Étude fonctionnelle des sous-domaines de Pcf11 : rôle du 2nd NTD dans la terminaison de transcription des snoRNAs et des motifs liant le zinc dans les activités de maturation de l'extrémité 3' des ARN messagers.

Julia Guéguéniat

▶ **To cite this version:**

Julia Guéguéniat. Étude fonctionnelle des sous-domaines de Pcf11 : rôle du 2nd NTD dans la terminaison de transcription des snoRNAs et des motifs liant le zinc dans les activités de maturation de l'extrémité 3' des ARN messagers.. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université de Bordeaux, 2015. Français. NNT : 2015BORD0236 . tel-01552297v2

HAL Id: tel-01552297

https://theses.hal.science/tel-01552297v2

Submitted on 3 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE SCIENCES DE LA VIE ET DE LA SANTÉ

SPÉCIALITÉ BIOCHIMIE

Par Julia GUÉGUÉNIAT

ÉTUDE FONCTIONNELLE DES SOUS-DOMAINES DE PCF11 :

Rôle du 2nd NTD dans la terminaison de transcription des snoRNAs et des motifs liant le zinc dans les activités de maturation de l'extrémité 3' des ARN messagers.

Sous la direction de : Lionel MINVIELLE-SÉBASTIA

Soutenue publiquement le 03 Décembre 2015

Membres du jury :

M. CULLIN Christophe	Professeur des Universités – Univ. de Bordeaux	Président
Mme. MILLEVOI Stéfania	Chargée de Recherche – INSERM	Rapporteur
M. LIBRI Domenico	Directeur de Recherche – CNRS	Rapporteur
Mme LENGRONNE Armelle	Chargée de Recherche – CNRS	Examinateur
M. MCCUSKER Derek	Chargé de Recherche – CNRS	Examinateur
M. CAVAILLÉ Jérôme	Directeur de Recherche – CNRS	Examinateur

À M. P. pour sa sagacité
À mes parents pour leur soutien
À mon frère « au fond c'était mon meilleur ami, il me soutenait, voyez-vous »
Aux copains, les meilleurs amis que l'on puisse imaginer
À (certains) des autres thésards
À Bérénice

« De douces et fraternelles amitiés m'entouraient déjà de sollicitudes et de dévouements dont je ne méconnaissais pas le prix » G. Sand, Histoire de ma vie

REMERCIEMENTS

En premier lieu, je tiens à remercier le Docteur Lionel Minvielle-Sébastia pour m'avoir accueillie dans son laboratoire, encadrée et fait confiance tout au long de ce travail de thèse. Je lui suis également reconnaissante pour ses conseils, pour les discussions scientifiques surtout sur la terminaison de transcription, pour avoir partager toutes ses connaissances et pour le temps qu'il m'a accordé pour la correction de ce manuscrit.

Je tiens à remercier le Docteur Christophe Cullin d'avoir accepté de présider ce jury de thèse. Je tiens également à remercier les Docteurs Stéfania Millevoi et Domenico Libri d'avoir accepté d'étudier et de juger ce travail de thèse en tant que rapporteurs. Enfin je remercie les Docteurs Armelle Lengronne, Derek McCusker et Jérôme Cavaillé d'avoir accepté d'examiner et de participer à ce jury de thèse.

J'aimerais remercier le Docteur Sébastien Fribourg, pour m'avoir accueillie dans son laboratoire pendant la première partie de cette thèse en ne cessant de me prodiguer conseils et encouragements, pour les discussions scientifiques ainsi que son oeil critique. Je souhaiterais également remercier le Docteur Cameron Mackereth, pour toute la partie RMN de ce projet, sa gentillesse et son expertise scientifique.

J'exprime toute ma gratitude à Thierry Noël, super/meilleur directeur du monde du super/meilleur Master, pour rire à mes blagues, pour me faire avec autant de gentillesse que de compétence, un grand nombre de suggestions très attentivement élaborées, pour sa disponibilité et sa sollicitude.

Un grand merci, mille énormes mercis à « THE BigBoss » la chef-de-tous-les instants qui m'a appris à aimé les membranes lipidiques, à me poser (parfois) les bonnes questions et à m'extasier devant un Wb, pour ta disponibilité, les pauses café/clopes et enfin surtout pour être aussi géniale ! « Un chef génial, capable par miracle d'un coup d'œil d'ensemble » Et c'est St. Exupéry qui le dit.

