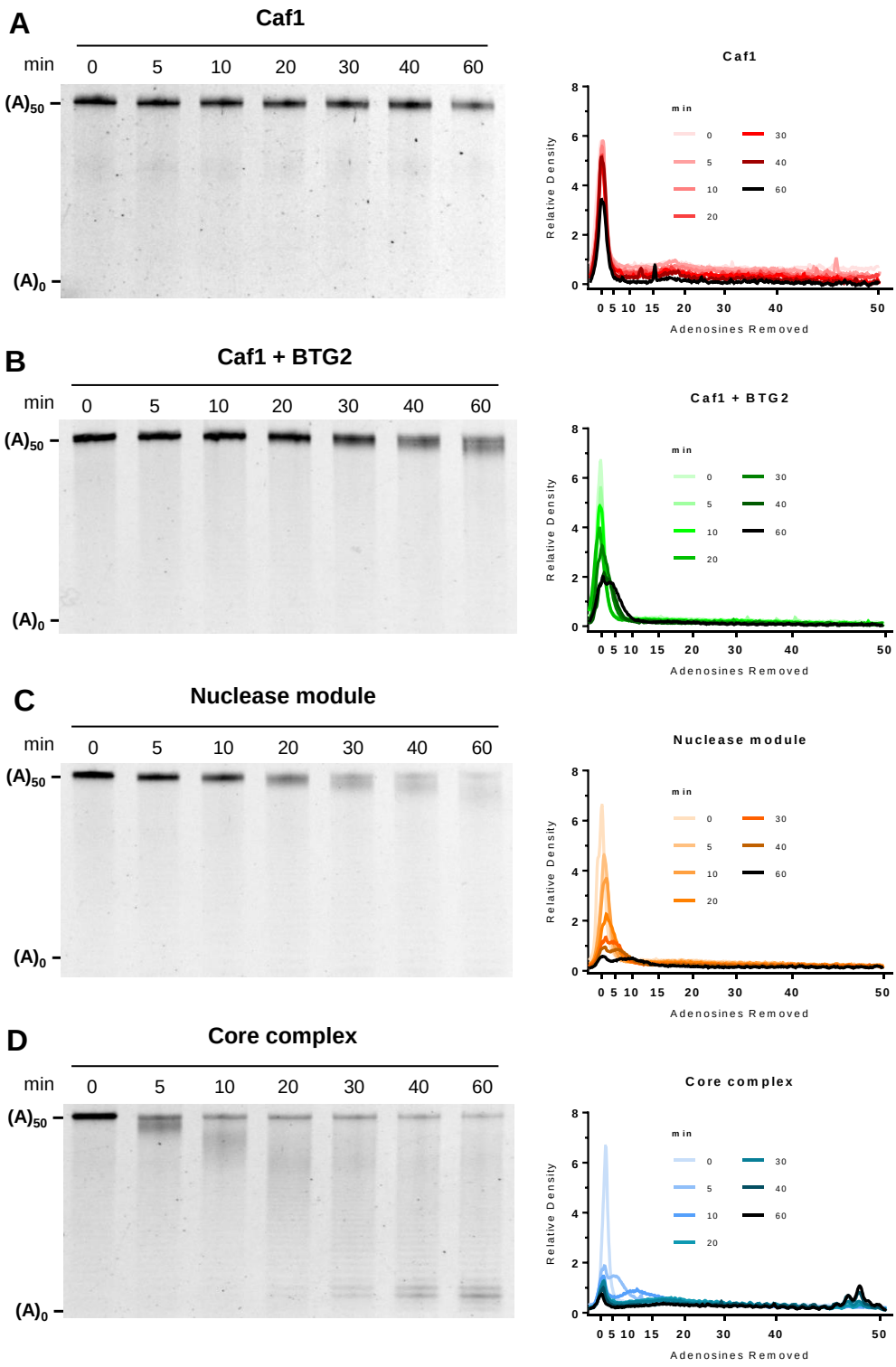



**Figure 5.5** –Deadenylation activities of the Caf1 enzyme and the Caf1-containing complexes. Reactions were started upon addition of the deadenylase enzymes/complexes (100 nM) to RNA substrate (200 nM). Aliquots from the

reactions were collected at specific time points and analysed by 50% UREA/16% PAGE (left panels). Densitometry analysis of the scanned gels is provided (right panels). The number of adenosines removed is estimated based on the densitograms.

## 5.5 Expression and purification of the human Poly(A)-Binding Protein (PABPC1)

PABPC1 is a well-characterised protein that binds to the mRNA poly(A) tails. Specifically, it displays four RNA-recognition motifs (RRM 1-4) that are recognising poly(A)$_{10}$ stretches. To determine the ability of Caf1 and Caf1-containing complexes to degrade PABPC1-poly(A) protein/RNA substrates, full length and two PABPC1 fragments were expressed and purified using *E. coli* BL21 DE3.

### *5.5.1 Vectors for protein expression in* E. coli *BL21 DE3*

To express PABPC1, the full length CDS was cloned from total mRNA extracted from HEK293 cells using reverse transcriptase-PCR. The cDNA sequence was amplified using primers containing a BamHI site at the 5' end and a SalI site at the 3' end. The amplification product was digested with BamHI and SalI and cloned into vector pQE80L linearised with identical restriction enzymes. Vectors expressing the PABPC1 N-terminal fragments were generated using a similar strategy using alternative 3' primers. The presence of the cloned cDNAs in vectors P11 (PABPC1 1-99), P12 (PABPC1 1-190) and P13 (full length PABPC1) was confirmed by Sanger sequencing (Fig. 5.6A).

### *5.5.2 PABPC1 (RRM1, RRM1-2 and FL) expression and purification*

A single colony of *E. coli* BL21 DE3 transformed with either vector P11, P12 or P13 was used to inoculate a 2 mL LB culture with appropriate antibiotic for 4-6 h. Twenty-five milliliters of LB were added to this culture and bacteria were incubated for 16 h at 37°C. Then, 1 L of LB was added and cultures were grown at 37°C until the $OD_{600}$ was between 0.7-0.8. IPTG (final concentration of 0.2 mM) was added to induce protein expression. Cultures were incubated for 16 h at 18°C and cells were harvested by centrifugation. Cells were resuspended in 5 mL Buffer HA (250 mM NaCl, 20

mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol, 5% glycerol and 20 mM imidazole) and lysed by sonication. The soluble lysate was isolated by centrifugation and then loaded on a HisTrap FF pre-packed column (GE) equilibrated with Buffer HA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol, 5% glycerol and 20 mM imidazole). After washing the unbound fraction, three elution fractions of 1 mL were collected using Buffer HB (as Buffer HA, but containing 250 mM imidazole). PABPC1 proteins were further purified by size-exclusion chromatography (Fig. 5.6B). After fractions pooling and reconcentration, approximately 1 mL of sample was applied to a SEC column (GE Sephacryl S-100 HR) and 2 mL fractions were collected. Typically, 4-5 fractions were pooled and re-concentrated. Five micromoles of sample were analysed by SDS-PAGE (Fig. 5.6C).

### *5.5.3 poly(A) tails coating with PABPC1 purified proteins*

Electrophoretic mobility shift assays (EMSA) were set up to confirm the binding of PABPC1 to poly(A) tails of RNA molecules and to identify the lowest excess of PABPC1 required to shift the majority of the RNA substrate in assay conditions (Fig. 5.7). A 6-fold excess of PABPC1 (FL) was sufficient to bind all RNA molecules, both the poly(A)$_{20}$ substrate (Fig. 5.7A) and the poly(A)$_{50}$ (Fig. 5.7B). A 8-fold excess was necessary for reaching an equilibrium where at least one PABPC1 (1-190) was bound to the poly(A)$_{20}$ substrate (Fig. 5.7C) and the poly(A)$_{50}$ substrate (Fig. 5.7D). Upon incubation with PABPC1 (1-99), RNA substrates were rapidly degraded, probably due to a nuclease contamination (data not shown). Therefore, this protein sample was not used in subsequent analysis.

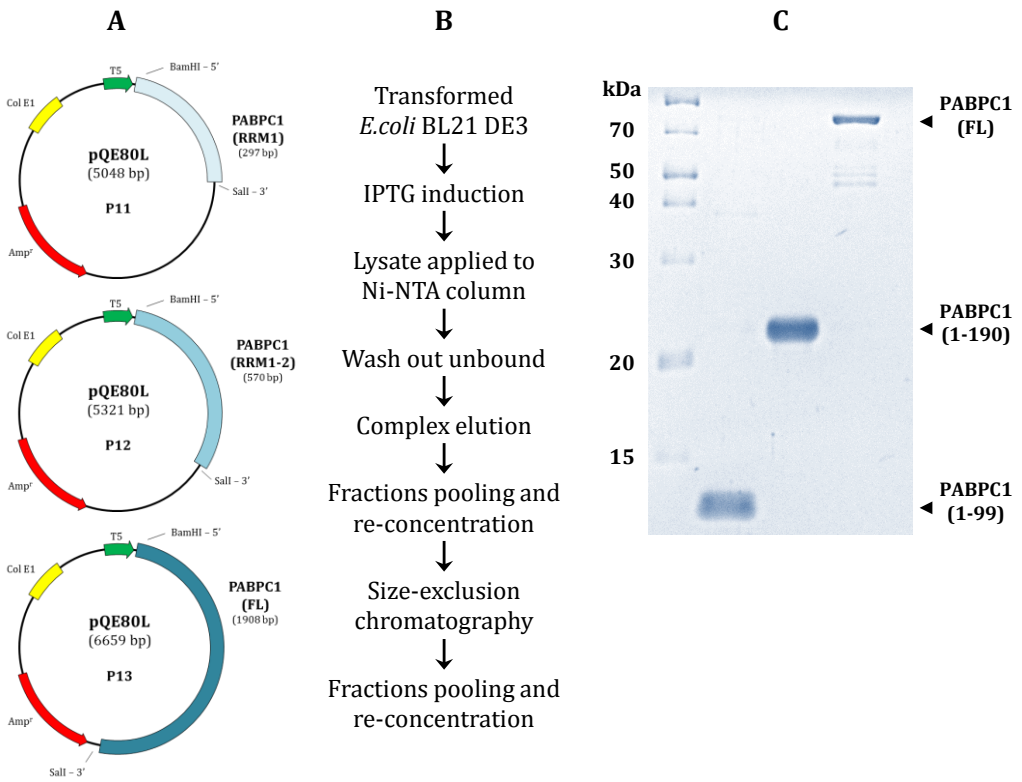



**Figure 5.6** – PABPC1 expression in *E. coli* BL21 DE3. (A) map of vectors P11, P12 and P13, expressing 6His-tagged PABPC1 (1-99, 1-190 and FL). (B) Outline of the purification strategy designed for PABPC1 expression and purification. (C) Analysis of the purified proteins using 14% SDS-PAGE. Proteins were visualised by Coomassie staining.

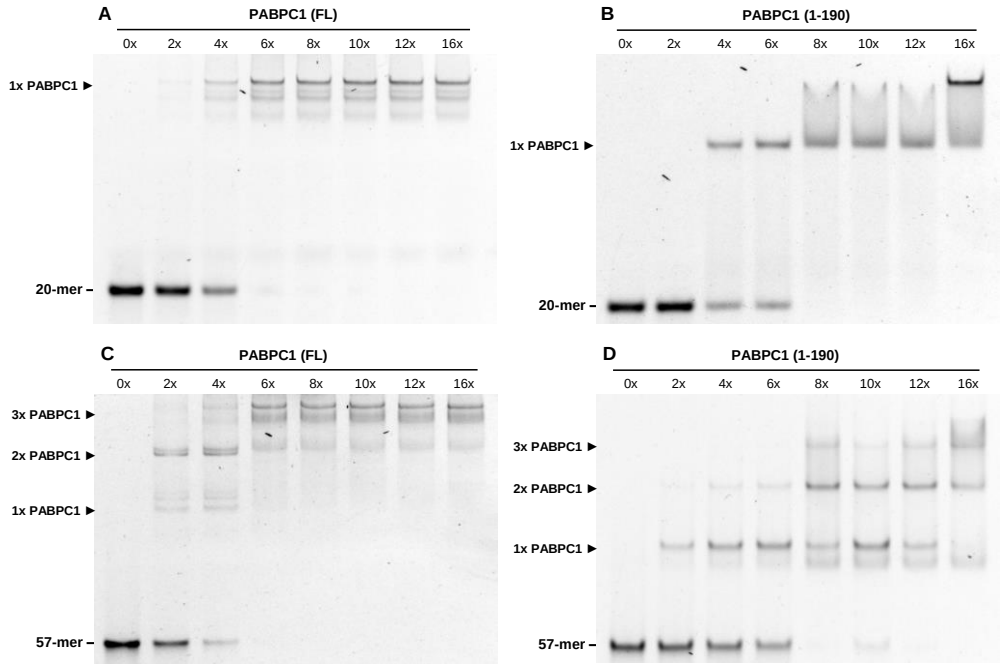



**Figure 5.7** – Electrophoretic mobility shift assay of PABPC1-poly(A) protein:RNA complexes. (A) Analysis of PABPC1 (full length) binding to poly(A)$_{20}$. (B) Analysis of PABPC1 (RRM1-RRM2; amino acids 1-190) binding to poly(A)$_{20}$. (C) Analysis of PABPC1 (full length) binding to poly(A)$_{50}$. (D) Analysis of PABPC1 (RRM1+RRM2; amino acids 1-190) binding to poly(A)$_{50}$. Several molar excesses of PABPC1 (0-16x) were incubated with RNA substrate (200 nM) for 20 min at room temperatures. Then, non-denaturing loading dye was added and reactions were analysed by 8% PAGE. A 6-fold excess of PABPC1 was necessary to shift both substrate types, whereas a 8-fold excess of PABPC1 1-190 was needed. These conditions were utilised in subsequent assays.

## 5.6 Deadenylase activity of Caf1 and Caf1-complexes using poly(A)$_{20}$-PABPC1 substrates

In this assay, the RNA substrate was a fluorescein fluorophore conjugated with a poly(A)$_{20}$ and coated with either full length PABPC1, or an N-terminal fragment containing RRM1 and RRM2. PAGE analysis in denaturing conditions (50% UREA) was carried out to visualise and characterise the deadenylation patterns.

### *5.6.1 A PAGE-based assay to characterise deadenylation activity*

This assay had a similar design to the one described above (Fig. 5.3). However, prior addition of the enzymes, RNA was incubated with an excess of PABPC1 (1-190) or of full length PABPC1. Then, a mixture of deadenylase enzyme (100 nM and 50 mM NaCl) and buffer containing $Mg^{2+}$ (2 mM) was added and reactions were incubated for up to 60 minutes at 30°C. Aliquots were taken at specific time points and mixed with RNA denaturing dye and subsequently analysed by PAGE (20% poly-acrylamide and 50% UREA).

### *5.6.2 Activity of enzymes and complexes using a poly(A)$_{20}$ substrate*

In the presence of RNA coated with PABPC1 (1-190 or FL), the activity of Caf1 was hardly detectable after one hour (Fig. 5.8A and 5.9A). However, PABPC1 (1-190 or FL) stimulated the activity of the Caf1-BTG2 complex (Fig 5.8B and 5.9B), when compared with uncoated poly(A)$_{20}$ (Fig. 5.4B). In a similar fashion, the nuclease module showed enhanced deadenylation activity with RNA bound to PABPC1 (Fig. 5.8C and 5.9C). The activity of the module seems to slow down after having removed few bases. The core complex displayed a similar degradation pattern to the one generated by the nuclease module, but substantially faster rates in presence of PABPC1 1-190 (Fig. 5.8D), but not PABPC1 FL (Fig. 5.9D). However, only a minor increase in the rate (PABPC1 1-190) or a decrease (PABPC1 FL) can be detected when compared with uncoated RNA reactions (Fig. 5.4D).

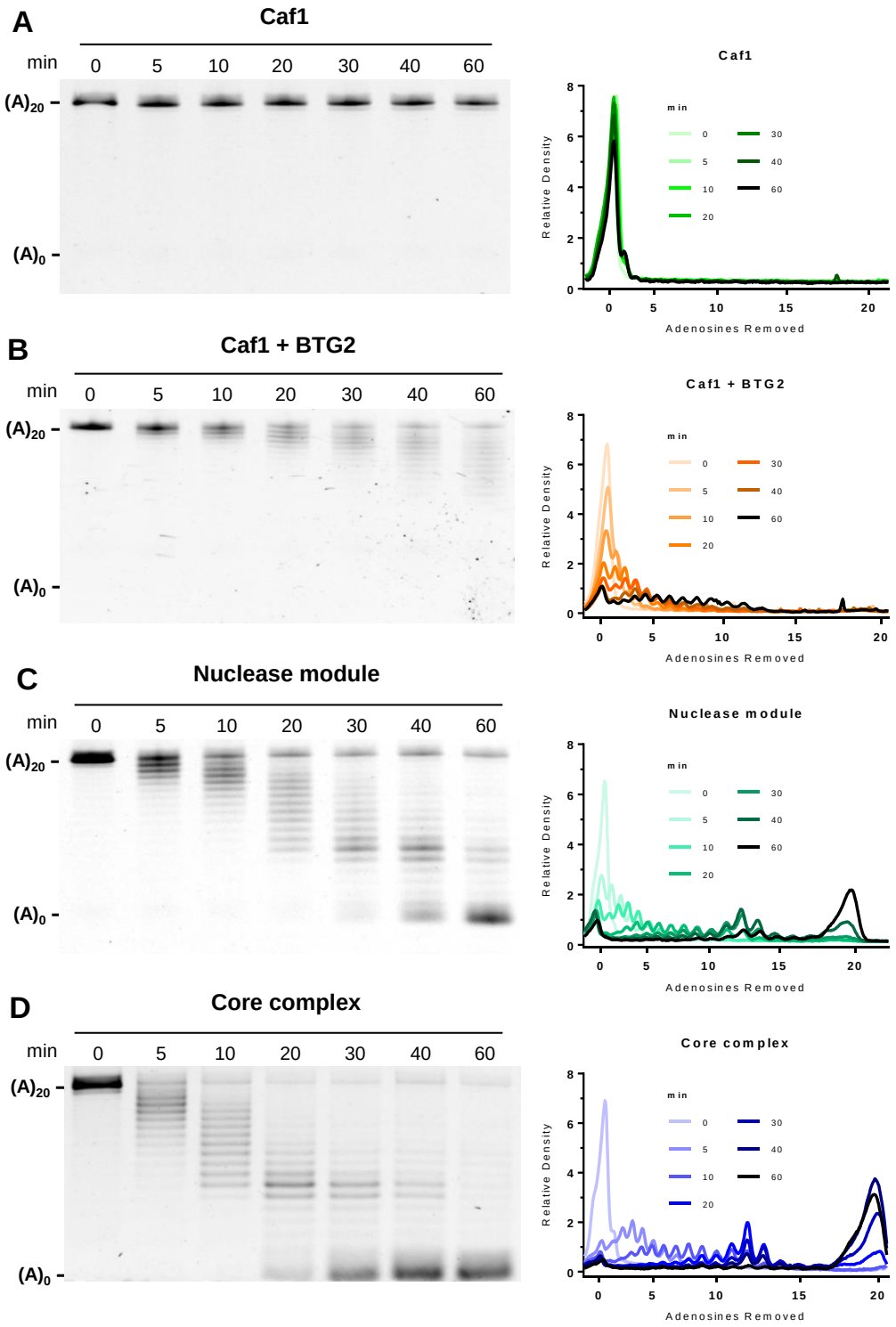



**Figure 5.8** – Comparison of the deadenylation activities of the Caf1 enzyme and the Caf1-containing complexes in presence of a 8-fold excess of PABPC1 (1-190). Aliquots from the reactions were collected at specific time points and analysed

by 50% UREA/20% PAGE (left panels). Densitometry analysis of the scanned gels is provided (right panels). The number of adenosines removed is estimated based on the densitograms.

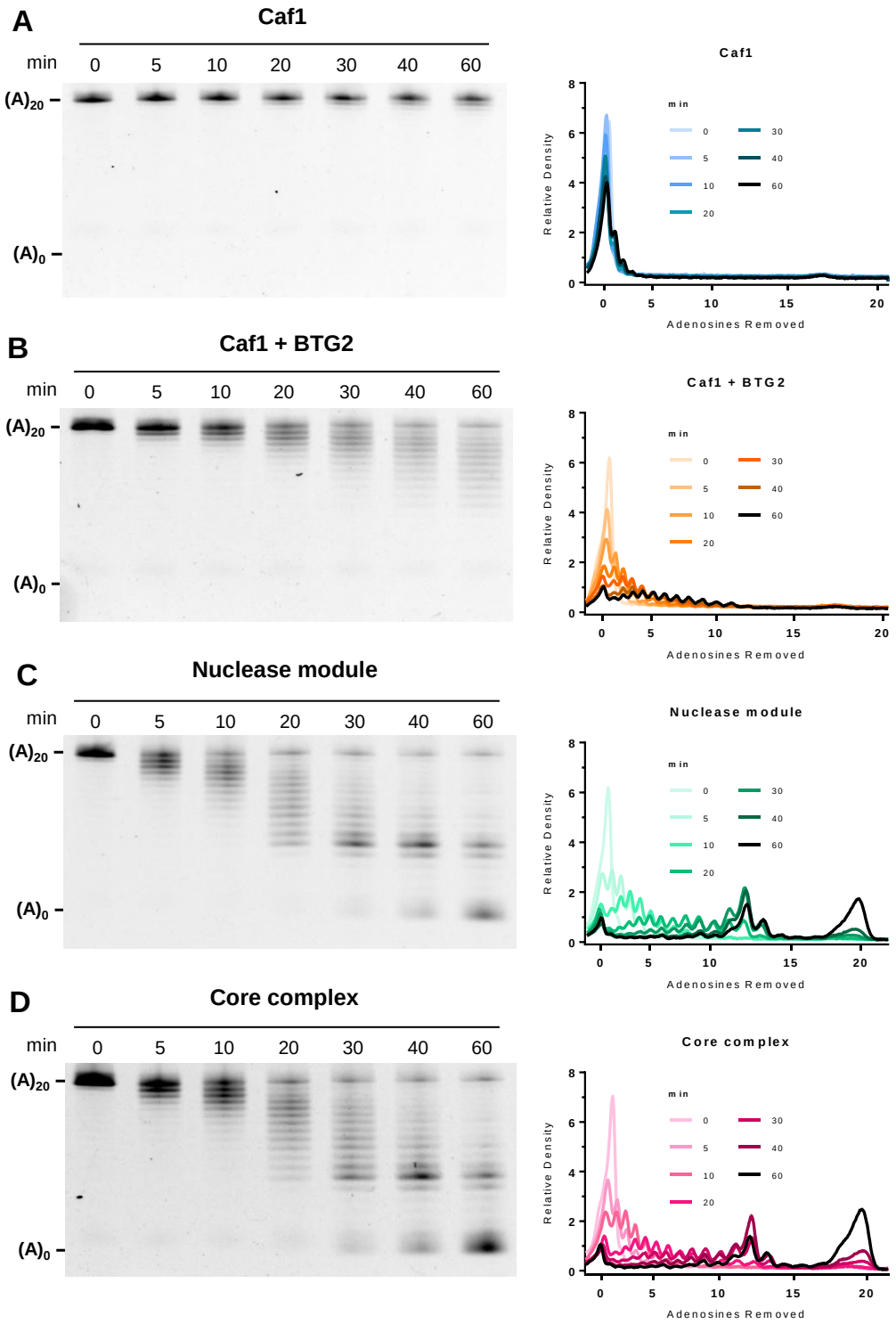



**Figure 5.9** – Comparison of the deadenylation activities of the Caf1 enzyme and the Caf1-containing complexes in presence of a 6-fold excess of PABPC1 (FL). Aliquots from the reactions were collected at specific time points and analysed

by 50% UREA/20% PAGE (left panels). Densitometry analysis of the scanned gels is provided (right panels). The number of adenosines removed is estimated based on the densitograms.

## 5.7 Deadenylase activity of Caf1 and Caf1-complexes using poly(A)$_{50}$-PABPC1 substrates

In this assay, the RNA substrate was a 57-mer composed of a fluorescein fluorophore conjugated to a 7-base sequence (CCUUUCC) followed by poly(A)$_{50}$. This substrate was complexed with full length PABPC1, or an N-terminal fragment of PABPC1 with RRM1 and RRM2 (1-190). Denaturing PAGE analysis was carried out to characterise deadenylation patterns.

### *5.7.1 A PAGE-based assay to characterise deadenylation activity*

This assay had a similar design to the one described above (Fig. 5.3). However, prior addition of the enzymes, RNA was incubated with an excess of PABPC1 (1-190) or PABPC1 FL. Then, a mixture of deadenylase enzyme (100 nM and 50 mM NaCl) and buffer containing $Mg^{2+}$(2 mM) was added and reactions were incubated for up to 60 minutes at 30°C. Aliquots were taken at specific time points and mixed with RNA denaturing dye and subsequently analysed with PAGE (16% poly-acrylamide and 50% UREA).

### *5.7.2 Activity of enzymes and complexes using a poly(A)$_{50}$ substrate*

In the presence of RNA coated with PABPC1 (1-190 or FL), the activity of Caf1 was hardly detectable after one hour (Fig. 5.10A and 5.11A). However, PABPC1 (1-190 or FL) stimulated the activity of the Caf1-BTG2 complex (Fig. 5.10B and 5.11B), when compared with uncoated poly(A)$_{50}$ (Fig. 5.5B). Similarly, the nuclease module showed increased deadenylation activity in presence of PABPC1 (Fig. 5.10C and 5.11C) when compared to uncoated poly(A)$_{50}$ substrate (Fig. 5.5C). Interestingly, the activity of the module seems to slow down after having removed about 40 bases (Fig. 5.10C and 5.11C). The core complex displayed a similar degradation pattern to the one generated by the nuclease module, but marginally faster rates (Fig. 5.10D and 5.11D). However, a noticeable decrease in the rate of deadenylation can be detected when compared with uncoated RNA reactions (Fig. 5.5D).

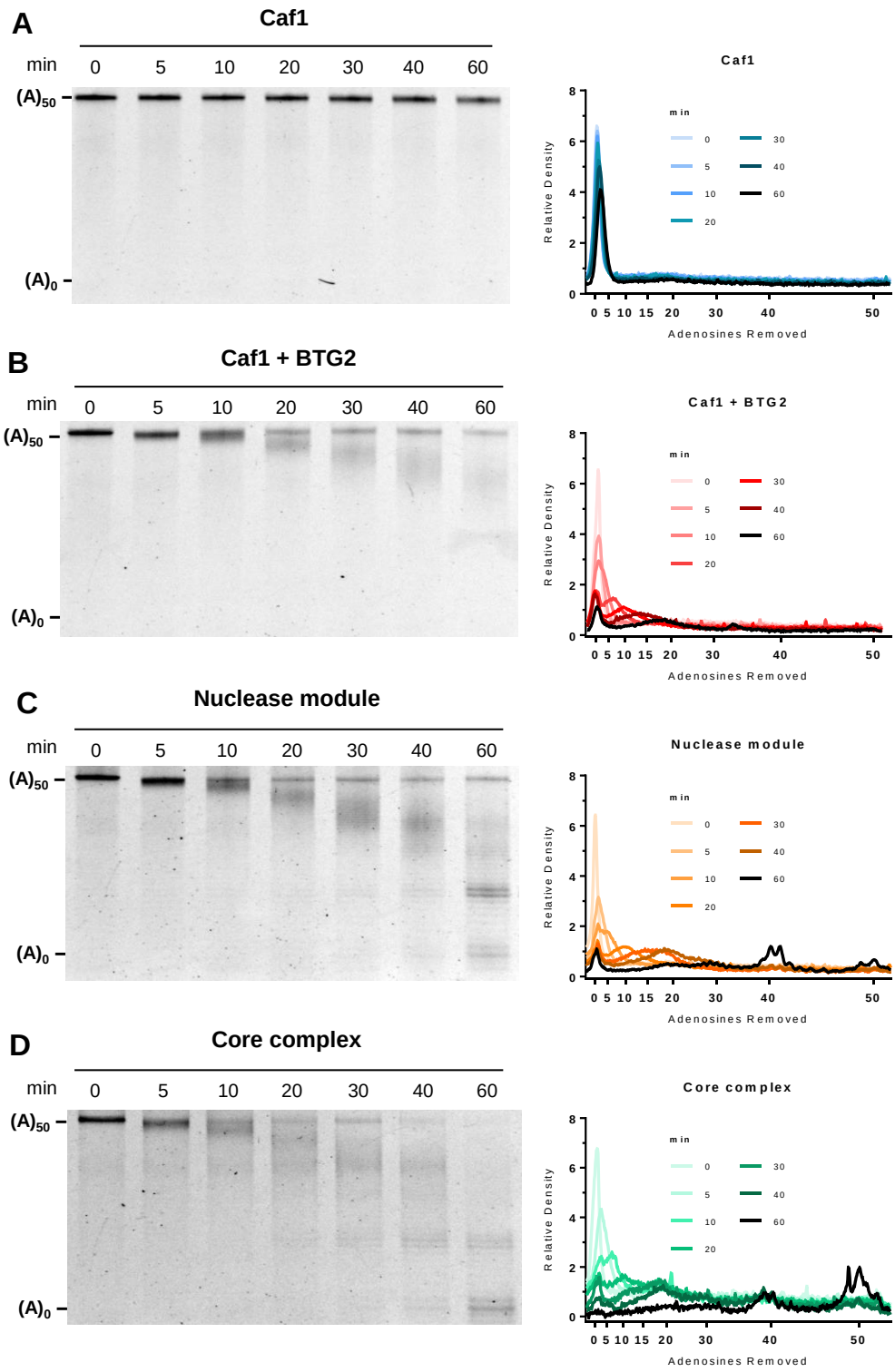



**Figure 5.10** –Deadenylation activities of Caf1 and Caf1-containing complexes using PABPC1-poly(A)$_{50}$ substrates. Protein-RNA complexes were prepared by using an 8-fold excess of PABPC1 (1-190) relative to poly(A)$_{50}$. Aliquots from

the reactions were collected at the indicated time points and analysed by 50% UREA/16% PAGE (left panels). Densitometry analysis of the scanned gels is provided (right panels). The number of adenosines removed is estimated based on the densitograms.

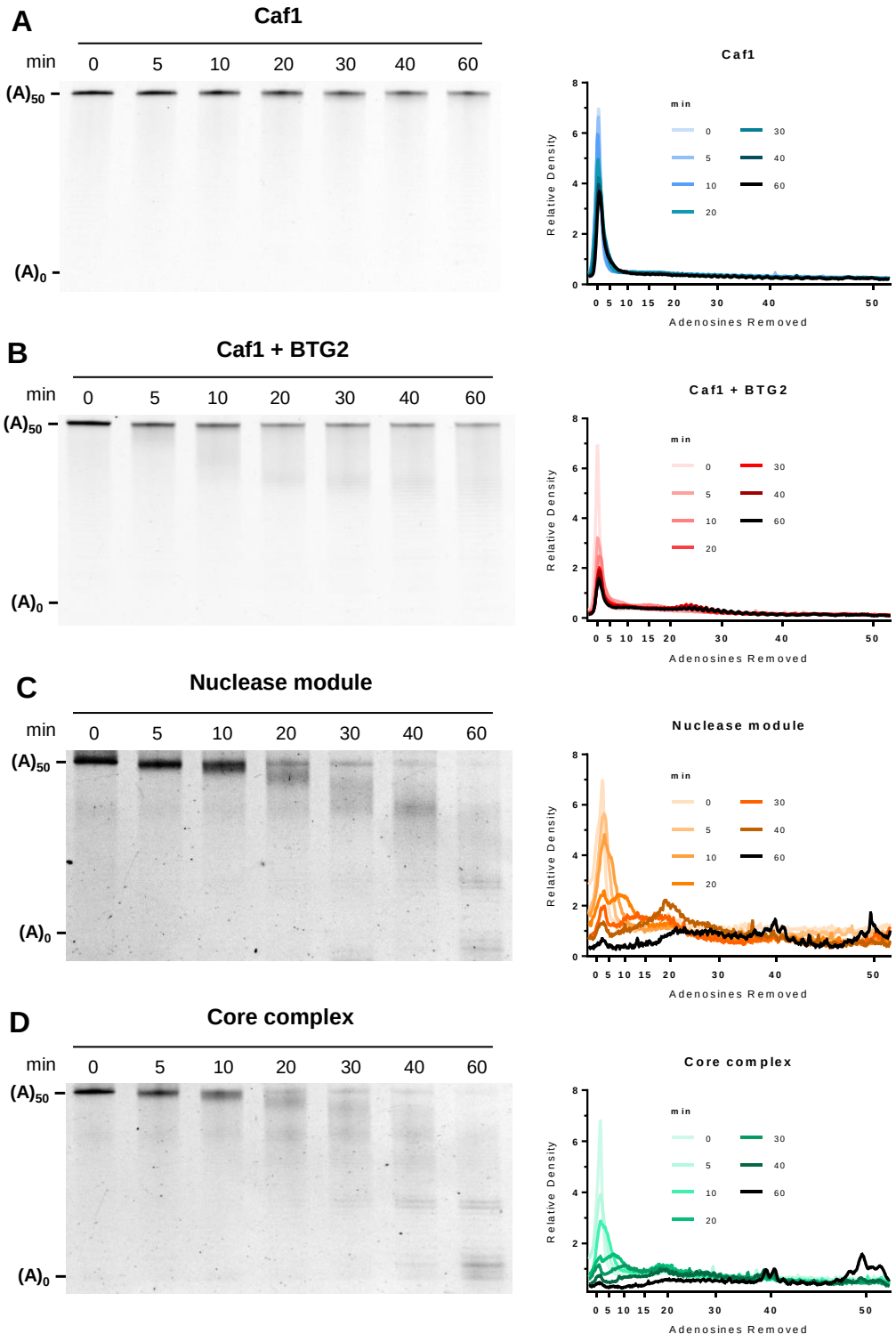



**Figure 5.11** – Deadenylation activities of Caf1 and Caf1-containing complexes using PABPC1-poly(A)$_{50}$ substrates. Protein-RNA substrates were prepared by adding a 6-fold excess of PABPC1 (FL) relative to poly(A)$_{50}$. Aliquots from the

reactions were collected at the indicated time points and analysed by 50% UREA/16% PAGE (left panels). Densitometry analysis of the scanned gels is provided (right panels). The number of adenosines removed is estimated based on the densitograms.

## 5.8 Selectivity for poly(A), poly(U), poly(C) and poly(G) RNA substrates

In this assay, different RNA substrates were used to assess the selectivity of the Caf1-containing deadenylase complexes. Specifically, $poly(A)_{20}$, $poly(U)_{20}$, $poly(C)_{20}$, $poly(G)_{20}$ substrates were selected to test the activity of Caf1, the nuclease module and the core complex, to try to understand whether a particular subunit/module contributes in selecting poly(A) tails as the preferential substrate for RNA degradation.

### *5.8.1 A PAGE-based assay to characterise deadenylation activity*

In this assay an equimolar amount of Caf1, the nuclease module and the Ccr4-Not complex were incubated at 30°C for 60 minutes with four different RNA substrates. After stopping the reactions by the addition of RNA loading dye and incubation at 85°C for 3 minutes, products were analysed by 50% UREA/20% PAGE (Fig. 5.12).

### *5.8.2 The Ccr4-Not core complex displays greater selectivity towards the poly(A) oligo when compared to the nuclease module*

An equimolar concentration of Caf1 enzyme and $poly(A)_{20}$ substrate was not sufficient to detect deadenylation activity (Fig. 5.12A). However, when reactions were set up with a 4-fold excess of Caf1 nuclease, full deadenylation of the substrate was achieved, while maintaining high selectivity towards adenosines (Fig. 5.12B). As expected, the nuclease module displayed higher deadenylation activity, but also showed partial removal of uridines and cytosines (Fig. 5.12C). Interestingly, even tough the core complex composed of the nuclease module and a CNOT1 (MIF4G-DUF2819)-CNOT9 dimer showed higher deadenylase activity compared to the nuclease module, the Ccr4-Not core complex showed increased selectivity towards poly(A) (Fig. 5.12D).

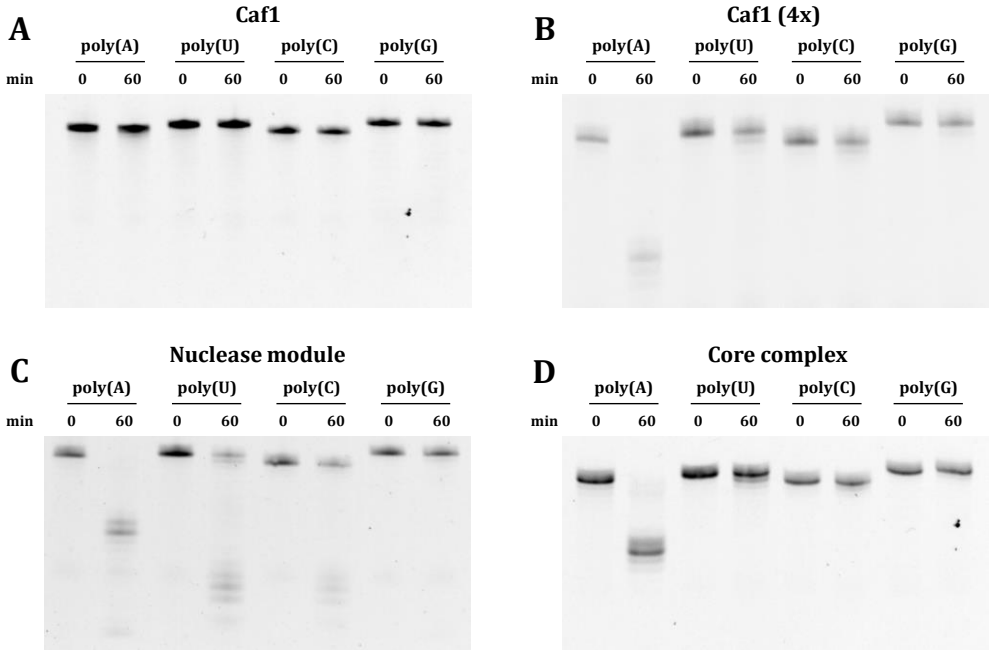



**Figure 5.12** – Comparison of the substrate selectivity of the Ccr4-Not core complexes. (A) Activity of Caf1. Reactions contained equimolar amounts of Caf1 and RNA substrates. (B) Activity of Caf1. Reactions contained a 4-fold excess of Caf1 with respect to the RNA substrates. (C) Activity of the nuclease module. Reactions contained equimolar amounts of the nuclease module and RNA substrates. (D) Activity of the nuclease module. Reactions contained equimolar amounts of the core complex and RNA substrates.

## 5.9 Discussion

The aim of the experiments described in this chapter was to understand the contribution towards deadenylation of different (non-catalytic) subunits and modules of the human Ccr4-Not complex. It is well established that the complex includes two catalytic subunits (Caf1 and Ccr4) but their relative contributions are not well understood to date: in *D. melanogaster* (Niinuma et al. 2016) and *S. pombe* (Stowell et al. 2016), the activities of the subunits appears additive, while the subunits are both required for deadenylation in the human nuclease module (Maryati et al. 2015). In addition, the non-catalytic subunits, such as CNOT1 and CNOT9, may have supplementary roles other than recruiting the complex according to the targeted mRNA degradation pathway mechanisms. For example, in *S. pombe*, the intact complex displays higher deadenylation activity as compared to the nuclease module, indicating that the non-catalytic subunits may stimulate the activity of Caf1 and Ccr4 (Stowell et al. 2016).