Avant de continuer, et afin de n'oublier personne, je tiens à remercier toutes les personnes que j'ai pu côtoyer durant ces trois années entre l'IECB et Carreire. Cette liste n'est pas exhaustive ; si vous n'êtes pas cité, ce n'est pas que je vous aie oublié ; je vous suis reconnaissante pour tout ce que vous m'avez apporté.

À l'IECB, un immense merci à Natacha, Cécile, Sabrina et Aurélie pour votre soutien, à Adrien pour ton accueil, tes blagues et tout le reste, à Amandine, à Lionel B., à Amélie, Britta, Caro, Leyla, Oscar et Paolo, à Mathilde & William (« *es un bato la ba* » me fera toujours rire), à Denis Super-chef et à Heddy. Merci à Jonathan et Ulf-Peter G, de me prouver que la route est moins difficile quand on est accompagnée (et aussi pour m'avoir fait découvrir toutes ces vidéos que je n'aurais jamais souhaité connaître, la Jean Lain en cannette, et ta sincère gentillesse). Merci également à Camimi, de me prouver que la route est moins difficile quand on est accompagnée et/ou quand on est quatre, mais aussi pour m'avoir fait manger n'importe quoi, ta bonne humeur, tes conseils et ta gentillesse. Merci enfin à Éric, de me prouver que la route est moins difficile quand on est accompagnée, pour être aussi drôle, de bons conseils, et pour n'avoir aucun filtre.

Enfin Carreire… Je crains qu'il ne soit trop long d'exprimer toute ma gratitude surtout en cette fin d'année. Un immense merci pour votre bienveillance, les blagues du midi, et ce côté famille, indescriptible. À Sandrine, pour rire de mes malheurs et TOUT le reste, à Kati pour TOUT, à Sonia et Laetitia pour vos conseils, Guillaume & Olivier (je l'admets, sans vous cela n'aurait pas été aussi savoureux), Éric, Fabien, Laurent & SuperNico, Samantha, Isa, Cathy, Aref, les nouvelles têtes, Lorena, Khalida et Sara (oula) et aux deux Michelles (vous êtes les meilleures M2, irremplaçables). Un merci tout particulier à Claire qui m'a fourni d'heureuses suggestions, pour ses conseils, sa relecture et

son avis de pro. Je n'oublie pas « les gens d'à côté » pour leur gentillesse et pour me rappeler à intervalle régulier que le labo n'est pas ma maison !

J'aimerais tout particulièrement remercier deux personnes importantes. Émilie, pour les soirées jusqu'à des heures indécentes, pour son soutien, ses conseils et sa disponibilité, sa gaine et les petits veuxche, pour avoir été là même quand j'ai mauvais caractère, et pour tout ce qui arrive !
Hélène Chat, ma super voisine, pour avoir été toujours là, pour m'avoir appris à attendre après avoir mis du vernis, pour être mon meilleur public (le Spa Nuxe restera), pour tes conseils et nos grands débats philosophiques et pour avoir pris soin de moi tout au long de ces trois ans.

J'adresse également toute ma gratitude et mon amour aux extérieurs du labo. À Sisley, même quand c'était mort, qu'il n'y avait pas de swell et que le vent soufflait offshore tu étais là, avec tes cheveux blonds. À Lulu, soutien sans faille. À Kenza (ô ma Kenza). À SuperMamar, Jeannette & Max, Camillha & Karine, Alice et Alix pour être aussi mignonnes, Floriane pour tes douces pensées, Baye, LoukLouk, Vincent & Pihi, Thomas, Thomas, Dima, Fred, Félix (oui, c'est toujours un plaisir les bières avec toi) et tout ceux non cités.

J'adresse tout mes remerciements et mon amour aux « copains » pour m'avoir soutenu depuis toutes ces années dans cette chasse incessante pour les attraper tous, ils savent que ça demande du courage. C'est grâce à vous que je fais partie des meilleures dresseuses.

Pour terminer, j'aimerais remercier ma famille ; mes oncles et tantes Bordelais, Brestois, St-Thonnantois, Toulonnais et Malgache ; mes Grands-mères et mes Grands-pères. Et surtout mes parents. Mon père pour son soutien, les nombreuses portes ouvertes enfoncées par ses soins, son calme, son humour et sa confiance absolue. Ma mère, pour son soutien rassurant, son écoute et sa prévenance. C'est V. Hugo qui en parle le mieux « Ô l'amour d'une mère ! Amour que nul n'oublie ! Pain merveilleux qu'un Dieu partage et multiplie ».

Mille tendres mercis à mon frère Pierre, pour ses appels quotidiens, son aide durant cette rédaction, sa relecture littéraire, sa patience et ses mots de réconfort, et pour être capable de m'énerver en moins de 30 sec. Ta présence et ton influence sont inestimables. « Un frère est un ami donné par la nature ».