To try to dissect the different contributions of the Ccr4-Not complex subunits in the mRNA degradation process, four different Caf1-containing complexes were compared under a number of different conditions. First, a poly(A)$_9$-containing RNA substrate was used in a fluorescence-based deadenylation assay, to simulate a short substrate and to have a quantitative readout. Caf1 activity was found to be the modest, with an noticeable increase when BTG2 was present. The nuclease module, having an additional catalyitically active enzyme, showed faster deadenylation rates in agreement with experiments described before (Maryati et al. 2015). Interestingly, activity was further increased activity by the presence of a module composed of CNOT1M (encompassing the MIF4G and DUF3819 domains) and CNOT9 (Fig. 5.2). Similar results were obtained when using poly(A)$_{20}$ and poly(A)$_{50}$ substrates (Fig. 5.4 and 5.5), suggesting a regulatory role for the CNOT9-CNOT1M dimer.

Longer RNA substrates containing 20 or 50 adenosine residues bound by PABPC1 or PABPC1 RRM1-2 (Fig. 5.7) were used to assess how the BTG2-PABPC1 interaction was impacting the deadenylation rates. The poly(A)$_{20}$ was selected because it could accommodate one PABPC1 (RRM1-2 or FL) moiety, while the poly(A)$_{50}$ could be bound by at least two, and this could lead to differences in the BTG2-PABPC1 regulation. The synthesis of longer poly(A) substrates, for instance by Poly(A) polymerase, was investigated but then discarded as it was producing a population of different mRNAs, that would complicate data interpretation. Moreover, fluorescence-based assays were not performed for longer substrates as intermediate products would not be detected in such an end-point assay format; for longer substrates, the gel-based assay was selected as it was proving to give consistent data in different conditions. In general, no significant difference was detected between reactions containing full length PABPC1 or a fragment encompassing RRM1-RRM2, suggesting that the latter was a sufficient portion of the protein to trigger the BTG2-mediated mechanisms as reported by Stuplfer et al. (Stupfler et al. 2016). However, while accelerating the processivity of the Caf1-BTG2 dimer and the nuclease module, PABPC1 had a more modest effect on the activity of the Ccr4-Not core complex, if not reducing its degradation rates (Fig. 5.9D and 5.11D).

Poly(A)-tails degradation is the result of a combination of at least two activities by these enzymes and complexes. On one hand, they bind the target substrate and on the other they have a deadenylase catalytic activity. The assays showed in this chapter are mainly focusing on the latter, but further RNA-binding assays, such as EMSA, will be required to dissect the contribution of the former.

Finally, the degradation activity of these complexes was tested against non-poly(A)-containing RNA substrates, particularly to compare the selectivity of Caf1 and Caf1-containing complexes (Fig. 5.12). The Caf1 nuclease shows

a strict preference for poly(A) substrates. However, the nuclease module seems to display reduced selectivity and be able to partially degrade poly(U) and poly(C) substrates. This less stringent activity could be either caused by Ccr4 or by structural modifications occurring at the Caf1 active site when in complex with Ccr4 and BTG2. Interestingly, the human Ccr4-Not core complex shows an identical degradation pattern as the Caf1 enzyme alone, displaying high selectivity towards poly(A) RNA molecules.

Together, the experiments described in this chapter therefore suggest that the presence of the CNOT1M/CNOT9 module further increases the deadenylase activity of the BTG2-Caf1-Ccr4 nuclease module and improves its selectivity towards poly(A).

# Chapter 6

## Biophysical and biochemical characterisation of the melanoma-associated P131L variant of CNOT9

# Chapter 6. Biophysical and biochemical characterisation of the melanoma-associated P131L variant of CNOT9

## 6.1 Introduction

CNOT9 (also known as RQCD1 or Caf40) is one of the canonical non-catalytic subunits of the Ccr4-Not complex. Several crystal structures of CNOT9 are available: the apoprotein (Garces et al. 2007), CNOT9 in complex with the scaffold protein CNOT1 (Mathys et al. 2014) and in complex with short peptide derived from interacting co-factors such as Roquin (Sgromo, Raisch, Bawankar, et al. 2017) or Bag-of-marbles (Sgromo, Raisch, Backhaus, et al. 2017). Initial characterisation studies identified an interesting surface binding capability of CNOT9 for ribonucleic acids, with a preference for poly(C), poly(G) and poly(U), but not poly(A) oligo (Garces et al. 2007). Recent studies have identified tryptophan-binding pockets on the surface of CNOT9 that may mediate crucial protein-protein interactions with GW182 proteins to direct the Ccr4-Not complex towards miRNA targets (Behm-Ansmant et al. 2006; Fabian et al. 2012; Braun et al. 2011; Mathys et al. 2014).

Recently, a CNOT9 P131L variant has been detected with a 4% prevalence in primary melanomas (Wong et al. 2015). Such an important prevalence is likely to be underpinned by a considerable impact on either the structure of the protein or its interactions with deadenylation activators.

In order to investigate the potential effect of this variant on the structure of CNOT9, we isolated a recombinant human CNOT9 P131L. Moreover, to assess whether this mutation is affecting the interaction between CNOT9 and CNOT1, two additional complexes were purified; the first was a complex of the CNOT1 DUF3819 domain and CNOT9 containing the P131L variant. In addition, a dimeric complex composed of a CNOT1 fragment spanning the MIF4G and DUF3819 domains was isolated to reconstitute a

pentameric core complex containing a variant CNOT9 P131L subunit. To investigate the effect of the P131L substitution, the following approaches were used: (i) computational analysis by modeling the P131L amino acid substitution; (ii) crystallisation of the CNOT9 P131L protein and a complex of CNOT1 (DUF3819) and CNOT9 P131L; (iii) thermal stability analysis of CNOT9 and CNOT9 P131L; and (iv) deadenylation activity of the pentameric core complex containing the CNOT9 P131L variant. While the attempt to obtain structural insight of the CNOT9 P131L protein was not successful, the results indicate that the P131L substitution may reduce the stability of the ARM domain of CNOT9. However, it appears that this effect does not impact on the deadenylation activity of the pentameric core complex under the conditions used.

## 6.2 Computational modelling of the CNOT9 P131L variant

Using the UCSF Chimera package (Pettersen et al. 2004), computational analyses were performed on the structure of the CNOT9 P131L variant to assess a potential impact of this amino acid change on the local structure of the protein. In position 131, the wild-type proline was substituted with the leucine rotamer with the highest conformational probability (57%) according to the Dunbrack library (Dunbrack & Karplus 1993).

Due to the position within CNOT9, it was not likely that the P131L variant would affect the CNOT1-CNOT9 interaction, or any of the two tryptophan-binding pockets (Fig. 6.1A). However, the P131L variant could potentially impact the structure of CNOT9 because of the secondary structure where this residue is positioned. Pro-131 is part of α-helix 8, which is one of the 18 α-helices of CNOT9. Proline residues are frequently found in α-helices, as their rigid structure supports the correct folding of such structures. By contrast, a leucine residue is able to rotate more freely and therefore may affect the α-helix structure. Moreover, the leucine lateral chain is predicted to clash with the neighbouring α-helix 5 (Fig. 6.1B). Minimisation routines highlightened a local rearrangment of these two α-helices, in order to accommodate the substituted residue (Fig. 6.1B).

The modelling analysis provided evidence supporting a potential effect of the P131L variant on the structure of this protein. On this basis, CNOT9 P131L proteins and CNOT9 P131L-containing complexes were expressed and purified to obtain protein crystals for X-ray analysis, in order to investigate the local rearrangement upon the P131L amino acid substitution.

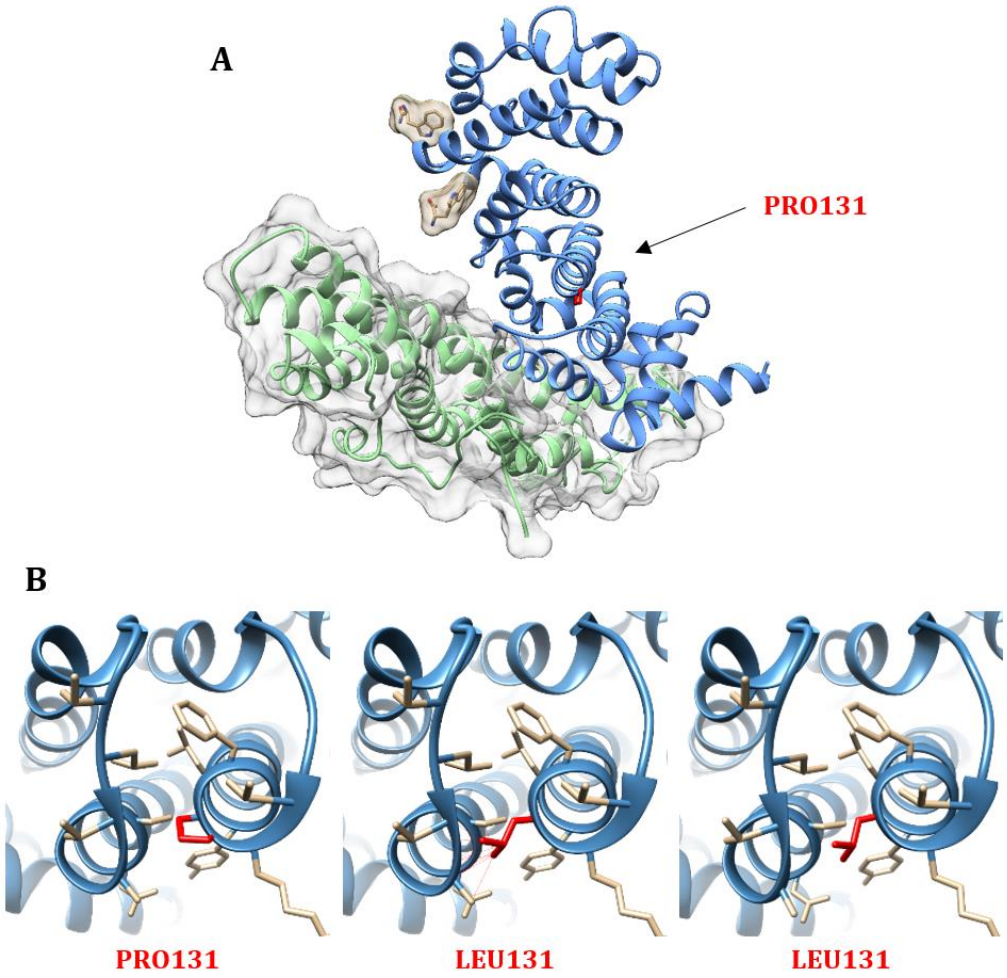



**Figure 6.1** – Structural analysis of the CNOT9 P131L variant. (A) Cartoon representation of the complex including the DUF3819 domain of CNOT1 (light green) and CNOT9 (blue), Pro-131 is depicted in red; tryptophans binding CNOT9 are depicted as surfaces (gold). (B) Modelling of the CNOT9 P131L variant. Proline-131 was substituted with a leucine rotamer (red stick) with the highest probability from the Dunbrack library of rotamers (57%). Clashes and contacts (red lines) with neighbouring atoms were then identified (B, middle panel) and the structure was minimised to accommodate the substituted rotamer (B, right panel). UCSF Chimera was used for structure visualisation and analysis. For energy minimisation, MMTK routines (Hinsen 2000) and Amber parameters (Case et al. 2005) were used.

## 6.3 Expression and purification of CNOT9 P131L protein complexes

### *6.3.1 Vector for protein expression in* E. coli *BL21 DE3*

To express and purify CNOT9 P131L in *E. coli*, site-directed mutagenesis was performed using vector pUC57 containing a codon-optimised CNOT9 CDS as a template. The presence of the mutation was confirmed by Sanger sequencing. Then the CDS was digested with BamHI and SalI restriction enzymes, present at the 5' and 3' end respectively. The digested CDS was cloned into a modified vector pQE80L linearised with identical restriction enzymes and containing, after the 6His-tag, an in-frame TEV-tag (Tobacco Etch Virus). The sequence of the resulting plasmid P07 was confirmed by Sanger sequencing (Fig. 6.2A).

To express the CNOT9 module displaying the P131L variant, CNOT1M CDS was digested using BamHI and SalI from vector P03 (Fig. 4.3A). The digested CDS was then cloned into a modified vector pACYC-Duet1 containing, after the 6His-tag, an in-frame TEV-tag in MCS 1. In the second step, CNOT9 P131L CDS was digested with BamHI and SalI restriction enzymes from the vector pUC57 and cloned into the MCS 2 of vector pACYC-Duet1 containing CNOT1M, linearised with BglII and XhoII. The sequence of the resulting plasmid P08 was confirmed by Sanger sequencing (Fig. 6.2B).

Finally, in order to express the heterodimer including the DUF3819 domain of CNOT1M and CNOT9 P131L a cloning approach was designed starting from plasmid P08. The CDS of the DUF3819 domain was amplified by PCR using primers including a BamHI restriction site at the 5' and a SalI restriction site at the 3' end. DUF3819 CDS was digested using BamHI and SalI and cloned into vector P08 linearised with identical restriction enzymes, in order to remove the CDS of CNOT1M from the MCS 1. The sequence of the resulting plasmid P09 was confirmed by Sanger sequencing (Fig. 6.2C).

### *6.3.2 Expression and purification of CNOT9 P131L and CNOT9 P131L-containing complexes*

A single colony of *E. coli* BL21 DE3 transformed with either vector P07 or P08 or P09 was used to inoculate a 2 mL LB culture with appropriate antibiotics for 4-6h. Twenty-five milliliters of LB (with appropriate antibiotics) were added to this culture and bacteria were incubated for 16 h at 37°C. Then, the culture was added to 1 L of LB and cultures were grown at 37°C until the $OD_{600}$ was between 0.7-0.8. IPTG (final concentration of 0.2 mM) was added to induce protein expression. Cultures were incubated for 16 h at 18°C and cells were harvested by centrifugation. Cells were resuspended in 5 mL Buffer HA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol, 5% glycerol and 20 mM imidazole) per litre of culture, cells were lysed by sonication and the soluble lysate was isolated by centrifugation. The soluble lysate was then loaded on a HisTrap FF pre-packed column (GE) equilibrated with Buffer HA (250 mM NaCl, 20 mM Tris-HCl pH 7.9, 1 mM β-mercaptoethanol, 5% glycerol and 20 mM imidazole). After washing the unbound fraction, five elution fractions of 1 mL were collected using Buffer HB (As Buffer HA, but containing 250 mM imidazole). Typically, fraction number two, containing about 85% of the purified protein (complex), was used in subsequent steps. One microgram of sample was analysed by SDS-PAGE (Fig. 6.3).

### *6.3.3 Ccr4-Not core complex with CNOT9 P131L*

Plasmid P08M was generated to express the CNOT9 module displaying the P131L mutation. Briefly, primers to introduce the P131L mutation were used to set up an inverse PCR reaction using vector P08 as template. The presence of the mutation was confirmed by Sanger sequencing. Vector P08M was used to express and purify the CNOT9 P131L module following the same method described for the wild type CNOT9 module. The CNOT9 P131L module was then incubated in a 1.2:1 fold excess with either wild

type or double mutant nuclease module and the core complexes were isolated by affinity chromatography as described above. A core complex with the correct stoichiometry was isolated (Fig. 6.3D).

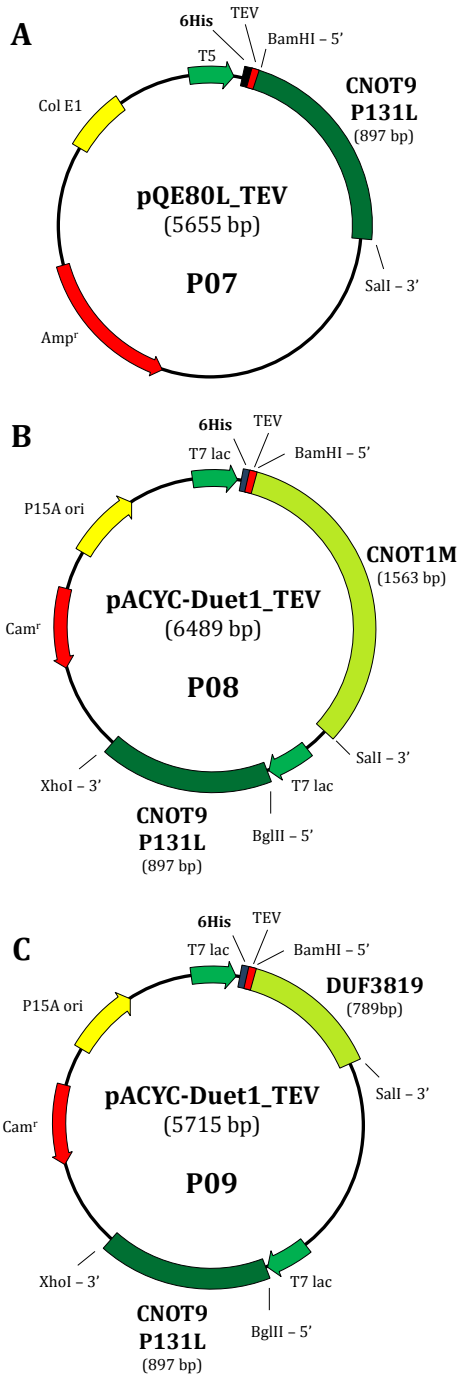



**Figure 6.2** – Vectors generated for the expression of CNOT9 P131L in *E. coli* BL21 DE3. (A) Map of plasmid pQE80L P07, expressing CNOT9 P131L fused with 6His and TEV. (B) Map of plasmid pACYC-Duet1 P08, co-expressing 6His and TEV-tagged CNOT1M and CNOT9 P131L. (C) Map of plasmid pACYC-Duet1 P08, co-expressing 6His and TEV-tagged DUF3819 and CNOT9 P131L.

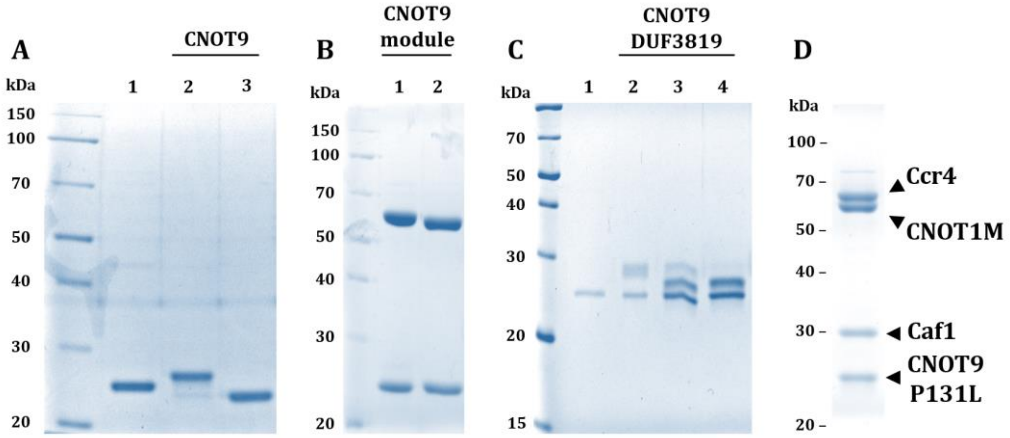



**Figure 6.3** – Expression of CNOT9 (P131L)-containing complexes. (A) Analysis of CNOT9L P131L expressed with plasmid P07 and purified by IMAC using 12% SDS-PAGE. Lanes 2 and 3 show CNOT9 P131L before and after proteolytic treatment with TEV protease (lane 1, for size reference). (B) Analysis of the CNOT9L P131L module with CNOT1M using 12% SDS-PAGE. The module was expressed with plasmid P08 and purified by IMAC. Lanes 1 and 2 show the module before and after proteolytic treatment with TEV protease. (C) Analysis of CNOT9L P131L in complex with DUF3819 using 14% SDS-PAGE. The heterodimer was expressed using plasmid P09 and purified by IMAC. Lane 2-4 show the dimer before (2) proteolytic treatment with TEV protease, after 24h (3) treatment and after 48h (4) treatment. (D) Analysis of the reconstituted Ccr4-Not core complex with CNOT9 displaying the P131L mutation using 12% SDS-PAGE.