TABLE DES MATIERES

LISTE DES ABREVIATIONS 3

LISTE DES FIGURES & TABLEAUX 6

PREFACE 10

INTRODUCTION 11

A. MATURATION DES PRECURSEURS DES ARNM 11
A.1. GENERALITES 11
A.1.1. ARN POLYMERASE II 11
A.1.2. MATURATION DE L'EXTREMITE 5' 12
A.1.3. ÉPISSAGE 14
A.1.4. EXPORT 15
A.2. MATURATION DE L'EXTREMITE 3' 18
A.3. LES AUTRES MECANISMES DE MATURATION DE L'EXTREMITE 3' 19
A.3.1. ARNM CODANTS POUR LES HISTONES 19
A.3.2. MATURATION DES SNORNAS ET SNRNAS 21
A.4. POLYADENYLATION ALTERNATIVE ET SELECTION DU SITE DE POLYA 23
A.5. LA POLYADENYLATION CHEZ LES PROCARYOTES 26
A.6. TRAMP ET POLYADENYLATION DE DEGRADATION 28
A.7. LES MALADIES ASSOCIEES 30
B. COMPOSITION DU COMPLEXE DE POLYADENYLATION 31
B.1. *SACCHAROMYCES CEREVISIAE*. 31
B.1.1. ÉLEMENTS *CIS* 31
B.1.2. FACTEURS *TRANS* 32
B.1.3. CTD DE L'ARN POLYMERASE II 45
B.2. COMPARAISON SYSTEME LEVURE ET MAMMIFERES 48
B.2.1. MACHINERIE MAMMIFERES 48
B.2.2. COMPARAISON 50
C. TERMINAISON TRANSCRIPTION 51
C.1. TERMINAISON DES ARNS MESSAGERS CHEZ LA LEVURE 52
C.1.1. MECANISME GENERAL 52
C.1.2. ALLOSTERIQUE – CLIVAGE INDEPENDANT 53
C.1.3. TORPEDO – CLIVAGE DEPENDANT 54
C.1.4. ALLOSTERIQUE/TORPEDO 55
C.2. TERMINAISON DES ARNS NON CODANTS CHEZ LA LEVURE 55
C.2.1. MECANISME DEPENDANT DU NNS 56
C.2.2. AUTRES MECANISMES DE TERMINAISON DES ARNNC CHEZ LA LEVURE 58
C.3. AUTRES MECANISMES DE TERMINAISON DE TRANSCRIPTION CHEZ *S. CEREVISIAE* 58
C.4. INTERACTION FONCTIONNELLE ENTRE LES DEUX MECANISMES 59
C.5. TERMINAISON DE LA TRANSCRIPTION CHEZ LES METAZOAIRES 61
C.6. TERMINAISON DE TRANSCRIPTION ET REPARATION DE L'ADN 61
D. MODIFICATIONS POST-TRADUCTIONNELLES DE LA MACHINERIE DE POLYADENYLATION 63
D.1. PHOSPHORYLATION 63
D.2. UBIQUITINATION 65
D.3. SUMOYLATION 66
D.4. ACETYLATION / METHYLATION 67

MATERIELS ET METHODES 68

A. LEVURES : SOUCHES UTILISEES, VECTEURS ET MILIEUX DE CULTURE 68
A.1. SOUCHES UTILISEES 68
A.2. VECTEURS 68
A.3. MILIEUX DE CULTURES 68
B. BACTERIES : SOUCHES UTILISEES, VECTEURS ET MILIEUX DE CULTURE 69
B.1. SOUCHES 69
B.2. VECTEURS 69
B.3. MILIEUX DE CULTURES 69
C. TECHNIQUES DE BIOLOGIE MOLECULAIRE 69
C.1. TRANSFORMATION DES BACTERIES 69
C.1.1. PREPARATION DES BACTERIES ELECTRO-COMPETENTES 69
C.1.2. TRANSFORMATION DES BACTERIES ELECTRO-COMPETENTES 70
C.2. TRANSFORMATION DES LEVURES 70
C.3. ISOLEMENT DES ACIDES NUCLEIQUES 70
C.3.1. PREPARATION D'ADN PLASMIDIQUE A PARTIR DES BACTERIES 70
C.3.2. PREPARATION D'ADN PLASMIDIQUE A PARTIR DE LEVURES 71
C.3.3. PREPARATION D'ARNS TOTAUX DE LEVURES 71
C.4. DOSAGE DES ACIDES NUCLEIQUES 72
C.5. DIGESTION DES PLASMIDES 72
C.6. AMPLIFICATION PAR PCR 72
C.7. CARACTERISATION DU PHENOTYPE 73
C.7.1. TEST EN GOUTTE / THERMOSENSIBILITE 73
C.7.2. TEST DE CROISSANCE 73
D. TECHNIQUES DE GENETIQUES 73
D.1. PRODUCTION DE DIPLOÏDES 73
D.2. DISSECTION D'ASQUES 74
D.3. STRATEGIE DE « GAP REPAIRED » 74
D.4. PLASMIDE SHUFFLING 74
E. TECHNIQUES DE BIOCHIMIE DES PROTEINES 75
E.1. PRODUCTION DE PROTEINES RECOMBINANTES 75
E.1.1. PREPARATION D'EXTRAITS PROTEIQUES 75
E.1.2. PURIFICATION GSH-SEPHAROSE 75
E.1.3. PURIFICATION NI-NTA AGAROSE 76
E.2. PREPARATION D'EXTRAITS PROTEIQUES DE LEVURES 77
E.2.1. EXTRACTION RAPIDE 77
E.2.2. PREPARATION D'EXTRAITS PROTEIQUES DITS « 0-40 » 77
E.3. PURIFICATION DES FACTEURS PROTEIQUES PAR *TAP-TAGGING* 78
E.4. WESTERN BLOT 79
E.5. TEST D'ACTIVITE KINASE 79
E.5.1. ACTIVITE ASSOCIEE AUX FACTEURS 79
E.5.2. PHOSPHORYLATION CTD PAR LA MAP KINASE 80
E.6. TEST D'ACTIVITE D'HYDROLYSE DE L'ATP 80
E.7. TEST D'INTERACTION CF IA 81
E.7.1. GST-CTD 81
E.7.2. GST-YRA1 81
F. TECHNIQUES DE BIOCHIMIE DES ARNS 81
F.1. PLASMIDES 81
F.2. PRODUCTION D'ARNS *IN VITRO* 82
F.2.1. TRANSCRIPTION D'ARNS *IN VITRO* 82
F.2.2. TRANSCRIPTION *IN VITRO* D'ARNS RADIOMARQUES 82
F.3. REACTION DE RECONSTITUTION DES ACTIVITES DE MATURATION EN 3' DES ARN MESSAGERS 84
F.4. NORTHERN BLOT 85
F.4.1. PREPARATION ET MIGRATION DES ECHANTILLONS 85
F.4.2. TRANSFERT SUR MEMBRANE DE NYLON 85

F.4.3. PREPARATION DES SONDES 85
F.4.4. HYBRIDATION ET REVELATION 86
F.5. ESSAI DE PROTECTION CONTRE LA DIGESTION PAR LA S1 NUCLEASE 86
F.5.1. PREPARATION DES MARQUEURS DE TAILLES 86
F.5.2. PREPARATION DES SONDES 86
F.5.3. ESSAI DE PROTECTION CONTRE LA DIGESTION PAR LA S1 NUCLEASE 87

CHAPITRE 1 : ETUDE FONCTIONNELLE DES SOUS-DOMAINES DE PCF11 88

A. RAPPEL SUR PCF11 88
A.1. REGION N-TERMINALE 88
A.2. REGION C-TERMINALE 90
B. OBJECTIFS DE L'ETUDE 92
C. ÉTUDE DU 2ND NTD DE PCF11 93
C.1. RESULTATS 93
C.1.1. OBTENTION ET CARACTERISATION DES MUTANTS 93
C.1.2. ASSEMBLAGE DU COMPLEXE CF IA 98
C.1.3. TEST D'ACTIVITE DE MATURATION DE L'EXTREMITE 3' DES ARN MESSAGERS 100
C.1.4. TERMINAISON DE LA TRANSCRIPTION PAR L'ARN POLYMERASE II 102
C.2. DISCUSSION ET PERSPECTIVES 108
D. ÉTUDE DES MOTIFS EN « DOIGT DE ZINC » DE PCF11 116
D.1. DONNEES PRELIMINAIRES (S. FRIBOURG – IECB, BORDEAUX) 116
D.2. RESULTATS 117
D.2.1. OBTENTION DES MUTANTS 117
D.2.2. RECHERCHE D'UNE LETALITE SYNTHETIQUE AVEC CLP1 119
D.2.3. ASSEMBLAGE DU COMPLEXE CF IA 121
D.2.4. TEST D'ACTIVITE DE MATURATION DE L'EXTREMITE 3' 123
D.2.5. TERMINAISON DE LA TRANSCRIPTION PAR L'ARN POLYMERASE II 124
D.3. DISCUSSION ET PERSPECTIVES 125