## 6.4 Structural characterisation of the CNOT9 P131L variant

### *6.4.1 Differential scanning fluorimetry*

Differential scanning fluorimetry (DSF) is a relatively simple assay that is routinely used to characterise protein stability or protein-ligand interactions (Niesen et al. 2007). The denaturation midpoint ($T_m$) of a protein is measured by increasing the temperature of a protein solution in presence of a fluorescent dye with affinity for hydrophobic regions (Sypro Orange). As the temperature increases, the protein starts unfolding and therefore exposing its hydrophobic regions, allowing binding of the fluorescent dye.

DSF assays were used to investigate whether the P131L amino acid change was impacting the stability of CNOT9. Initially, four different concentrations of CNOT9 (wild-type and P131L variant) were tested, in the same reaction buffer, containing 250 mM NaCl (Fig. 6.4). In all the conditions tested, a temperature shift ($\Delta T_m$) greater than 2°C was observed. Then, two protein concentrations (5 μM and 10 μM) were tested in two additional buffer conditions (100 mM and 500 mM NaCl) to confirm that the observed $\Delta T_m$ was not dependent on the ionic strength of the solution (Fig. 6.5).

These observations indicate that the P131L amino acid substitution causes a reduction in the thermal stability of the ARM domain of CNOT9, which is consistent with the predictions based on computational modelling.

### *6.4.2 Crystallisation of the DUF3910-CNOT9 (P131L) complex*

The data from the DSF assays was suggesting an impact of the P131L amino acid substitution on the structure of CNOT9. To investigate this further, a crystallisation buffer screening was performed on both CNOT9 and the CNOT9-DUF3819 complex. Crystals for CNOT9 P131L were obtained but the overall structure does not appear to be different from the wild-type

CNOT9 protein. A limited electron density was identified in the P131L region, precluding conclusive analysis on the structural role of this amino acid substitution at this stage (Appendix 3).

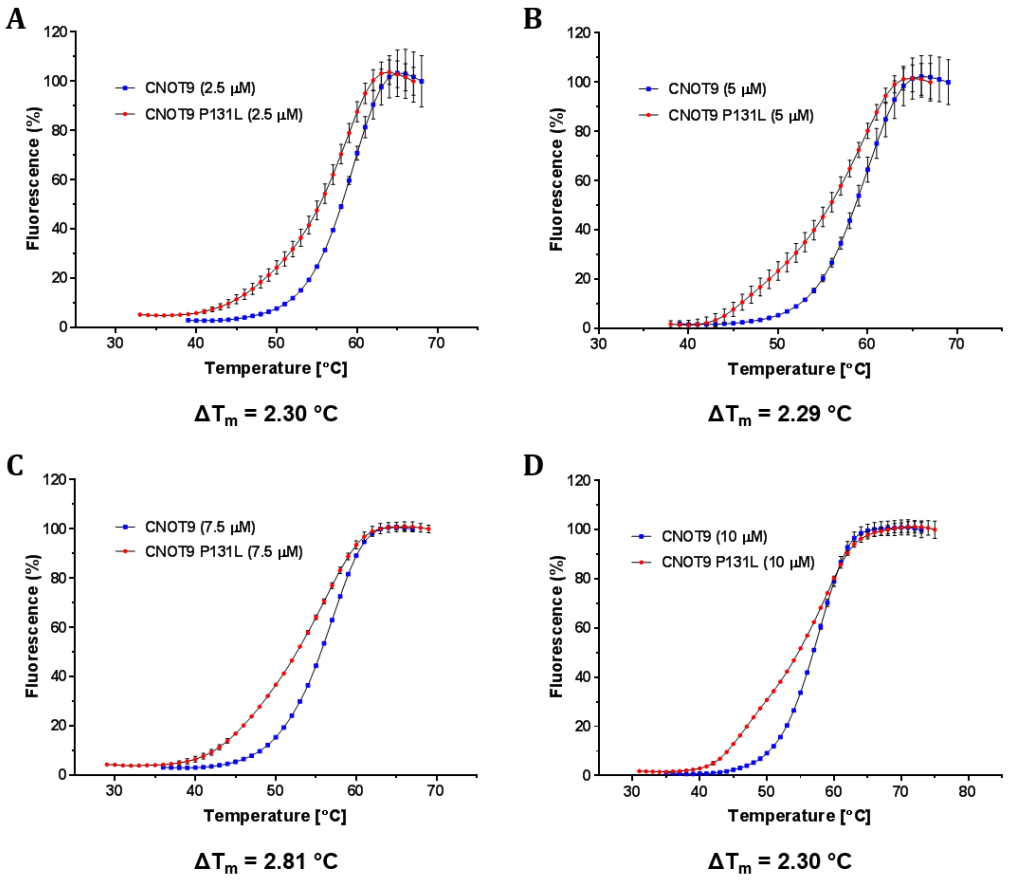



**Figure 6.4** – Differential scanning fluorimetry (DSF) analysis comparing CNOT9 wild-type with the P131L variant. Different protein concentrations were tested in the same buffering condition and comparable thermal shifts were detected.

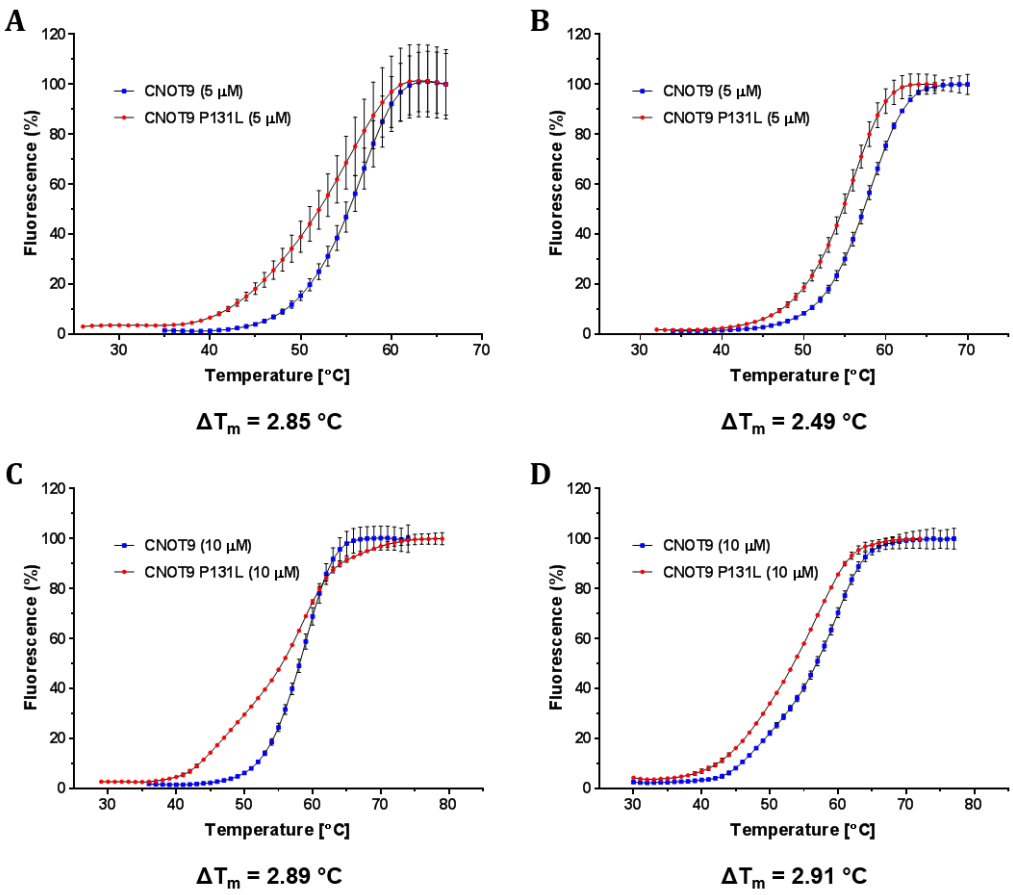



**Figure 6.5** – Differential scanning fluorimetry (DSF) analysis comparing wild-type CNOT9 and the P131L variant. Two protein concentrations (5 and 10 µM) were tested in two different buffers (100 mM NaCl for A and C; 500 mM NaCl for B and D).

## 6.5 Characterisation of the activity and selectivity of the Ccr4-Not core complex containing the CNOT9 P131L variant

### *6.5.1 Deadenylation assays*

To determine the effect of the P131L variant of CNOT9, the deadenylation activities of equimolar amounts of mutated and wild type Ccr4-Not core complexes were compared (Fig. 6.6). As described in Chapter 5, the catalytic activity was compared using an RNA poly(A)$_{20}$ substrate. PAGE analysis in denaturing conditions (50% urea) was carried out to visualise and characterise the deadenylation patterns with identical time point described above (Fig. 6.7). Under these conditions, no obvious differences in the deadenylation patterns were detected.

A similar assay was set up to compare the activities of these complexes in presence of a poly(A)$_{20}$ coated with either full length PABPC1, or an N-terminal fragment containing RRM1 and RRM2 only, as described in Chapter 5. PAGE analysis in denaturing conditions (50% urea) was carried out to visualise and characterise the deadenylation patterns (Fig. 6.8 and 6.9). Similarly to the uncoated poly(A)$_{20}$-containing assay, no significant differences were found.

Finally, the activity of the CNOT9 P131L-containing Ccr4-Not core complex was compared to the wild type in presence of a long poly(A)$_{50}$ oligo, under identical conditions as described in Chapter 5. PAGE analysis (50% urea) was carried out to visualise and characterise the deadenylation patterns, for which no significant difference was found (Fig. 6.10).

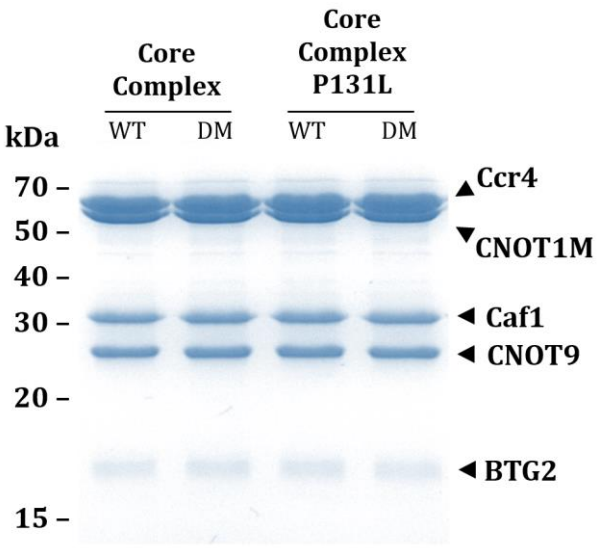



**Figure 6.6** - Analysis of the purified Ccr4-Not core complexes (Caf1, Ccr4, BTG2, CNOT1M and CNOT9). Protein samples (5 µg) were analysed by 14% SDS-PAGE and visualised by Coomassie staining. WT, wild type Caf1 and Ccr4; DM, Caf1 D40A/Ccr4 E240A double mutant.

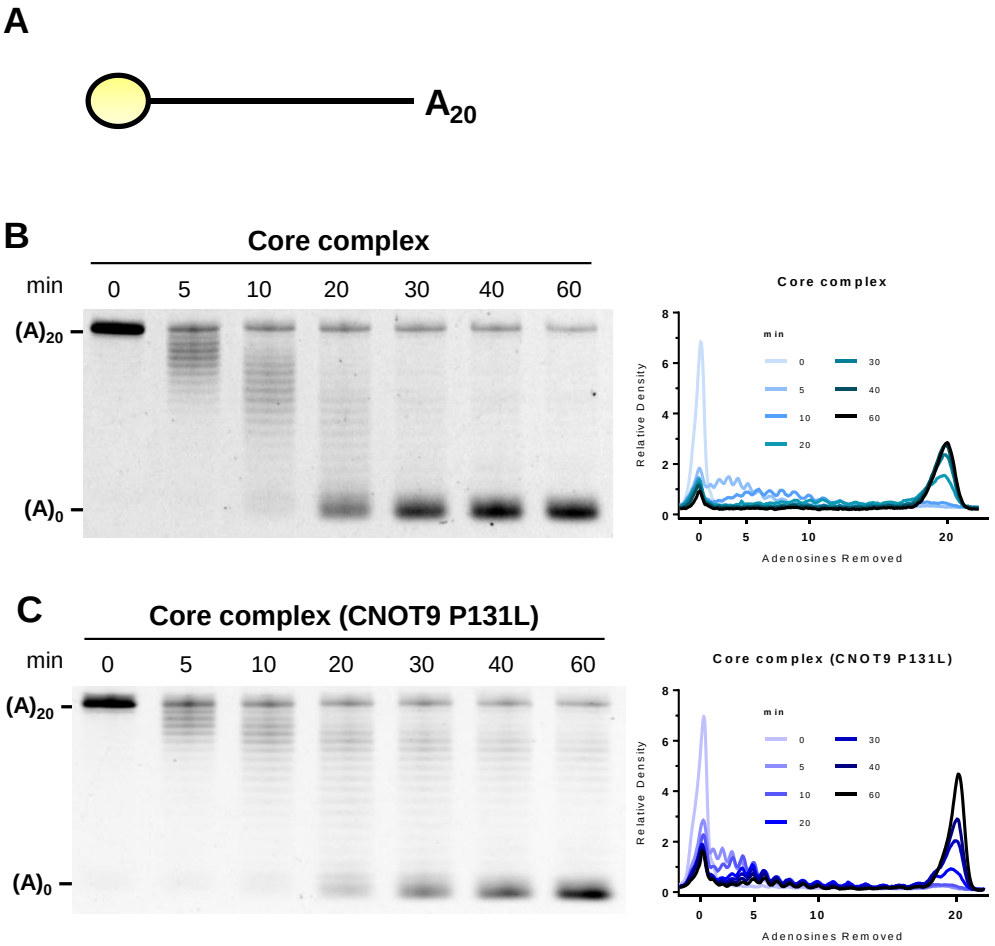



**Figure 6.7** - Comparison of the deadenylation activities of the Cc4-Not core complexes using a poly(A)$_{20}$ oligo as a substrate (A). Reactions were started upon addition of the deadenylase complexes (100 nM) to RNA substrate (200 nM). Aliquots from the reactions were collected at specific time points and analysed by 50% UREA/20% PAGE (B and C, left panels). Densitometry analysis of the scanned gels is provided (B and C, right panels). The number of adenosines removed is estimated based on the densitograms.

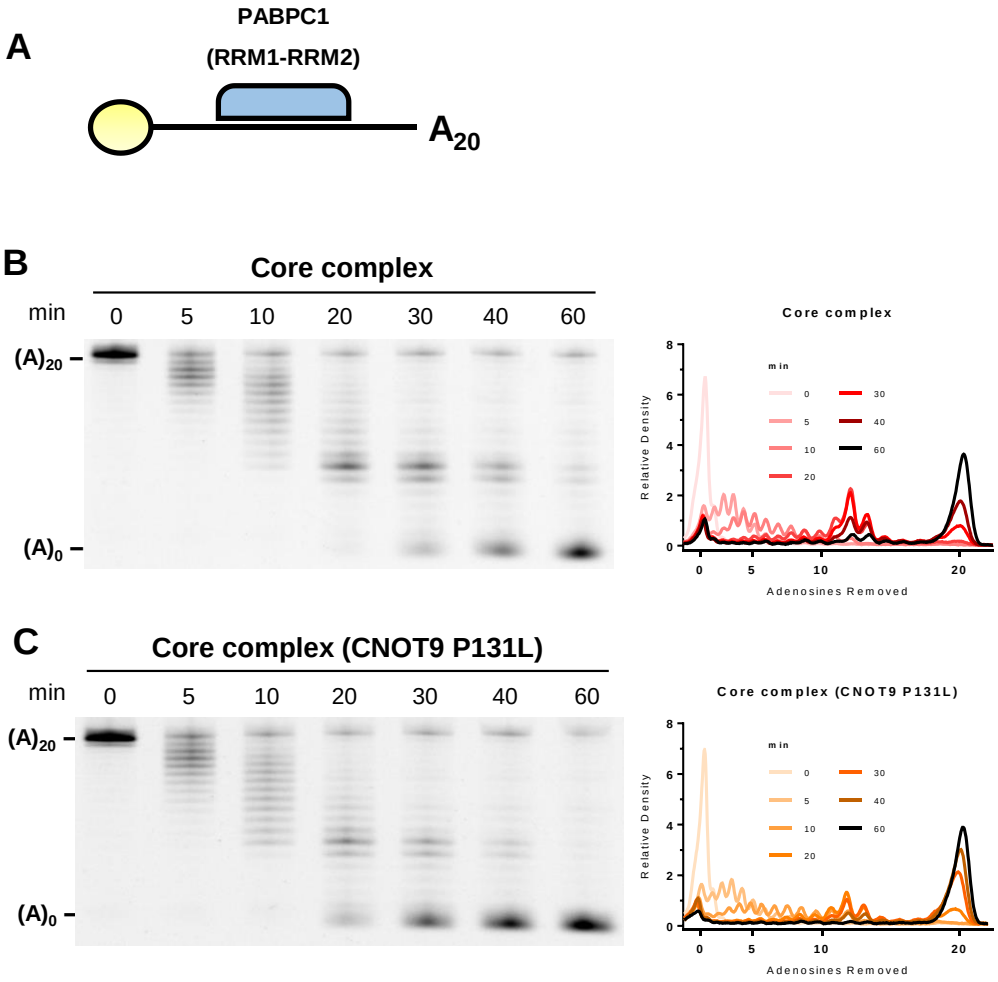



**Figure 6.8** – Comparison of the deadenylation activities of the Cc4-Not core complexes using a poly(A)$_{20}$ oligo coated with the RRM1-RRM2 domains of PABPC1 (A). Reactions were started upon addition of the deadenylase complexes (100 nM) to RNA substrate (200 nM). Aliquots from the reactions were collected at specific time points and analysed by 50% UREA/20% PAGE (B and C, left panels). Densitometry analysis of the scanned gels is provided (B and C, right panels). The number of adenosines removed is estimated based on the densitograms.

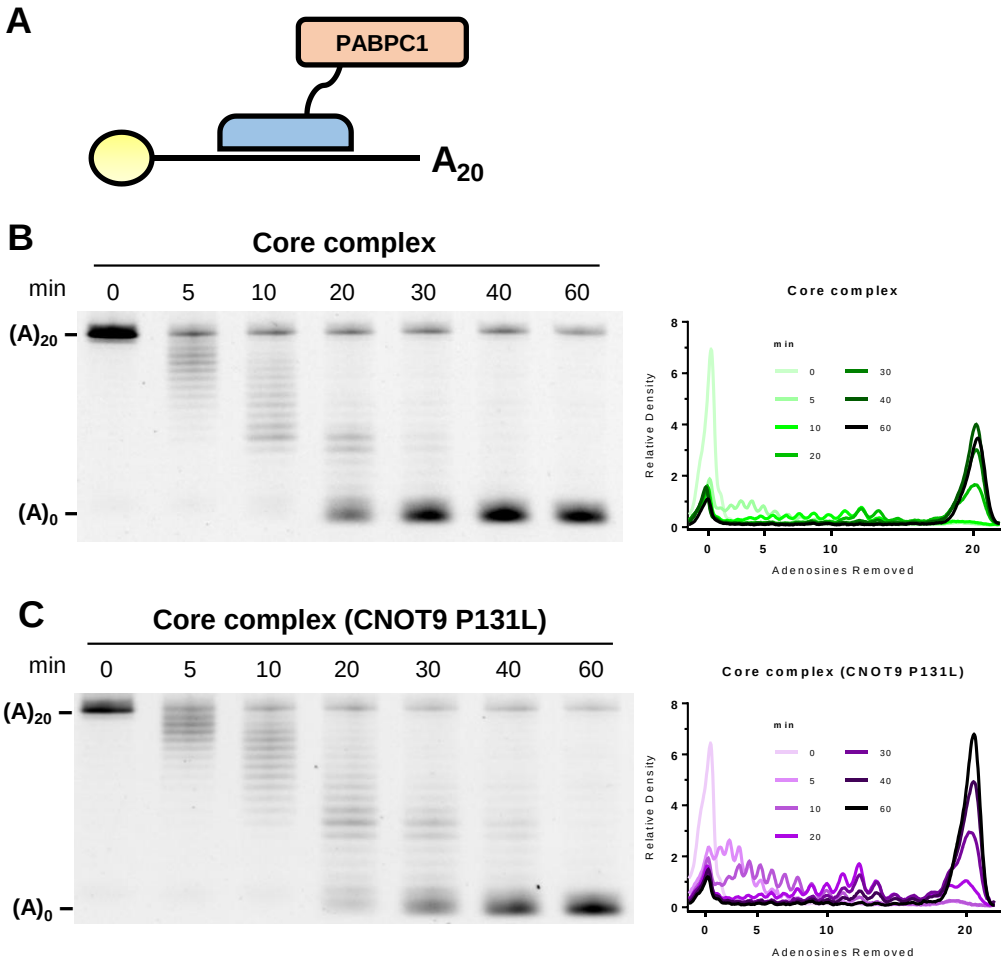



**Figure 6.9** – Comparison of the deadenylation activities of the Cc4-Not core complexes using a poly(A)$_{20}$ oligo coated with PABPC1 (A). Reactions were started upon addition of the deadenylase complexes (100 nM) to RNA substrate (200 nM). Aliquots from the reactions were collected at specific time points and analysed by 50% UREA/20% PAGE (B and C, left panels). Densitometry analysis of the scanned gels is provided (B and C, right panels). The number of adenosines removed is estimated based on the densitograms.