CHAPITRE 2 : MISE EN EVIDENCE D'UNE ACTIVITE KINASE ASSOCIEE AUX FACTEURS DE MATURATION EN 3' DES ARNS MESSAGERS 130

A.1. ORGANISATION DE CLP1 130
A.2. FONCTIONS ATTRIBUEES AUX HOMOLOGUES DE CLP1 131
C.1. RESULTATS OBTENUS 135
C.2. MODIFICATION DES TESTS D'ACTIVITE KINASE 136
C.3. ANALYSE DES DIFFERENTES ETAPES DE PURIFICATION 137
C.4. MODIFICATION DU PROTOCOLE DE PURIFICATION 138
C.5. CONCENTRATIONS DES FACTEURS 140
C.6. TEST D'ACTIVITE ATPASE 141

CONCLUSION GENERALE 151

BIBLIOGRAPHIE 154

LISTE DES ABREVIATIONS

GENERALITES

ARN :	Acide RiboNucléique
ADN :	Acide DésoxyriboNicléoque
snRNAs :	*Small nuclear RNAs*
snRNPs :	*Small nuclear RiboNucleoproteins Particules*
snoRNAs :	*Small nucleolar RNAs*
snoRNPs :	*Small nucleolar RiboNucleoproteins Particules*
pré-ARNm :	Pré-ARN messagers
mRNPs :	*Messenger Ribonucleoproteins Particules*
ARNmi :	Micro ARN
ARNt :	ARN de transfert
ARNr :	ARN ribosomiques
SUTs :	*Stable unannotated transcripts*
CUTs :	*Cryptic unstable transcripts*
XUTs :	*Xrn1-dependent unstable transcripts*
RUTs :	*Reb1-dependent unstable transcripts*
CTD :	C-terminal Domain de Rpb1
GTFs :	*General transcription factors*
PIC :	*Pre-initiation complex*
UTR :	*UnTranslated Region*
Pol II :	ARN Polymérase II
***S. cerevisiae* :**	*Saccharomyces cerevisiae*
D. melanogaster :	*Drosophila melanogester*
***C. elegans* :**	*Caenorhabditis elegans*
***A. thaliana* :**	*Arabidopsis thaliana*
***B. subtilis* :**	*Bacillus subtilis*
BP :	Boite de Branchement
NPC :	*Nuclear Pore Complex*
CF IA :	*Cleavage Factor IA*
CF IB :	*Cleavage Factor IB*
CPF :	*Cleavage and Polyadenylation Factor*
CF II :	*Cleavage Factor II*
PF I :	*Polyadenylation Factor I*
APT :	*Associated with PTa1*
CPSF :	*Cleavage and Polyadenylation Specificity Factor*
CstF :	*Cleavage stimulation Factor*
CF I_m :	*Cleavage Factor I*
CF II_m :	*Cleavage Factor II*
SETX :	Senataxin
PAS :	*PolyAdénylation Signal*
DES :	*Downstream Sequence Element*
USE :	*Upstream Sequence Element*
DSE :	*Downstream Sequence Element*
PAN :	*Pab1-dependant poly(A) nuclease*
HDE :	*Histone Downstream Element*
SLBP :	*Stem-Loop Binding Protein*
APA :	Polyadénylation alternative
IgM :	Immunoglobuline M
IgG :	Immunoglobuline G
PAP I :	*Polyribonucleotides Adenylyl Transférase*
PNPase :	Polynucléotide phosphorylase
PAPnc :	*Polyribonucleotide Adenylyl Transférase* non canonicale
EE :	*Efficiency Element*
PE :	*Positioning Element*
UUE :	*Upstream U-rich Element*
DUE :	*Downstream U-rich Element*
CK1 :	Caséine Kinase 1

CK2 : Caséine Kinase 2
PP1 : *Protein Phosphatase 1*
UBL : domaine « *ubiquitine-like* »
H2B : Histone 2B
RRMs : *RNA Recognition Motifs*
hnRNPs : *heterogeneous nuclear ribonucleoprotein Particules*
HAT : *Half a TRP*
CID : Domaine d'interaction avec le CTD
TSS : *Transcription Start Signal*
NNS : *Nrd1 Nab3 Sen1 complex*
ChIP : *Chromatin ImmunoPrecipitation*
PAR-CLIP : *Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation*
CBC : *Cap Binding Complex*
DDR : *DNA Damage Response*
MMR : *MisMatch Repair*
BER : *Base Excision Repair*
NER : *Nucleotide Excision Repair*
GG-NER : *Global Genome Nucleotide Excision Repair*
TC-NER : *Transcription-Coupled Nucleotide Excision Repair*
SSBR : *Single-Strand Break Repair*
NHEJ/HR : *NonHomologous end joining Homologous Recombination*
PTMs : *Post-Translational Modifications*
DUBs : déubiquitinases
Phase S : Phase de synthèse
SUMO : *small ubiquitin-like modifier*
TSEN : *tRNA Splicing endonuclease*
RISC : *RNA-induced silencing complex*
TRAPP : *TRAsport Portein Particle*
ZF : Zinc Finger
PNK : *PolyNucleotide Kinase*
RPNK : *RNA-specific PolyNucleotide Kinase*

CORRESPONDANCE ADN/ARN ET ACIDES AMINES

U : Uracile
A : Adénine
G : Guanine
C : Cytosine
R : A ou G

Cys (C) : Cystéine
His (H) : Histidine
Ser (S) : Sérine
Thr (T) : Thréonine
Tyr (Y) : Tyrosine
Pro (P) : Proline
Asp (D) : Aspartate ou Acide Aspartique
Iso (I) : Isoleucine
Ala (A) : Alanine
Phé (F) : Phénylalanine
Gly (G) : Glycine
Arg (R) : Arginine
Trp (W) : Tryptophane
Glu (E) : Glutamate ou Acide Glutamique
Leu (L) : Leucine
Lys (K) : Lysine
Val (V) : Valine
X : Acides aminés non identifiés
P-… : Phosphorylation

MATERIELS & METHODES - UNITES

RMN :	Résonance Magnétique Nucléaire
***TAP-tagging* :**	Tandem Affinity Purification
5-FOA :	Acide 5-Fluoro-orotique
DO_{600nm} :	Densité Optique à 600nm
BET :	Bromure d'Éthidium
YPD(A) :	*Yeast Peptone Dextrose Adenine*
SC :	*Synthetic complete*
SDS-PAGE :	SDS PolyAcrylamide Gel Electrophoresis
EDTA :	Acide Éthylène Diamine Tétracétique
SSC :	Tampon Sodium Sodium Citrate
SDS :	Sodium Sodecyl Sulfate
µg :	micro-gramme
TRO :	*Transcriptional Run-On*
RT-qPCR :	Reverse Transcription quantitative Polymerase Chain Reaction
FSBA :	5′-fluorosulfonylbenzoyl 5′-adenosine
Mn^{2+} :	Manganèse
Mg^{2+} :	Magnésium
ATP :	Adénosine TriPhosphate
-CO-CH3 :	groupement acéthyl
-CH3 :	groupement méthyl
m^6A :	N^6-méthyladensine
TEV :	*Tobacco etch virus protéase*
CalFT :	*Calmodulin Flow Through*
EGTA :	*Ethylene Glycol Tetraacetic Acid*
CAMK :	Protéine kinase calmoduline-dépendante
ADP :	Adénosine DiPhosphate
Pi :	OrthoPhosphate inorganique
CBP :	*Calmoduline Binding Peptide*
GAL :	Galactose
NMWL :	*Nominal Molecular Weight Limit*
kDa :	kilo Dalton
MDa :	Méga Dalton
Nt :	Nucléotides
µL :	Microlitre
mL :	Millilitre
mM :	Millimolaire
M :	Molaire
rpm :	Tour par minute
nm :	Nanomolaire
Ci :	Curie
µCi :	MicroCurie
mmole :	millimolaire
FA :	Tampon de charge Formamide EDTA

LISTE DES FIGURES & TABLEAUX

INTRODUCTION

Figure 1.01 : Organisation de l'ARN polymérase II chez *S. cerevisiae*.

Figure 1.02 : Représentation des différentes étapes de maturation des pré-ARNm transcrits par l'ARN polymérase II.

Figure 1.03 : L'ajout de la coiffe est une réaction séquentielle faisant intervenir trois activités enzymatiques.

Figure 1.04 : Les enzymes impliquées dans l'ajout de la coiffe sont recrutées co-transcriptionnellement.

Figure 1.05 : L'élimination des introns se fait par deux étapes de trans-estérification.

Figure 1.06 : L'épissage est permis par un complexe de 1MDa, le « spliceosome ».

Figure 1.07 : Organisation des protéines impliquées dans l'export des mRNPs.