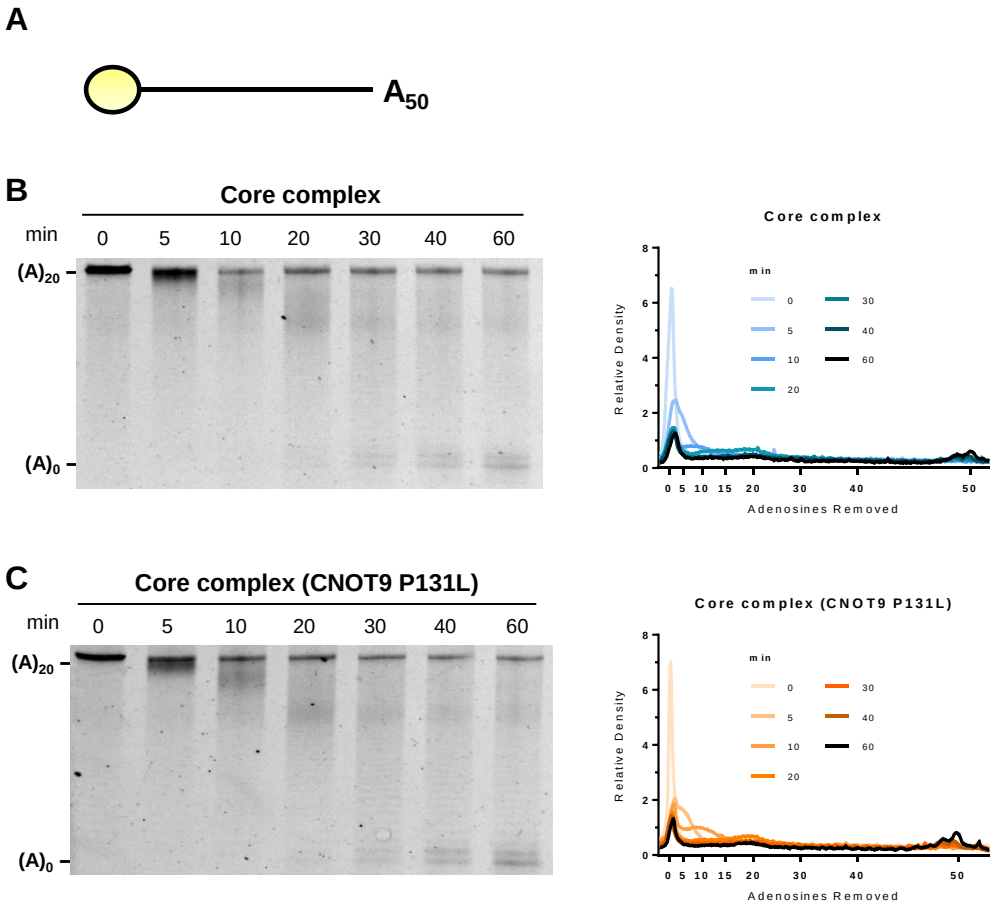



**Figure 6.10** - Comparison of the deadenylation activities of the Cc4-Not core complexes using a poly(A)$_{50}$ oligo as a substrate (A). Reactions were started upon addition of the deadenylase complexes (100 nM) to RNA substrate (200 nM). Aliquots from the reactions were collected at specific time points and analysed by 50% UREA/16% PAGE (B and C, left panels). Densitometry analysis of the scanned gels is provided (B and C, right panels). The number of adenosines removed is estimated based on the densitograms.

### *6.5.2 Selectivity assays*

To understand whether the P131L variant displayed an altered substrate preference, poly(A)$_{20}$, poly(U)$_{20}$, poly(C)$_{20}$, poly(G)$_{20}$ substrates were used to compare the activities of the wild type and CNOT9 P131L-containing Ccr4-Not core complexes.

An equimolar amount of each Ccr4-Not core complex was incubated at 30°C for 60 minutes with four different RNA substrates, as described above. After stopping the reactions by the addition of RNA loading dye and an incubation at 85°C for 3 minutes, products were analysed by 50% UREA/20% PAGE (Fig. 6.8). Under these conditions, the P131L variant does not appear to have a significant effect on substrate selectivity of the Ccr4-Not core complex.

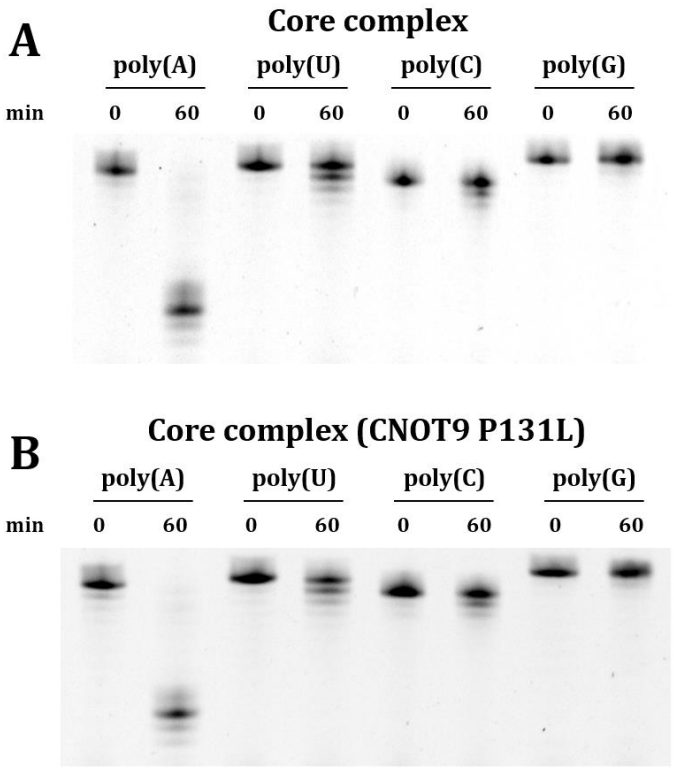



**Figure 6.11** - Comparison of the substrate selectivity of the Ccr4-Not core complexes. (A) Substrate selectivity of pentameric core complexes containing wild type CNOT9. (B) containing the CNOT9 P131L variant. Reactions contained equimolar concentration (100 nM) of the core complex and RNA substrates (200 nM). Reactions were incubated at 30°C for 60 minutes and aliquots were analysed by 50% UREA/18% PAGE.

## 6.6 Discussion

CNOT9 is one of the six non-catalytic subunits of the human Ccr4-Not complex (Wahle & Winkler 2013). Originally, CNOT9 was thought to directly interact with RNA molecules, with a preference for poly(U), poly(G) and poly(C) residues, but not poly(A) tails (Garces et al. 2007). However, more recently it has become clear that CNOT9 is a regulatory factor that influences Ccr4-Not activity by directing it to target mRNA molecules via interactions with regulatory RNA-binding proteins, including GW182 (Braun et al. 2011), Roquin (Sgromo, Raisch, Bawankar, et al. 2017), Bag-of-marbles (Bag) (Sgromo, Raisch, Backhaus, et al. 2017) and tristetraprolin (TTP) (Sandler et al. 2011; Fabian et al. 2013).

An interesting CNOT9 P131L variant has been recently detected in a whole exome sequencing study, with 4% prevalence in primary melanomas (Wong et al. 2015). The aim of this chapter was to further characterise this variant from both a structural and a biochemical point of view.

Initially, a modelling study was carried out to closely assess the potential impact on the structure of the protein. In CNOT9 wild type, the proline in position 131 is part of an α-helix structure exposed to the solvent but not in proximity of the CNOT1-binding surface. This suggests that an amino acid substitution in this position is not likely to directly impact the CNOT9-CNOT1 interaction. However, prolines are rigid amino acids, which help maintaining the correct folding of secondary structures such as α-helices. It can be envisaged that a P131L substitution, introducing a relatively flexible residue such as leucine, affects the stability of secondary structure elements and hence the thermal stability of protein domains. Moreover, the computational modelling detected several clashes of Leu-131 with the neighbouring α-helix, suggesting a local rearrengment of the protein structure to accommodate the substituted residue.

Following these findings, CNOT9 P131L was expressed and isolated in order to assess the impact of the residue change by thermal shift assays. Differential scanning fluorimetry experiments were optimised to compare CNOT9 wild type with P131L variant, and a negative shift of approximately 2°C was reproducible in all conditions. This indicated a moderate but significant impact of the amino acid substitution on the thermal stability of the ARM domain of CNOT9.

To further investigate this, protein crystals were obtained for X-ray analysis to obtain the structure of CNOT9 P131L. In order to enhance protein solubility and improve crystal packing, two additional protein complexes were expressed and purified. The first included CNOT9 (P131L) and its binding partner, the DUF3819 domain of CNOT1. The second consisted of CNOT9 (P131L) and CNOT1M, encompassing the MIF4G and DUF3819 domains. After two initial screens for crystallisation conditions, two conditions for crystallisation were found and followed-up with several rounds of optimisation plates. During the optimisations, flat plane-shaped crystals were obtained in several conditions. X-ray diffraction patterns were retrieved and analysed by Dr. Ingrid Dreveny and Emma Gunnell (School of Pharmacy, University Of Nottingham). The overall structure of CNOT9 P131L appears to closely overlap to that of the wild type protein, as expected. However, the limited local electron density precluded conclusive analysis on the effect of the P131L substitution.

The potential impact of the P131L variant was further investigated biochemically. The deadenylation activity of Ccr4-Not core complexes containing CNOT9 P131L was compared with wild type ones using short poly(A)$_{20}$ substrates in presence or absence of the coating protein PABPC1 (full length or RRM1-RRM2) and long poly(A)$_{50}$ oligos. No significant differences were detected under all conditions. Furthermore, selectivity assays were carried out to establish whether CNOT9 P131L was affecting

the preferential substrate of the Ccr4-Not core complex. However, the Ccr4-Not core complex including CNOT9 P131L showed identical deadenylation patterns compared to the wild type one.

It can therefore be concluded that the P131L variant of CNOT9 does not lead to obvious biochemical defects as far as the deadenylation activity of the core complex is concerned. It can be speculated that the defect caused by the P131L substitution impacts the non-catalytic roles of the complex where CNOT9 is involved. It may be that reduced thermal stability, as supported by computational modelling and thermal shift assays, leads to less efficient recruitment of the Ccr4-Not complex by the miRNA machinery or regulatory RNA-binding proteins such as Roquin or TTP. Alternatively, the P131L substitution may affect interactions with (yet unidentified) specific regulatory RNA-binding proteins.

# Chapter 7

## Concluding remarks

# Chapter 7. Concluding remarks

Cytoplasmic mRNA degradation is a key mechanism for regulated gene expression in eukaryotic cells. The initial step for this mechanism is the shortening and removal of the poly(A) tail, or the deadenylation process. In eukaryotes, the PAN2-PAN3 and the Ccr4-Not complexes are pivotal components in deadenylation.

PAN2-PAN3 is a trimer composed of two regulatory subunits (PAN3) and a catalytic enzyme (PAN2). While several studies have shed light on the structure (Jonas et al. 2014; Schäfer et al. 2014; Wolf et al. 2014), its precise role in eukaryotes is not clear yet. Similarly, only little is known about the regulatory mechanisms that target this complex to specific mRNAs.

The Ccr4-not complex is composed of at least eight subunits, constituting four modules, that bind the scaffold protein CNOT1: the N-terminal module (CNOT10, CNOT11) (Bawankar et al. 2013; Mauxion et al. 2013); the catalytic nuclease module comprising Caf1 and Ccr4 binding to the MIF4G domain of CNOT1 (Petit et al. 2012; Maryati et al. 2015); CNOT9 subunit associated with the DUF3819 domain of CNOT1 (Chen et al. 2014); and the NOT module, which is composed of CNOT2 and CNOT3 binding to a C-terminal region of CNOT1 (Boland et al. 2013). A few mechanisms that direct the Ccr4-Not complex to mRNA targets are known (Fabian et al. 2013; Sgromo, Raisch, Bawankar, et al. 2017), and a picture is emerging in which this complex is recruited by proteins inducing mRNA degradation and translational repression. However, the role and contribution of some of the non-catalytic subunits is not clear.

## 7.1 PAN2-PAN3 complex characterisation

In Chapter 3, several potential assays are described that can be used to characterise the PAN2-PAN3 complex in more detail from both a cellular and a biochemical prospective. To better understand its role in human cells

and in P-body formation, genetically modified cell lines were designed by using the recent CRISPR/Cas9 technology for genomic engineering. A HEK293 cell line displaying a knock-out for PAN2 was created. An interesting characterisation for these cell lines was assessing the impact of gene knock-out and/or point mutations on P-body formation. In order to help selecting relevant point-mutations, a yeast two-hybrid experiment was designed and optimised. With this biochemical assay, several mutations linked with cancer could be screened and amino acid changes impacting the interaction between different subunits in the complex identified to be reproduced with CRISPR/Cas9 and test in a cell-based assay. To this end, an immunocytochemistry-based assay was developed to readily detect and count P-bodies in HEK293 cells (Appendix 2). Using this assay, fewer P-bodies were detected in HEK293 cells expressing exogenous inactive mutant Caf1, when compared to wild type Caf1. This assay will be also interesting in order to test the cellular effects of small drug-like molecules targeting Caf1 (Maryati et al. 2014; Jadhav et al. 2015).

To characterise the biochemical activity of the PAN2-PAN3 complex, an extensive set of experiments was designed and carried out to express and purify a recombinant human complex in *E. coli*. Despite the number of conditions explored, only insoluble complexes were expressed, preventing purification and testing in assays *in vitro*. Future studies will explore alternative protein expression systems, such as mammalian cells or baculovirus-insect cells (Uchida et al. 2004).

### 7.2 Reconstitution of a human Ccr4-Not core complex

In Chapter 4, the reconstitution of a recombinant human Ccr4-Not core complex is described. A series of purification strategies were designed, and a novel pentameric core complex including the nuclease module (BTG2, Caf1 and Ccr4) and the CNOT9 module (CNOT9 and CNOT1M) was expressed and purified. The successful purification strategy involved the

separate purification of the nuclease and CNOT9 modules followed by reconstitution by incubation and isolation of pentameric complexes by affinity purification.

Future studies will focus on trying to expand this core complex with additional subunits, such as the NOT module, to better mimic the complex in human cells. In addition, structural studies on the Ccr4-Not core complex here described could give insight on the spatial arrangement of the included subunits, possibly providing indications on their role. To this end, cryo-EM structural analysis or alternative methods such as protein EPR or single-molecule FRET will be pursued. Pilot experiments have been carried out to measure the distance between Caf1 and Ccr4 active sites in the human nuclease module of the Ccr4-Not complex by EPR spectroscopy (Appendix 4). On the other hand, single-molecule FRET could be used to detect a signal change between fluorophores introduced at specific sites in the complex, via click-chemistry on unnatural amino acids (Chin et al. 2002; Tyagi & Lemke 2013). A pilot experiment was successfully carried out to test the feasibility of introducing unnatural amino acids in proteins of interest (Appendix 5).

### 7.3 Deadenylation activity of the Ccr4-Not core complex

Using the reconstituted Ccr4-Not core complex, the contribution of the non-catalytic CNOT1 and CNOT9 subunits towards the deadenylation activity was subsequently determined (Chapter 5). Several assays were designed to compare the Ccr4-Not core complex with the catalytic subunit Caf1 and the nuclease module (Ccr4, Caf1 and BTG2) using model poly(A) substrates. Interestingly, the CNOT9 module, composed of CNOT9 and a middle fragment of CNOT1 (MIF4G and DUF3819 domains), has a remarkable impact on deadenylation rates. When compared to the nuclease module, the core complex displays higher catalytic activity with all types of substrates tested. This indicates that non-catalytic subunits contribute to the

enzymatic activity, which is in agreement with results obtained using *S. pombe* Ccr4-Not (Stowell et al. 2016).

In this chapter, the regulatory effects of PABPC1 on the Ccr4-Not complex, via its interaction with BTG2, were investigated by coating poly(A) substrates with PABPC1 full length or the RRM1-RRM2 domains only. Collectively, the data from these assays indicate a dual contribution of PABPC1 in poly(A) tail degradation. On one hand, it promotes deadenylation of the first 11-12 nucleotides via its interaction with BTG2; when reaching this threshold, the catalytic activity of the deadenylase enzymes seems to be remarkably reduced, possibly indicating that PABPC1 needs to be displaced before the rest of the poly(A) can be degraded. Accordingly, two 'pauses' in catalytic activity are detected when a longer $poly(A)_{50}$ substrate is coated with PABPC1. Interestingly, no difference between coating with full length PABPC1 or RRM1-RRM2 domains only is observed, confirming the dual function of these domains to both bind poly(A) tails and interact with BTG/Tob proteins (Stupfler et al. 2016). The next steps will focus on investigating more closely the molecular details of the PABPC1-BTG2 interaction.

Future studies will focus on increasing the complexity of this system, for instance by introducing additional components that mimic more closely the human targeted deadenylation process. Potential targeted deadenylation mechanisms to be reconstituted include TTP and ARE-containing mRNAs (Sandler et al. 2011; Fabian et al. 2013) or Roquin (Leppek et al. 2013; Janowski et al. 2016; Sgromo, Raisch, Bawankar, et al. 2017).

Moreover, small drug-like inhibitors targeting Caf1 (Maryati et al. 2014; Jadhav et al. 2015) will be tested under these conditions, where a more complete Ccr4-Not complex is likely to closely resemble the environment of the drug target in cells.

## 7.4 Investigation on CNOT9 P131L variant associated with melanoma

Finally, the effect of the CNOT9 P131L amino acid substitution was investigated (Chapter 6). This variant is associated with primary melanoma (Wong et al. 2015). The characterisation focused on both a structural and a biochemical level. Under the experimental conditions used, data indicate that a Ccr4-Not core complex including CNOT9 P131L retains identical catalytic activity compared to wild type CNOT9, both in terms of deadenylation rates and substrate selectivity. It is possible that P131L effects are too small to be detected in this assay format.

While no effect on the activity of the Ccr4-Not core complex was detected, this variant could potentially impact the structural stability of the protein. Differential scanning fluorimetry indicated that the P131L amino acid substitution caused a consistent negative shift of 2-3°C in thermal stability assays, showing that the variant was impacting CNOT9 folding to some extent. To further investigate the structural change, protein crystals of CNOT9 (P131L), CNOT9 (P131L)-DUF3819 and CNOT9 (P131L)-CNOT1M were grown. Crystallisation buffer screening helped identifying a condition where CNOT9 P131L was forming plane-like crystals (Appendix 3). Single plane crystals were collected and analysed at the Diamond Light Source facility (Didcot, UK). However, due to the low electron density around the amino acid position 131, this approach was inconclusive. Future studies will be needed to optimise crystallisation conditions for CNOT9 P131L and to promote crystal formation for CNOT1-CNOT9 complexes, to further investigate the impact of the P131L variant.