Figure 1.08 : Organisation et composition des pores nucléaires ou NPC chez *S. cerevisiae.*

Figure 1.09 : Organisation des éléments *cis* des gènes codants pour les histones en présence du snRNA U7 – modèle murin.

Figure 1.10 : Modèle d'organisation des éléments *cis* et des facteurs *trans* pour la maturation des ARN histones.

Figure 1.11 : Organisation des snoRNAs chez *S. cerevisiae.*

Figure 1.12 : Les isoformes formés par APA peuvent présenter une région codante différente et/ou une région 3'-UTR de taille variable.

Figure 1.13 : Chez les procaryotes, deux enzymes ont été identifiées comme présentant des activité de polymérase.

Figure 1.14 : Représentation des trois fonctions les mieux caractérisées du complexe TRAMP4/5.

Figure 1.15 : Organisation du complexe de polyadénylation chez *S. cerevisiae.*

Figure 1.16 : Le sous-complexe CPF peut-être divisé en trois sous-parties en fonction de leurs interactions.

Figure 1.17 : Représentation des interactions entre les sous-unités du complexe CPF.

Figure 1.18 : Modèle d'organisation hétéro-tétramérique du complexe CF IA – CF IB.

Figure 1.19 : Représentation des interactions entre les sous-unités des complexes CF IA et CF IB.

Figure 1.20 : Représentation schématique des différentes sous-unités du complexe CF IA et CF IB.

Figure 1.21 : Représentation des modifications du CTD durant la transcription d'un ARN codant par l'ARN polymérase II.

Figure 1.22 : Comparaison du complexe de polyadénylation chez *S.cerevisiae* et chez les mammifères.

TABLEAU 1.1 : Composition du complexe de polyadénylation dans le système mammifère.

Figure 1.23 : Terminaison des ARNs codants par la voie dépendante du complexe de polyadénylation.

Figure 1.24 : Représentation schématique de l'organisation des sous-unités du complexe NNS.

Figure 1.25 : Deux types de complexe NNS sont impliqués dans la terminaison de transcription des ARNs non codants.

Figure 1.26 : L'expression de *NRD1* est autorégulée par terminaison précoce induite par le complexe NNS.

Figure 1.27 : L'accessibilité de site de polyadénylation permet de réguler l'expression de *RPL9*.

Figure 1.28: Modèle de terminaison **torpédo** chez l'Homme.

Figure 1. 29: Glc7 est un acteur clef de la régulation de la maturation de l'extrémité 3'.

Figure 1. 30: L'ubiquitination de Swd2 permet l'export des ARNm.

CHAPITRE 1

Figure 2.01 : Organisation et structure de Pcf11.

Figure 2.02 : Organisation des différents mutants construits pour l'étude du 2nd NTD.

Figure 2.03 : La délétion du 2nd NTD induit un phénotype de croissance lente chez *S. cerevisiae.*

Figure 2.04 : Phénotypes associés aux différentes mutations du 2nd NTD.

Figure 2.05 : Caractérisation du phénotype de croissance lente du mutant délété pour le 2nd NTD.

Figure 2.06 : La *trans*-complémentation de la délétion du 2nd NTD par la protéine entière ou la région N-terminale permet de restaurer le phénotype sauvage.

Figure 2.07 : Étude de la sensibilité au formamide des différents mutants sur 2nd NTD.

Figure 2.08 : Étude des mutants du 2nd NTD et du CID.

Figure 2.09 : L'assemblage du complexe CF IA n'est pas perturbé par l'insertion des mutations au niveau du 2nd NTD de Pcf11.

Figure 2.10 : La double mutations *pcf11-17* (L189D, L213D) conduit à une instabilité de la partie N-terminale de Pcf11.

Figure 2.11 : Double purification par tap-tagging des complexes [CF IA+CPF] pour les différents mutants du 2nd NTD.

Figure 2.12 : Le 2nd NTD et le motif Q_{20} ne sont pas impliqués dans la maturation de l'extrémité 3' des ARN messagers *in vitro*.

Figure 2.13 : Étude du rôle du 2nd NTD dans la terminaison de transcription par Northern blot.

Figure 2.14 : Étude du rôle du 2nd NTD dans la terminaison de transcription par essai de protection contre la digestion par la S1 Nucléase – snr13 boîte C/D.

Figure 2.15 : Étude du rôle du 2nd NTD dans la terminaison de transcription par essai de protection contre la digestion par la S1 Nucléase – snR5 boîte H/ACA.

TABLEAU 2.1 : Récapitulatif des différents mutants obtenus dans l'étude du 2nd NTD.

Figure 2.16 : Données préliminaires sur les motifs en doigt de zinc présents dans la partie C-terminale de Pcf11.