## 7.5 Future outlook

The work described in this thesis presents a novel approach to reconstitute a recombinant human Ccr4-Not core complex, including two active deadenylases (Caf1 and Ccr4), a co-factor BTG2, a central fragment of

CNOT1 (MIF4G and DUF3819 domains) and the regulatory protein CNOT9. This procedure is scalable and yields milligrams of stable and active protein complex. When compared with single nucleases and the nuclease module only, the Ccr4-Not core complex displays faster deadenylation rates and higher substrate selectivity, indicating a strong contribution of the CNOT9 module. In the future, these experimental approaches could be expanded to investigate other non-catalytic subunits of the Ccr4-Not complex, such as the Not module, to highlight the role of these subunits in deadenylation.

## References


Ajiro, M. et al., 2010. Critical involvement of RQCD1 in the EGFR-Akt pathway in mammary carcinogenesis. *International Journal of Oncology*, 37(5), pp.1085–1093.

Ajiro, M. et al., 2009. Involvement of RQCD1 overexpression, a novel cancer-testis antigen, in the Akt pathway in breast cancer cells. *International Journal of Oncology*, 35(4), pp.673–681.

Amrani, N. et al., 2008. Translation factors promote the formation of two states of the closed-loop mRNP. *Nature*, 453(7199), pp.1276–1280.

Andrei, M.A. et al., 2005. A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. *RNA (New York, N.Y.)*, 11(5), pp.717–727.

Araki, Y. et al., 2001. Ski7p G protein interacts with the exosome and the ski complex for 3???-to-5??? mRNA decay in yeast. *EMBO Journal*, 20(17), pp.4684–4693.

Ashkenazy, H. et al., 2010. ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research*, 38(SUPPL. 2).

Ashkenazy, H. et al., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, 44(W1), pp.1–7.

Aslam, A. et al., 2009. The Ccr4-NOT deadenylase subunits CNOT7 and CNOT8 have overlapping roles and modulate cell proliferation. *Molecular biology of the cell*, 20(17), pp.3840–50.

Aslam, A., 2010. The Ccr4–Not Deadenylase Subunits CNOT7 and CNOT8 Have Overlapping Roles and Modulate Cell Proliferation. *Molecular biology of the cell*, 21, pp.4042–4056.

Ayache, J. et al., 2015. P-body assembly requires DDX6 repression complexes rather than decay or Ataxin2/2L complexes. *Molecular Biology of the Cell*, 26(14), pp.2579–2595.

Bartel, D.P., 2009. MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, 136(2), pp.215–233.

Basquin, J. et al., 2012. Architecture of the nuclease module of the yeast ccr4-Not complex: The not1-caf1-ccr4 interaction. *Molecular Cell*, 48(2), pp.207–218.

Bawankar, P. et al., 2013. NOT10 and C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto the NOT1 N-terminal domain. *RNA biology*, 10(2), pp.228–44.

Beelman, C. a et al., 1996. An essential component of the decapping enzyme required for normal rates of mRNA turnover. *Nature*, 382(6592), pp.642–646.

Behm-Ansmant, I. et al., 2006. MRNA degradation by miRNAs and GW182 requires both CCR4 : NOT deadenylase and DCP1 : DCP2 decapping complexes. *Genes & Development*, 20, p.1885–1898 ST–MRNA

degradation by miRNAs and GW1.
Bett, J.S. et al., 2013. The P-body component USP52/PAN2 is a novel regulator of HIF1A mRNA stability. *The Biochemical journal*, 451(2), pp.185–94. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3632086&tool=pmcentrez&rendertype=abstract.
Bhandari, D. et al., 2014. Structural basis for the Nanos-mediated recruitment of the CCR4-NOT complex and translational repression. *Genes and Development*, 28(8), pp.888–901.
Bhattacharyya, S.N. et al., 2006. Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress. *Cell*, 125(6), pp.1111–1124.
Boeck, R. et al., 1996. The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *Journal of Biological Chemistry*, 271(1), pp.432–438.
Boland, A. et al., 2013. Structure and assembly of the NOT module of the human CCR4-NOT complex. *Nature structural & molecular biology*, 20(11), pp.1289–97.
Bonneau, F. et al., 2009. The Yeast Exosome Functions as a Macromolecular Cage to Channel RNA Substrates for Degradation. *Cell*, 139(3), pp.547–559.
Bouveret, E. et al., 2000. A Sm-like protein complex that participates in mRNA degradation. *The EMBO journal*, 19(7), pp.1661–71.
Braun, J.E. et al., 2012. A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5′ exonucleolytic degradation. *Nature Structural & Molecular Biology*, 19(12), pp.1324–1331.
Braun, J.E. et al., 2011. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Molecular cell*, 44(1), pp.120–33.
Braun, J.E. et al., 2010. The C-terminal alpha-alpha superhelix of Pat is required for mRNA decapping in metazoa. *The EMBO journal*, 29(14), pp.2368–80.
Brengues, M., Teixeira, D. & Parker, R., 2005. Cell biology: Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science*, 310(5747), pp.486–489.
Brogna, S. & Wen, J., 2009. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology*, 16(2), pp.107–113.
Buratowski, S., 2009. Progression through the RNA Polymerase II CTD Cycle. *Molecular Cell*, 36(4), pp.541–546.
Burgess, H.M. & Gray, N.K., 2010. mRNA-specific regulation of translation by poly(A)-binding proteins. *Biochemical Society Transactions*, 38(6), pp.1517–1522. Available at: http://biochemsoctrans.org/lookup/doi/10.1042/BST0381517.
Case, D.A. et al., 2005. The Amber biomolecular simulation programs.

*Journal of Computational Chemistry*, 26(16), pp.1668–1688.
Celniker, G. et al., 2013. ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Israel Journal of Chemistry*, 53(3–4), pp.199–206.
Chang, C. Te et al., 2014. The activation of the decapping enzyme DCP2 by DCP1 occurs on the EDC4 scaffold and involves a conserved loop in DCP1. *Nucleic Acids Research*, 42(8), pp.5217–5233.
Chang, H. et al., 2014. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Molecular cell*, 53(6), pp.1044–52.
Charlesworth, A., Meijer, H.A. & De Moor, C.H., 2013. Specificity factors in cytoplasmic polyadenylation. *Wiley Interdisciplinary Reviews: RNA*, 4(4), pp.437–461.
Chekulaeva, M. et al., 2011. miRNA repression involves GW182-mediated recruitment of CCR4–NOT through conserved W-containing motifs. *Nature Structural & Molecular Biology*, 18(11), pp.1218–1226.
Chen, C.-Y.A., 2013. Deadenylation and P-bodies. *Advances in Experimental Medicine and Biology*, 768(713), pp.1–12.
Chen, C.-Y.A. & Shyu, A.-B., 2011. Mechanisms of deadenylation-dependent decay. *Wiley Interdisciplinary Reviews: RNA*, 2(2), pp.167–183.
Chen, C. et al., 2011. Distinct expression patterns of the subunits of the CCR4-NOT deadenylase complex during neural development. *Biochemical and Biophysical Research Communications*, 411(2), pp.360–364.
Chen, C.Y. et al., 2001. AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell*, 107(4), pp.451–464.
Chen, C.Y.A. et al., 2009. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nature Structural and Molecular Biology*, 16(11), pp.1160–1166.
Chen, J. et al., 2001. Purification and characterization of the 1.0 MDa CCR4-NOT complex identifies two novel components of the complex. *Journal of molecular biology*, 314(4), pp.683–94.
Chen, N. et al., 2005. Crystal Structures of Human DcpS in Ligand-free and m7GDP-bound forms Suggest a Dynamic Mechanism for Scavenger mRNA Decapping. *Journal of Molecular Biology*, 347(4), pp.707–718.
Chen, Y. et al., 2014. A DDX6-CNOT1 Complex and W-Binding Pockets in CNOT9 Reveal Direct Links between miRNA Target Recognition and Silencing. *Molecular Cell*, 54(5), pp.737–750.
Chin, J.W. et al., 2002. Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *Journal of the American Chemical Society*, 124(31), pp.9026–9027.
Chowdhury, a. et al., 2012. Both Sm-domain and C-terminal extension of Lsm1 are important for the RNA-binding activity of the Lsm1-7-Pat1 complex. *Rna*, 18(5), pp.936–944.
Chowdhury, A., Kalurupalle, S. & Tharun, S., 2014. Pat1 contributes to the RNA binding activity of the Lsm1-7-Pat1 complex. *RNA (New York,*

*N.Y.)*, 20(9), pp.1465–75.
Chowdhury, A., Mukhopadhyay, J. & Tharun, S., 2007. The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA (New York, N.Y.)*, 13(7), pp.998–1016.
Chowdhury, A. & 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 (New York, N.Y.)*, 14(10), pp.2149–58.
Christie, M. et al., 2013. Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. *Molecular Cell*, 51(3), pp.360–373.
Conaway, R.C. & Conaway, J.W., 2011. Function and regulation of the Mediator complex. *Current Opinion in Genetics and Development*, 21(2), pp.225–230.
Cougot, N., Babajko, S. & Séraphin, B., 2004. Cytoplasmic foci are sites of mRNA decay in human cells. *Journal of Cell Biology*, 165(1), pp.31–40.
Cramer, P., 2001. Structural Basis of Transcription: RNA Polymerase II at 2.8 Angstrom Resolution. *Science*, 292(5523), pp.1863–1876. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1059493.
Das, R. et al., 2007. SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. *Molecular Cell*, 26(6), pp.867–881.
Deo, R.C. et al., 1999. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell*, 98(6), pp.835–845.
Deshmukh, M. V. et al., 2008. mRNA Decapping Is Promoted by an RNA-Binding Channel in Dcp2. *Molecular Cell*, 29(3), pp.324–336.
Díaz-Peña, R. et al., 2012. A high density SNP genotyping approach within the 19q13 chromosome region identifies an association of a CNOT3 polymorphism with ankylosing spondylitis. *Annals of the rheumatic diseases*, 71(5), pp.714–7.
Doidge, R. et al., 2012. The Anti-Proliferative Activity of BTG/TOB Proteins Is Mediated via the Caf1a (CNOT7) and Caf1b (CNOT8) Deadenylase Subunits of the Ccr4-Not Complex. *PLoS ONE*, 7(12).
Doma, M.K. & Parker, R., 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, 440(7083), pp.561–564.
Draper, M.P., Salvadore, C. & Denis, C.L., 1995. Identification of a mouse protein whose homolog in Saccharomyces cerevisiae is a component of the CCR4 transcriptional regulatory complex. *Molecular and Cellular Biology*, 15(7), pp.3487–3495.
Du, H. et al., 2016. YTHDF2 destabilizes m 6 A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nature Communications*, 7.
Dunbrack, R.L. & Karplus, M., 1993. Backbone-dependent rotamer library

for proteins: Application to side-chain prediction. *Journal of Molecular Biology*, 230(2), pp.543–574.
Dziembowski, A. et al., 2007. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nature Structural & Molecular Biology*, 14(1), pp.15–22.
Eberle, A.B. et al., 2009. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nature Structural & Molecular Biology*, 16(1), pp.49–55.
Egloff, S. & Murphy, S., 2008. Cracking the RNA polymerase II CTD code. *Trends in Genetics*, 24(6), pp.280–288.
Eulalio, A., Behm-Ansmant, I., Schweizer, D., et al., 2007. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Molecular and cellular biology*, 27(11), pp.3970–81.
Eulalio, A., Behm-Ansmant, I. & Izaurralde, E., 2007. P bodies: at the crossroads of post-transcriptional pathways. *Nature Reviews Molecular Cell Biology*, 8(1), pp.9–22.
Ezzeddine, N. et al., 2007. Human TOB, an Antiproliferative Transcription Factor, Is a Poly(A)-Binding Protein-Dependent Positive Regulator of Cytoplasmic mRNA Deadenylation. *Molecular and Cellular Biology*, 27(22), pp.7791–7801.
Ezzeddine, N., Chen, C.-Y.A. & Shyu, A.-B., 2012. Evidence providing new insights into TOB-promoted deadenylation and supporting a link between TOB's deadenylation-enhancing and antiproliferative activities. *Molecular and cellular biology*, 32(6), pp.1089–98.
Fabian, M.R., 2011. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. *Nature structural & molecular biology*, 18(11).
Fabian, M.R. et al., 2012. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. *Nature Structural & Molecular Biology*, 19(3), pp.1211–1217.
Fabian, M.R. et al., 2013. Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. *Nature structural & molecular biology*, 20(6), pp.735–9.
Fabrega, C. et al., 2003. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Molecular Cell*, 11(6), pp.1549–1561.
Faraji, F. et al., 2016. Post-transcriptional Control of Tumor Cell Autonomous Metastatic Potential by CCR4-NOT Deadenylase CNOT7. *PLoS Genetics*, 12(1).
Färber, V. et al., 2013. Trypanosome CNOT10 is essential for the integrity of the NOT deadenylase complex and for degradation of many mRNAs. *Nucleic Acids Research*, 41(2), pp.1211–1222.
Frazão, C. et al., 2006. Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature*, 443(7107), pp.110–4.

Garces, R.G., Gillon, W. & Pai, E.F., 2007. Atomic model of human Rcd-1 reveals an armadillo-like-repeat protein with in vitro nucleic acid binding properties. *Protein science : a publication of the Protein Society*, 16(2), pp.176–88.

Garneau, N.L., Wilusz, J. & Wilusz, C.J., 2007. The highways and byways of mRNA decay. *Nature reviews. Molecular cell biology*, 8(February), pp.113–126.

Glaser, F. et al., 2003. ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, 19(1), pp.163–164.

Glover-Cutter, K. et al., 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature Structural and Molecular Biology*, 15(1), pp.71–78.

Goldstrohm, A.C. et al., 2006. PUF proteins bind Pop2p to regulate messenger RNAs. *Nature Structural & Molecular Biology*, 13(6), pp.533–539.

Greenblatt, J., Nodwell, J.R. & Mason, S.W., 1993. Transcriptional antitermination. *Nature*, 364(6436), pp.401–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8332211.

Gutierrez-Camino, A. et al., 2014. Noncoding RNA-related polymorphisms in pediatric acute lymphoblastic leukemia susceptibility. *Pediatric research*, 75(6), pp.767–773.

Ha, M. & Kim, V.N., 2014. Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology*, 15(8), pp.509–524.

Haas, G. et al., 2010. HPat provides a link between deadenylation and decapping in metazoa. *Journal of Cell Biology*, 189(2), pp.289–302.

Hemsley, A. et al., 1989. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Research*, 17(21), p.8915.

Hernández, H. et al., 2006. Subunit architecture of multimeric complexes isolated directly from cells. *EMBO reports*, 7(6), pp.605–10.

Hinsen, K., 2000. The Molecular Modeling Toolkit: A New Approach to Molecular Simulations. *Journal of Computational Chemistry*, 21(2), pp.79–85.

Hiroi, N. et al., 2002. Mammalian Rcd1 is a novel transcriptional cofactor that mediates retinoic acid-induced cell differentiation. *EMBO Journal*, 21(19), pp.5235–5244.

Hirose, Y. & Manley, J.L., 2000. RNA polymerase II and the integration of nuclear events. *Genes and Development*, 14(12), pp.1415–1429.

Horiuchi, M. et al., 2009. Structural basis for the antiproliferative activity of the Tob-hCaf1 complex. *Journal of Biological Chemistry*, 284(19), pp.13244–13255.

Huang, K.-L. et al., 2013. Phosphorylation at intrinsically disordered regions of PAM2 motif-containing proteins modulates their interactions with PABPC1 and influences mRNA fate. *RNA (New York,*

*N.Y.)*, 19(3), pp.295–305.
Huang, Y. et al., 2003. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Molecular Cell*, 11(3), pp.837–843.
Huntzinger, E. & Izaurralde, E., 2011. Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. *Nature Reviews Genetics*, 12(2), pp.99–110.
Hutvágner, G. et al., 2001. A cellular function for the RNA-interference enzyme dicer in the maturation of the let-7 small temporal RNA. *Science*, 293(5531), pp.834–838.
Imataka, H., Gradi, A. & Sonenberg, N., 1998. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO Journal*, 17(24), pp.7480–7489.
Ingelfinger, D. et al., 2002. The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. *Rna*, 8(12), pp.1489–1501.
Ito, K., Inoue, T., et al., 2011. CNOT2 depletion disrupts and inhibits the CCR4-NOT deadenylase complex and induces apoptotic cell death. *Genes to Cells*, 16(4), pp.368–379.
Ito, K., Takahashi, A., et al., 2011. The role of the CNOT1 subunit of the CCR4-NOT complex in mRNA deadenylation and cell viability. *Protein and Cell*, 2(9), pp.755–763.
Izaurralde, E. et al., 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell*, 78(4), pp.657–668.
Jadhav, G.P. et al., 2015. Discovery, synthesis and biochemical profiling of purine-2,6-dione derivatives as inhibitors of the human poly(A)-selective ribonuclease Caf1. *Bioorganic and Medicinal Chemistry Letters*, 25(19), pp.4219–4224.
Janowski, R. et al., 2016. Roquin recognizes a non-canonical hexaloop structure in the 3'-UTR of Ox40. *Nature communications*, 7, p.11032.
Januszyk, K., Liu, Q. & Lima, C.D., 2011. Activities of human RRP6 and structure of the human RRP6 catalytic domain. *RNA*, 17(8), pp.1566–1577.
Jayne, S. et al., 2006. Involvement of the SMRT/NCoR-HDAC3 complex in transcriptional repression by the CNOT2 subunit of the human Ccr4-Not complex. *The Biochemical journal*, 398, pp.461–467.
Jinek, M., Coyle, S.M. & Doudna, J.A., 2011. Coupled 5' nucleotide recognition and processivity in Xrn1-mediated mRNA decay. *Molecular cell*, 41(5), pp.600–8.
Joly, W. et al., 2013. The CCR4 deadenylase acts with Nanos and Pumilio in the fine-tuning of Mei-P26 expression to promote germline stem cell self-renewal. *Stem cell reports*, 1(5), pp.411–24.
Jonas, S. et al., 2014. An asymmetric PAN3 dimer recruits a single PAN2 exonuclease to mediate mRNA deadenylation and decay. *Nature structural & molecular biology*, 21(7), pp.599–608.

Kawamata, T. & Tomari, Y., 2010. Making RISC. *Trends in Biochemical Sciences*, 35(7), pp.368–376.
Kim, I. et al., 2012. B-Catenin recognizes a specific RNA motif in the cyclooxygenase-2 mRNA 3′-UTR and interacts with HuR in colon cancer cells. *Nucleic Acids Research*, 40(14), pp.6863–6872.
Köhler, A. & Hurt, E., 2007. Exporting RNA from the nucleus to the cytoplasm. *Nature Reviews Molecular Cell Biology*, 8(10), pp.761–773.
Kuehner, J.N., Pearson, E.L. & Moore, C., 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews Molecular Cell Biology*, 12(5), pp.283–294.
Landau, M. et al., 2005. ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Research*, 33(SUPPL. 2).
Lau, N.-C. et al., 2009. Human Ccr4-Not complexes contain variable deadenylase subunits. *The Biochemical journal*, 422(3), pp.443–453.
Lebreton, A. et al., 2008. Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature*, 456(7224), pp.993–996.
Lee, Y. et al., 2002. MicroRNA maturation: Stepwise processing and subcellular localization. *EMBO Journal*, 21(17), pp.4663–4670.
Lee, Y. et al., 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956), pp.415–419.
Leppek, K. et al., 2013. Roquin promotes constitutive mrna decay via a conserved class of stem-loop recognition motifs. *Cell*, 153(4), pp.869–881.
Liu, Q., Greimann, J.C. & Lima, C.D., 2006. Reconstitution, Activities, and Structure of the Eukaryotic RNA Exosome. *Cell*, 127(6), pp.1223–1237.
Lorentzen, E. et al., 2008. Structure of the Active Subunit of the Yeast Exosome Core, Rrp44: Diverse Modes of Substrate Recruitment in the RNase II Nuclease Family. *Molecular Cell*, 29(6), pp.717–728.
Mahajan, M.A., Murray, A. & Samuels, H.H., 2002. NRC-interacting factor 1 is a novel cotransducer that interacts with and regulates the activity of the nuclear hormone receptor coactivator NRC. *Mol.Cell Biol.*, 22(19), pp.6883–6894.
Maryati, M. et al., 2014. A fluorescence-based assay suitable for quantitative analysis of deadenylase enzyme activity. *Nucleic acids research*, 42(5), p.e30.
Maryati, M., Airhihen, B. & Winkler, G., 2015. The enzyme activities of Caf1 and Ccr4 are both required for deadenylation by the human Ccr4-Not nuclease module. *Biochemical Journal*, 4922873455(May), p.Epub ahead of print.
Mathys, H. et al., 2014. Structural and Biochemical Insights to the Role of the CCR4-NOT Complex and DDX6 ATPase in MicroRNA Repression. *Molecular Cell*, 54(5), pp.751–765.
Mauxion, F. et al., 2009. BTG/TOB factors impact deadenylases. *Trends in*

*Biochemical Sciences*, 34(12), pp.640–647.
Mauxion, F., Prève, B. & Séraphin, B., 2013. C2ORF29/CNOT11 and CNOT10 form a new module of the CCR4-NOT complex. *RNA biology*, 10(2), pp.267–76.
McCracken, S. et al., 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, 385(6614), pp.357–361.
Midtgaard, S.F. et al., 2006. Structure of the nuclear exosome component Rrp6p reveals an interplay between the active site and the HRDC domain. *Proceedings of the National Academy of Sciences of the United States of America*, 103(32), pp.11898–11903.
Milac, A.L., Bojarska, E. & Wypijewska del Nogal, A., 2014. *Decapping Scavenger (DcpS) enzyme: Advances in its structure, activity and roles in the cap-dependent mRNA metabolism*,
Mittal, S. et al., 2011. The Ccr4a (CNOT6) and Ccr4b (CNOT6L) deadenylase subunits of the human Ccr4-Not complex contribute to the prevention of cell death and senescence. *Molecular biology of the cell*, 22(6), pp.748–758.
Morita, M. et al., 2011. Obesity resistance and increased hepatic expression of catabolism-related mRNAs in Cnot3(+/-) mice. *The EMBO journal*, 30(22), pp.4678–4691.
Mugridge, J.S. et al., 2016. Structural basis of mRNA-cap recognition by Dcp1–Dcp2. *Nature Structural & Molecular Biology*, 23(11), pp.987–994.
Mukherjee, D. et al., 2002. The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO Journal*, 21(1–2), pp.165–174.
Nasertorabi, F. et al., 2011. Insights into the structure of the CCR4-NOT complex by electron microscopy. *FEBS Letters*, 585(14), pp.2182–2186.
Neely, G.G. et al., 2010. A Global In Vivo Drosophila RNAi Screen Identifies NOT3 as a Conserved Regulator of Heart Function. *Cell*, 141(1), pp.142–153.
Niesen, F.H., Berglund, H. & Vedadi, M., 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, 2(9), pp.2212–2221.
Niinuma, S., Fukaya, T. & Tomari, Y., 2016. CCR4 and CAF1 deadenylases have an intrinsic activity to remove the post-poly(A) sequence. *RNA (New York, N.Y.)*, pp.1–10.
Nissan, T. et al., 2010. Decapping Activators in Saccharomyces cerevisiae Act by Multiple Mechanisms. *Molecular Cell*, 39(5), pp.773–783.
Ozgur, S., Chekulaeva, M. & Stoecklin, G., 2010. Human Pat1b connects deadenylation with mRNA decapping and controls the assembly of processing bodies. *Molecular and cellular biology*, 30(17), pp.4308–23.

Page, A.M. et al., 1998. Mutational analysis of exoribonuclease I from Saccharomyces cerevisiae. *Nucleic acids research*, 26(16), pp.3707–16.

Panepinto, J.C., Heinz, E. & Traven, A., 2013. The cellular roles of Ccr4-NOT in model and pathogenic fungi-implications for fungal virulence. *Frontiers in Genetics*, 4(DEC).

Parker, R. & Sheth, U., 2007. P Bodies and the Control of mRNA Translation and Degradation. *Molecular Cell*, 25(5), pp.635–646.

Pellegrini, O. et al., 2008. Chapter 9 In Vitro Assays of 5′ to 3′-Exoribonuclease Activity. *Methods in Enzymology*, 448, pp.167–183.

Petit, A.P. et al., 2012. The structural basis for the interaction between the CAF1 nuclease and the NOT1 scaffold of the human CCR4-NOT deadenylase complex. *Nucleic Acids Research*, 40(21), pp.11058–11072.

Pettersen, E.F. et al., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), pp.1605–12.

Piccirillo, C., Khanna, R. & Kiledjian, M., 2003. Functional characterization of the mammalian mRNA decapping enzyme hDcp2. *RNA (New York, N.Y.)*, 9(9), pp.1138–47.

Pilkington, G.R. & Parker, R., 2008. Pat1 contains distinct functional domains that promote P-body assembly and activation of decapping. *Molecular and cellular biology*, 28(4), pp.1298–1312.

Proudfoot, N.J., 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science*, 352(6291).

Proudfoot, N.J., Furger, A. & Dye, M.J., 2002. Integrating mRNA processing with transcription. *Cell*, 108(4), pp.501–512.

Quesada, V. et al., 2004. Cloning and enzymatic analysis of 22 novel human ubiquitin-specific proteases. *Biochemical and Biophysical Research Communications*, 314(1), pp.54–62.

Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11), pp.2281–308.

Ruan, L. et al., 2010. Quantitative characterization of tob interactions provides the thermodynamic basis for translation termination-coupled deadenylase regulation. *Journal of Biological Chemistry*, 285(36), pp.27624–27631.

Safaee, N. et al., 2012. Interdomain Allostery Promotes Assembly of the Poly(A) mRNA Complex with PABP and eIF4G. *Molecular Cell*, 48(3), pp.375–386.

Sakuno, T. et al., 2004. Decapping reaction of mRNA requires Dcp1 in fission yeast: Its characterization in different species from yeast to human. *Journal of Biochemistry*, 136(6), pp.805–812.

Sandler, H. et al., 2011. Not1 mediates recruitment of the deadenylase Caf1 to mRNAs targeted for degradation by tristetraprolin. *Nucleic Acids Research*, 39(10), pp.4373–4386.

Schaeffer, D. et al., 2009. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nature structural & molecular biology*, 16(1), pp.56–62.
Schäfer, I.B. et al., 2014. The structure of the Pan2-Pan3 core complex reveals cross-talk between deadenylase and pseudokinase. *Nature structural & molecular biology*, 21(7), pp.591–8.
Schlundt, A. et al., 2014. Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nature Structural & Molecular Biology*, 21(8), pp.671–678.
Schneider, C. et al., 2009. The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. *Nucleic Acids Research*, 37(4), pp.1127–1140.
Schoenberg, D.R. & Maquat, L.E., 2012. Regulation of cytoplasmic mRNA decay. *Nature reviews. Genetics*, 13(4), pp.246–59.
Schuler, B., 2005. Single-molecule fluorescence spectroscopy of protein folding. *ChemPhysChem*, 6(7), pp.1206–1220.
Schweingruber, C. et al., 2013. Nonsense-mediated mRNA decay — Mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(6), pp.612–623.
Sgromo, A., Raisch, T., Bawankar, P., et al., 2017. A CAF40-binding motif facilitates recruitment of the CCR4-NOT complex to mRNAs targeted by Drosophila Roquin. *Nature Communications*, 8, p.14307.
Sgromo, A., Raisch, T., Backhaus, C., et al., 2017. Drosophila Bag-of-marbles directly interacts with the CAF40 subunit of the CCR4-NOT complex to elicit repression of mRNA targets. *RNA*, p.rna.064584.117.
Sharif, H. et al., 2013. Structural analysis of the yeast Dhh1-Pat1 complex reveals how Dhh1 engages Pat1, Edc3 and RNA in mutually exclusive interactions. *Nucleic Acids Research*, 41(17), pp.8377–8390.
Sharif, H. & Conti, E., 2013. Architecture of the Lsm1-7-Pat1 Complex: A Conserved Assembly in Eukaryotic mRNA Turnover. *Cell Reports*, 5(2), pp.283–291.
She, M. et al., 2008. Structural Basis of Dcp2 Recognition and Activation by Dcp1. *Molecular Cell*, 29(3), pp.337–349.
Shen, B. et al., 1997. Functional analysis of point mutations in human flap endonuclease-1 active site. *Nucleic acids research*, 25(16), pp.3332–8.
Sheth, U., 2003. Decapping and Decay of Messenger RNA Occur in Cytoplasmic Processing Bodies. *Science*, 300(5620), pp.805–808.
Siddiqui, N. et al., 2007. Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein. *The Journal of biological chemistry*, 282(34), pp.25067–75.
Singh, J. & Padgett, R.A., 2009. Rates of in situ transcription and splicing in large human genes. *Nature Structural and Molecular Biology*, 16(11), pp.1128–1133.

Steiger, M. et al., 2003. Analysis of recombinant yeast decapping enzyme. *RNA (New York, N.Y.)*, 9(2), pp.231–8.

Stevens, a, 1980. Purification and characterization of a Saccharomyces cerevisiae exoribonuclease which yields 5'-mononucleotides by a 5' leads to 3' mode of hydrolysis. *The Journal of biological chemistry*, 255(7), pp.3080–5.

Stevens, A., 2001. 5′-Exoribonuclease 1: Xrn1. In pp. 251–259.

Stowell, J.A.W. et al., 2016. Reconstitution of Targeted Deadenylation by the Ccr4-Not Complex and the YTH Domain Protein Mmi1 Correspondence Reconstitution of Targeted Deadenylation by the Ccr4-Not Complex and the YTH Domain Protein Mmi1. *Cell Reports MRC Laboratory of Molecular Biology*, 17.

Stupfler, B. et al., 2016. BTG2 bridges PABPC1 RNA-binding domains and CAF1 deadenylase to control cell proliferation. *Nature communications*, 7, p.10811.

Sun, M. et al., 2013. Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Molecular cell*, 52(1), pp.52–62.

Suzuki, A. et al., 2012. Interaction between NANOS2 and the CCR4-NOT deadenylation complex is essential for male germ cell development in mouse. *PLoS ONE*, 7(3).

Suzuki, A. et al., 2010. NANOS2 interacts with the CCR4-NOT deadenylation complex and leads to suppression of specific RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), pp.3594–3599.

Tan, D. et al., 2014. The ROQ domain of Roquin recognizes mRNA constitutive-decay element and double-stranded RNA. *Nature Structural and Molecular Biology*, 21(8), pp.679–685.

Tarun, S.Z. & Sachs, A.B., 1996. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *The EMBO journal*, 15(24), pp.7168–77.

Temme, C. et al., 2010. Subunits of the Drosophila CCR4-NOT complex and their roles in mRNA deadenylation. *RNA*, 16(7), pp.1356–1370.

Tharun, S. et al., 2005. Mutations in the Saccharomyces cerevisiae LSM1 gene that affect mRNA decapping and 3??? end protection. *Genetics*, 170(1), pp.33–46.

Tharun, S. et al., 2000. Yeast Sm-like proteins function in mRNA decapping and decay. *Nature*, 404(6777), pp.515–8.

Tharun, S. & Parker, R., 2001. Targeting an mRNA for decapping: Displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Molecular Cell*, 8(5), pp.1075–1083.

Tucker, M. et al., 2001. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. *Cell*, 104(3), pp.377–386.

Tyagi, S. & Lemke, E.A., 2013. Genetically Encoded Click Chemistry for Single-Molecule FRET of Proteins. *Methods in Cell Biology*, 113, pp.169–187.

Uchida, N., Hoshino, S.I. & Katada, T., 2004. Identification of a Human Cytoplasmic Poly(A) Nuclease Complex Stimulated by Poly(A)-binding Protein. *Journal of Biological Chemistry*, 279(2), pp.1383–1391.

Ukleja, M. et al., 2016. The architecture of the Schizosaccharomyces pombe CCR4-NOT complex. *Nature Communications*, 7, p.10433.

Valkov, E. et al., 2016. Structure of the Dcp2–Dcp1 mRNA-decapping complex in the activated conformation. *Nature Structural & Molecular Biology*, 23(6), pp.574–579.

Vasudevan, S., Peltz, S.W. & Wilusz, C.J., 2002. Non-stop decay?a new mRNA surveillance pathway. *BioEssays*, 24(9), pp.785–788.

Villalba, A., Coll, O. & Gebauer, F., 2011. Cytoplasmic polyadenylation and translational control. *Current Opinion in Genetics and Development*, 21(4), pp.452–457.

Viswanathan, P. et al., 2004. Mouse CAF1 can function as a processive deadenylase/3???-5??? -exonuclease in vitro but in yeast the deadenylase function of CAF1 is not required for mRNA poly(A) removal. *Journal of Biological Chemistry*, 279(23), pp.23988–23995.

Wahle, E. & Winkler, G.S., 2013. RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. *Biochimica et biophysica acta*, 1829(6–7), pp.561–570.

Wang, H. et al., 2010. Crystal structure of the human CNOT6L nuclease domain reveals strict poly(A) substrate specificity. *The EMBO journal*, 29(15), pp.2566–76.

Wang, X. et al., 2002. Modular recognition of RNA by a human Pumilio-homology domain. *Cell*, 110(4), pp.501–512.

Wang, Z. et al., 2002. The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proceedings of the National Academy of Sciences*, 99(20), pp.12663–12668.

Wasmuth, E. V. & Lima, C.D., 2012. Exo- and Endoribonucleolytic Activities of Yeast Cytoplasmic and Nuclear RNA Exosomes Are Dependent on the Noncatalytic Core and Central Channel. *Molecular Cell*, 48(1), pp.133–144.

Weidmann, C.A. et al., 2016. Drosophila Nanos acts as a molecular clamp that modulates the RNA-binding and repression activities of Pumilio. *eLife*, 5.

Weill, L. et al., 2012. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nature Structural & Molecular Biology*, 19(6), pp.577–585.

Wells, S.E. et al., 1998. Circularization of mRNA by Eukaryotic Translation Initiation Factors. *Molecular Cell*, 2(1), pp.135–140.

Will, C.L. & Lührmann, R., 2011. Spliceosome structure and function. *Cold*

*Spring Harbor Perspectives in Biology*, 3(7), pp.1–2.
Winkler, G.S. & Balacco, D.L., 2013. Heterogeneity and complexity within the nuclease module of the Ccr4-Not complex. *Frontiers in genetics*, 4, p.296.
Wolf, J. et al., 2014. Structural basis for Pan3 binding to Pan2 and its function in mRNA recruitment and deadenylation. *The EMBO Journal*, 33(14), pp.1514–1526.
Wong, S.Q. et al., 2015. Whole exome sequencing identifies a recurrent RQCD1 P131L mutation in cutaneous melanoma. *Oncotarget*, 6(2), pp.1115–27.
Woychik, N.A. & Hampsey, M., 2002. The RNA polymerase II machinery: Structure illuminates function. *Cell*, 108(4), pp.453–463.
Wu, D. et al., 2014. Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. *Cell research*, 24(2), pp.233–46.
Yamashita, A. et al., 2005. Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nature Structural and Molecular Biology*, 12(12), pp.1054–1063.
Yang, X. et al., 2008. Crystal structures of human BTG2 and mouse TIS21 involved in suppression of CAF1 deadenylase activity. *Nucleic Acids Research*, 36(21), pp.6872–6881.
Young, T.S. et al., 2010. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *Journal of Molecular Biology*, 395(2), pp.361–374.
Yudkovsky, N., Ranish, J.A. & Hahn, S., 2000. A transcription reinitiation intermediate that is stabilized by activator. *Nature*, 408(6809), pp.225–229.
Zhao, J., Hyman, L. & Moore, C., 1999. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and molecular biology reviews : MMBR*, 63(2), pp.405–445.
Zheng, D. et al., 2008. Deadenylation is prerequisite for P-body formation and mRNA decay in mammalian cells. *Journal of Cell Biology*, 182(1), pp.89–101.
Zhou, M. et al., 2015. DcpS is a transcript-specific modulator of RNA in mammalian cells. *RNA (New York, N.Y.)*, 21(7), pp.1306–12.
Zuo, Y. & Deutscher, M.P., 2001. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic acids research*, 29(5), pp.1017–26.
Zwartjes, C.G.M. et al., 2004. Repression of Promoter Activity by CNOT2, a Subunit of the Transcription Regulatory Ccr4-Not Complex. *Journal of Biological Chemistry*, 279(12), pp.10848–10854.

# Appendix 1

## Investigation of the BTG2-PABP interaction

# Appendix 1. Investigating BTG2-PABPC1 interaction via BoxC domain of BTG2

Recently, the BoxC of BTG2 has been shown to be key for the interaction between BTG2 and PABPC1 (Stupfler et al. 2016). In this study, the aim was to further characterise this interaction by identifying the single residues that are mainly engaged when these proteins bind. As a first step, Caf1-BTG2 and nuclease module complexes (wild type and inactive mutant) diplaying an altered BoxC in BTG2, were expressed and purified, to compare their activity with the complexes including wild type BTG2 in various deadenylation assays formats. To impair the interaction with PABPC1, the sequence of the BTG2 BoxC (residues 116-120) was changed from DGSIC to the BoxC of Tob1 (KGPVK), which is not interacting with PABPC1. To introduce the change, reverse PCR was performed using P02 plasmid as template, and correct cloning of P02C plasmid was confirmed by sequencing (Figure A1.1).

Initial assays to compare the wild type complexes with the ones including BTG2 with the altered BoxC have lead to inconclusive results. Further assays, perhaps following bioinformatics docking simulations, will be needed to investigate which single residues of BTG2 and PABPC1 are involved in this interaction.

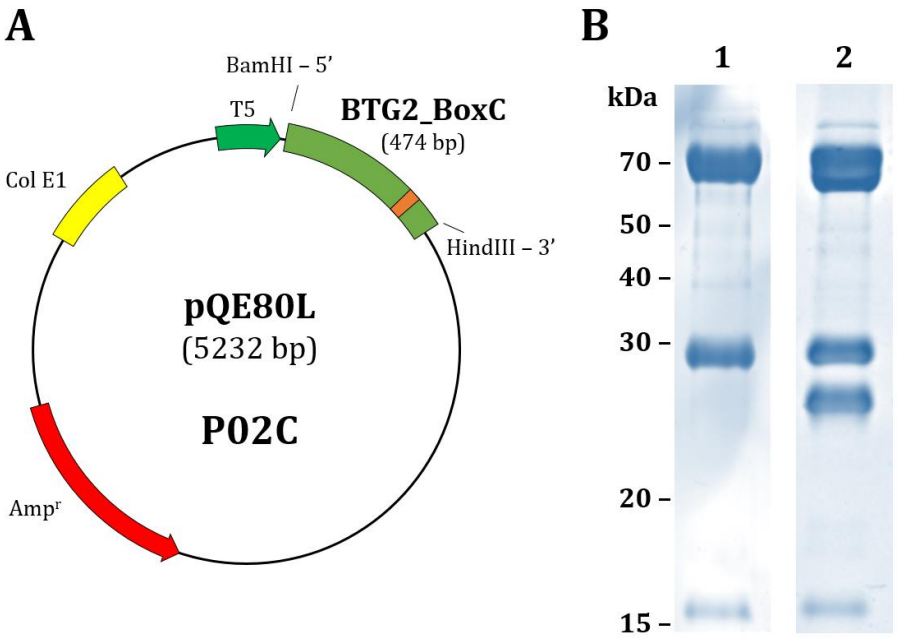



**Figure A1.1** – Purified complexes including BTG2 showing an altered BoxC (in orange, amino acid sequence KGPVK instead of the wild type DGSIC) expressed with plasmid P02C (A). Both the Nuclease module (lane 1) and the Ccr4-Not core complex (lane 2) were purified following the protocols described in Chapter 4 for the wild type and inactive mutant complexes (B).

# Appendix 2

## Characterisation of deadenylation complexes in human cells by immunocytochemistry

# Appendix 2. Characterisation of deadenylation complexes in human cells by immunocytochemistry

RNA processing bodies, or P-bodies, are dynamic foci in the cytoplasm where the mRNA decay process can occur. It has been established that deadenylation is a triggering event for P-bodies formation and, consequently, mRNA decay in eukaryotic cells (Zheng et al. 2008). Moreover, Ccr4 shows co-localisation with P-bodies, and its knock down is associated with their reduced number; however, it is not clear whether this is due to the impaired deadenylation process or to a potential structural function within the P-bodies (Cougot et al. 2004). Reduced deadenylation impairs P-bodies formation, but not the opposite: when PAN3, but not PAN2, is knocked down, P-bodies formation is strongly reduced, with only a minor effect on deadenylation (Zheng et al. 2008). PAN3 has been established to recruit PAN2, Ccr4 and Caf1 to P-bodies, therefore playing a major structural role in these foci (Zheng et al. 2008). Deadenylation is not only enhancing P-bodies formation because of its role in blocking translation initiation; in fact, a highly impacting consequence of the degradation of the poly(A) tail is the release of PABP, which could prevent the formation of the P-bodies (Chen 2013).

With P-bodies being so closely linked with deadenylation and yet several mechanisms being still uncovered, the aim of this study was to investigate on the dualistic functional and structural role of enzymes and complexes responsible for deadenylation in human cells. To do so, several cell lines displaying genomic modifications engineered with CRISPR/Cas9 technology were conceived. On one hand, several combined knock-outs of regulatory and/or catalytic enzymes were planned, with additional experiments involving rescuing of deadenylation activity by exogenous over-expression. Additionally, single point mutations could be introduced to generate cell lines expressing inactive enzymes.

The cell lines would then be analysed by immunocytochemistry, to assess the impact of the genomic modifications on P-bodies formation. To evaluate the feasibility of this approach, the effect of wild type and inactive mutant Caf1 over-expression was tested in HEK293 cells. Caf1 and Caf1 D40A CDS were cloned into pCMV5 vectors containing a FLAG tag at the 5' end of the MCS. Cells were transfected following the calcium phosphate protocol described and prepared for immunofluorescent staining. Recently, it was showed that DDX6 is required for P-bodies assembly (Ayache et al. 2015); in order to detect P-bodies, an antibody targeting the RNA helicase DDX6 was selected (Table A2.1)

The average P-bodies assembled when wild type Caf1 was over-expressed was three, whereas when Caf1 D40A was over-expressed, only one P-body per cell was formed on average (Figure A2.1 and A2.2). These results indicate that impairing deadenylation by over-expression of inactive mutant Caf1 has a strong effect on P-bodies formation under these conditions.

An additional purpose for this assay is to test the effect of small drug-like molecules targeting the Caf1 deadenylase enzyme (Maryati et al. 2014; Jadhav et al. 2015). Caf1 over-expression is associated with high metastatic events in mouse (Faraji et al. 2016), making this enzyme a potential target for novel anti-cancer drug discovery.

## A2.1 Immunofluorescence microscopy

A2.1.2 Reagents, stock solutions and buffers for use in immunofluorescence

4% paraformaldehyde (PFA): 4% PFS in $H_2O$ w/v, stored at -20°C.

0.5% Triton X-100: 0.5% Triton X-100 in PBS, stored at 4°C

3% Bovine serum albumin (BSA): 3% BSA in PBS, stored at 4°C

10% Bovine serum albumin (BSA): 10% BSA in PBS, stored at 4°C

Mounting medium: 90% glycerol, 10% PBS

Phosphate buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4$, 1.47 mM $KH_2PO_4$ (1x PBS without $Ca^{2+}$, $Mg^{2+}$; PAA, H15-002)

A2.2.2 Preparation of mammalian cells for immunofluorescence

Round glass cover slips (22 mm diameter, 0.15 mm thick) were sterilised with 70% ethanol and passed through a flame. Working in a class II safety cabinet, the cover slips were placed into 6-wells culture plates. HEK293 cells were seeded (250,000 cells per well) in 2 ml of DMEM onto the coverslips. 24 h after seeding, cells were transfected with the Calcium Phosphate method as described before. 16 h after transfection, cells were washed twice with complete medium. 24 h after transfection, the cells were fixed with 1 ml 4% paraformaldehyde (4% PFA) and incubated at room temperature for 10 min. The cells were then washed five times with 1 ml of PBS. After fixing, coverslips were incubated with 1 ml 0.2% Triton X-100 for 2 min at room temperature (permeabilisation) and then washed three times with 1 ml of PBS. For the blocking step, 1 ml of 3% bovine serum albumin (BSA) in PBS were added and cells were incubated for 3 h at room temperature. Coverslips were incubated with the primary antibody solution (3% BSA in PBS) at the appropriate dilution (Table 2.11). 200 µl of

antibody solution was added onto the coverslips inside the 6-wells plate and incubated for 1 h at room temperature. Coverslips were washed three times with 1 ml PBS for 5 min. Coverslips were then incubated with 200 µl of secondary antibody solution (3% BSA in PBS) for 30 min at room temperature. Coverslips were washed three times for 5 min with 1 ml PBS. To stain DNA, 1 ml of Hoechst stain (0.5 µg/ml in PBS, or Draq5 diluted 1:5000 in PBS) was added to each coverslip and incubated in the dark at room temperature for 5 min. Coverslips were shielded from the light. After removing the Hoechst stain, the coverslips were washed three times with 1 ml PBS. Using a scalpel, coverslips were carefully removed from the 6-wells plate and mounted facing down onto standard glass microscope slides with 10 µl of mounting medium. The edges of the coverslips were sealed with clear nail varnish and slides were kept in the dark at 4°C. Slides were imaged using a Zeiss LSM 510 Meta microscope and HeNe laser for excitation at 594 nm. For each slide, three representative Z-stack images were taken. Cells were analysed and imaged using a Zeiss LSM510 Meta confocal microscope. Three different lasers were used, with wavelengths of 488, 543 and 633nm. For each coverslip, three distinct Z-stack images were taken.

| Antibody | Target | Supplier | Cat # | Dilution |
|---|---|---|---|---|
| Primary | Anti-Flag (mouse) | Sigma Aldrich | F1804 – 50UG | 1:250 |
| | Anti-DDX6 (rabbit) | Bethyl Labs | A300-461A | 1:500 |
| Secondary | Alexa 594 (anti-mouse) | Invitrogen | A21441 | 1:500 |
| | Alexa 488 (anti-rabbit) | Invitrogen | A21203 | 1:500 |

**Table A2.1** - List of the antibodies used in the immunofluorescence microscopy experiments.

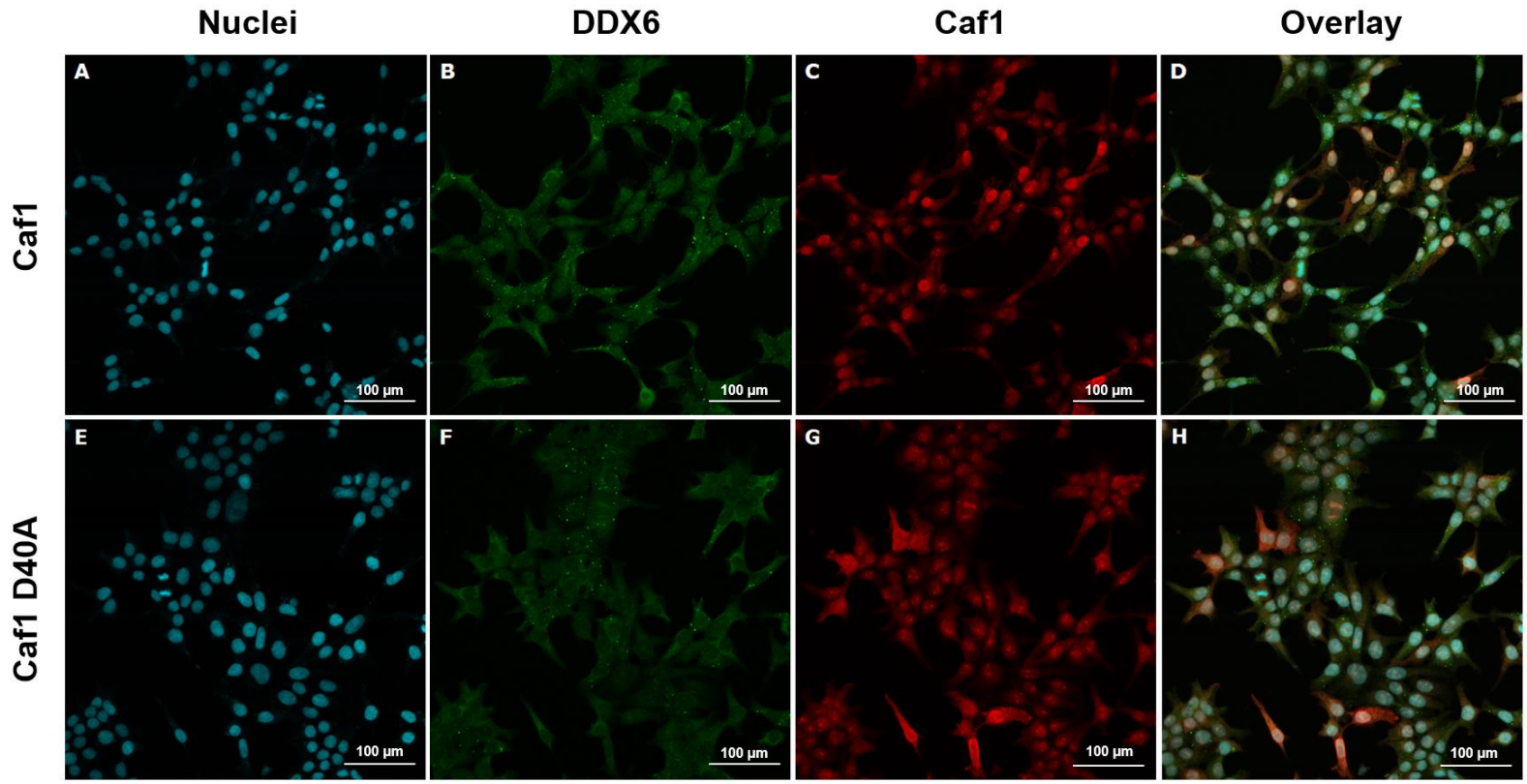



**Figure A2.2** – Representative panels comparing HEK293 cells over-expressing wild type Caf1 (A-D) or the inactive mutant D40A (E-H). Similar expression levels of Caf1 were observed (C and G) but P-bodies formation was reduced with inactive Caf1 (B and F). P-bodies were detected targeting DDX6, a RNA helicase required for their assembly.

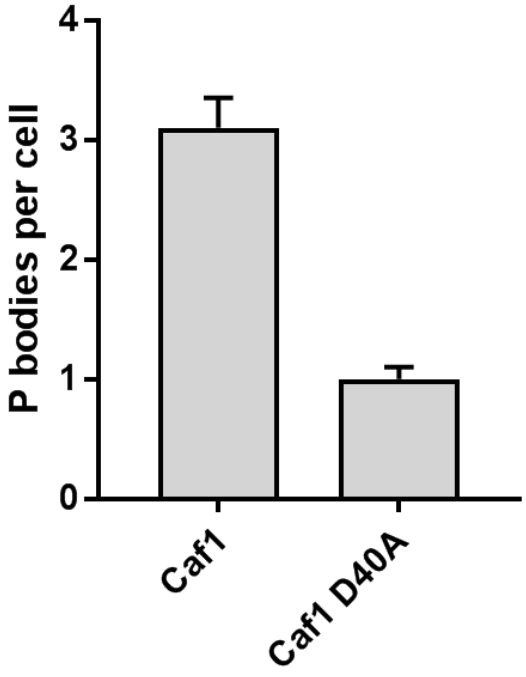



**Figure A2.3** – Analysis of the P-bodies formation in HEK293 cells over-expressing Caf wild type or D40A inactive mutant (ns = 3).

# Appendix 3

## Crystallisation of CNOT9 P131L

## Appendix 3. Crystallisation of CNOT9 P131L

Recently, an interesting CNOT9 P131L variant has been detected with 4% prevalence in a whole-exome sequencing study of primary cutaneous melanoma patients (Wong et al. 2015). The structural or biochemical impact of this amino acid substitution is still unclear.

To try to characterise this mutation from a structural point of view, CNOT9 P131L variant was expressed and purified. In addition, two complexes were expressed and purified, including either CNOT1M (MIF4G and DUF3819 domains) or the DUF3819 domain, and CNOT9 (P131L). TEV protease digestions were performed on the purified protein/complexes to remove the 6His tag and improve crystal packing.

A first crystallisation buffer screening was performed on the CNOT9 P131L protein and the CNOT1M-CNOT9 (P131L) complex, which included two commercially available screens (Morpheus and ProPlex, Molecular Dimensions), by 2-drops 96-wells plates (Molecular dimensions, MD11-00-10/40/100). CNOT9 P131L was screened at 4 and 8 mg/ml; DUF3819-CNOT9 (P131L) was screened at 4.8 and 9.7 mg/ml; CNOT1M-CNOT9 (P131L) was screened at 8.3 and 16.7 mg/ml. Plates were sealed and incubated for 3-4 weeks at 20°C.

After two weeks, one hit was identified in the ProPlex screen (condition F11), similar to the crystallisation conditions published for CNOT9 wild type (Garces et al. 2007). The initial hit (0.1 M Sodium acetate pH 5.0 and 1.5 M Ammonium sulfate) was further investigated in a 24-well optimisation plate (Molecular Dimensions, MD3-11) (Table A3.1). Several plate-shaped crystals were observed across the plate, and five of the best samples were fished and sent to the Diamond Light Source for X-ray analysis (Fig A3.1).

| Sodium acetate 0.1 M | Ammonium sulfate (M) | | | | | |
|---|---|---|---|---|---|---|
| | **1.3** | **1.4** | **1.5** | **1.6** | **1.7** | **1.8** |
| **pH 4.9** | A1 | A2 | A3 | A4 | A5 | A6 |
| **pH 5.0** | B1 | B2 | B3 | B4 | B5 | B6 |
| **pH 5.1** | C1 | C2 | C3 | C4 | C5 | C6 |
| **pH 5.2** | D1 | D2 | D3 | D4 | D5 | D6 |

**Table A3.1** - Optimisation plate set up after 'hit' in the screening plate. Proteins precipitated and formed crystals in several of the conditions above, recurrently at pH 5 (B2-B5). Crystals were grown for three weeks and fished for X-ray crystallography at the Diamond Light Source.



**Figure A3.1** – Plate crystals obtained from the optimisation step (condition B4). When fishing, crystal plates were separated and individually collected for X-ray analysis.

# Appendix 4

## Structural characterisation of the human Ccr4-Not complex by Electron Paramagnetic Resonance

# Appendix 4. Structural characterisation of the human Ccr4-Not complex by Electron Paramagnetic Resonance

EPR spectroscopy is a potential tool to further investigate the relative positions of Caf1 and Ccr4, the two active deadenylase enzymes of the nuclease module of the Ccr4-Not complex, and whether these change upon RNA binding and during the catalytic cycle,

In order to determine the distance between the catalytic centres, the nuclease module (containing Caf1, BTG2 and Ccr4) was purified as described in Chapter 4. At the end of the purification procedure, the buffer was exchanged to GF buffer prepared with $D_2O$, to reduce background signal during EPR spectroscopy, and without glycerol. After protein concentration, an equimolar amount of $MnCl_2$ and glycerol-d8 (20-30% final concentration) were added.

Samples were analysed by EPR spectroscopy by Dr Alexey Potapov, Sir Peter Mansfield Imaging Centre, School of Physics, University Of Nottingham.

# Appendix 5

## Incorporation of unnatural amino acids in the deadenylation enzymes of the Ccr4-Not complex

# Appendix 5. Incorporation of unnatural amino acids in the deadenylation enzymes of the Ccr4-Not complex

To evaluate whether the incorporation of non-natural amino acids can be carried out efficiently, the following experiment was designed. Codon at position 34 (TAA) in the Caf1 gene (vector P06 as template) was mutated into an amber codon (TAG) by reverse PCR. The inserted mutation (vector P06A) was confirmed by sequencing. *E. coli* BL21 DE3 competent cells were co-transformed with P06A and pEVOL-pAzF (Addgene #31186) vectors, for Caf1 (N-terminal amber) and tRNA/tRNA synthetase expression (Young et al. 2010), respectively. Bacterial cultures (200 ml) containing appropriate antibiotics were grown at 37°C and 200 rpm until they reached an $OD_{600}$ of 0.8. Then, they were incubated at 18°C and IPTG (0.2 mM) and p-Azido-L-phenylalanine (0.5 mM, Chem-Impex #06162) were added. Samples were taken after 3 h and 16 h to verify Caf1 expression (Fig. A5.1).

In presence of both the tRNA/tRNA synthetase pair and 0.5 mM p-azido-L-phenylalanine, recombinant Caf1 was expressed at similar levels to Caf1 wild-type, indicating the incorporation of the unnatural aminoacid in position 34. This proof of principle shows that unnatural amino acids can be incorporated in protein expressed in *E. coli* and therefore this technique can be used to measure potential conformational changes in the nuclease module of the Ccr4-Not complex by single-molecule FRET (Schuler 2005).

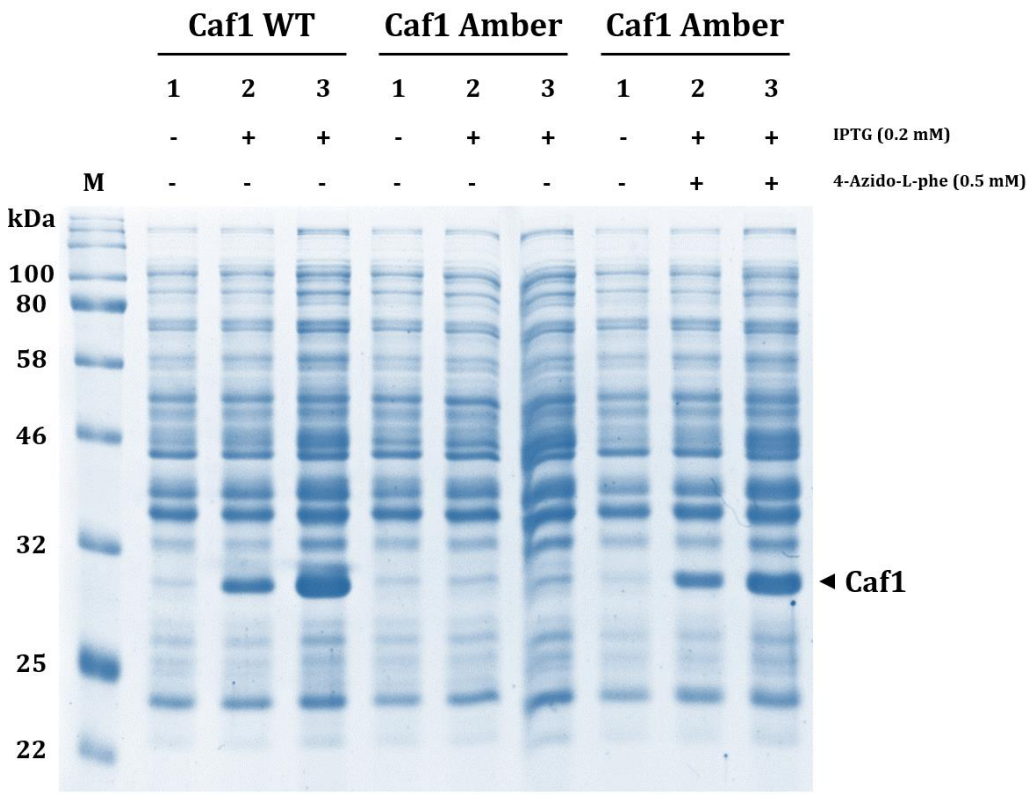



**Figure A5.1** – Analysis of Caf1 (N-term Amber codon) expression using 12% SDS-PAGE. Proteins were visualised by Coomassie staining. Samples were taken from bacterial cultures before IPTG addition (lane 1), 3 h (lane 2) and 16 h (lane 3) after IPTG addition. Unnatural amino acid 4-azido-L-phenylalanine was added immediately after IPTG.