Figure 2.17 : Représentation schématique des différents mutants construits pour l'étude des motifs en doigts de zinc de Pcf11.

Figure 2.18 : La mutation des deux motifs en doigt de zinc de Pcf11 est létale chez *S. cerevisiae.*

Figure 2.19 : La mutation du second motif en doigt de zinc induit un phénotype de croissance lente marqué.

Figure 2.20 : Recherche de létalité synthétique entre les mutations des ZF de Pcf11 et différentes mutations de Clp1.

Figure 2.21 : L'assemblage du complexe CF IA n'est pas modifié par l'insertion des mutations au niveau des deux motifs en doigt de zinc de Pcf11.

Figure 2.22 : Les motifs en doigt de zinc de Pcf11 sont nécessaires aux activités de maturation de l'extrémité 3' des ARN messagers.

Figure 2.23 : Les motifs en doigt de zinc de Pcf11 ne sont pas impliqués dans la terminaison de la transcription des snoRNAs.

Figure 2.24 : Représentation du locus de snR5 et des tailles attendues en essai de protection contre la S1 Nucléase.

TABLEAU 2.2 : Récapitulatif des différents mutants obtenus dans l'étude des motifs en doigts de zinc.

Figure 2.25 : Représentation du modèle d'action de Yra1 dans la maturation de l'extrémité 3'.

CHAPITRE 2

Figure 3.01 : Organisation de Pcf11.

Figure 3.02 : Résultats préliminaires – Une activité kinase est retrouvée associée aux complexes CF IA et CPF.

Figure 3.03 : Activité kinase associée aux complexes nouvellement purifiés.

Figure 3.04 : Modification du protocole de tests d'activité kinase des complexes CF IA témoins.

Figure 3.05 : Purification par TAP-tagging.

Figure 3.06 : Les différentes étapes de purification présentent une activité kinase.

Figure 3.07 : La concentration des facteurs permet de rétablir un marquage comparable à celui des facteurs témoins.

Figure 3.08 : Clp1 ne possède pas d'activité d'hydrolyse de l'ATP.

MATERIELS ET METHODES

TABLEAU 4.1 : Souches de *S. cerevisiae* utilisées dans cette étude.

TABLEAU 4.2 : Vecteurs levures utilisés dans cette étude.

TABLEAU 4.3 : Vecteurs bactéries utilisés pour l'expression des protéines recombinantes.

TABLEAU 4.4 : Oligonucléotides utilisés lors des clonages et séquençages.

TABLEAU 4.5 : Anticorps utilisés dans cette étude.

TABLEAU 4.6 : Tests d'activité kinase associée aux facteurs.

TABLEAU 4.7 : Plasmides et séquences des ARN synthétiques CYC1 et CYC1-pré.

TABLEAU 4.8 : Composition des mix réactionnel utilisés lors de la reconstitution des activités de maturation de l'extrémité 3'.

Figure 4.01 : Transfert sur membrane de Nylon.

PREFACE

PREFACE

L'ARN polymérase II permet la synthèse d'ARNs messagers codants (ARNm) et d'ARNs non codants comme les **snRNAs** et les **snoRNAs**. Directement après leur synthèse, les pré-ARNm sont maturés pour former les ARNm matures. Cette maturation est étroitement liée aux étapes de synthèse et est considérée comme une **modification co-transcriptionnelle**. La maturation du pré-ARNm englobe la maturation de l'extrémité 5', caractérisé par l'ajout d'une **coiffe**, l'élimination des introns non codants lors de l'**épissage** et l'ajout d'une **queue poly(A)** à l'extrémité 3'.

Ce projet de thèse a été réalisé au sein de l'équipe du Dr L. Minvielle-Sébastia qui s'intéresse à la maturation de l'extrémité 3' et au complexe protéique permettant cette modification. L'ajout de la queue poly(A) est une étape essentielle pour la **terminaison de transcription** et participe à la régulation **qualitative** et **quantitative** de l'expression d'un gène, permettant la régulation de la stabilité du transcrit, de **son export** ainsi que **la traduction** du message génétique. Deux projets sont décrits dans ce manuscrit ; le premier s'intéresse à une sous-unité du complexe de polyadénylation et le second traite du complexe dans son ensemble et à la régulation éventuelle de ce complexe par phosphorylation. Le premier chapitre de ce manuscrit s'intéresse plus spécifiquement à **Pcf11**. **Pcf11** est une protéine essentielle pour la **maturation de l'extrémité 3'** ainsi que la **terminaison de transcription** des ARNs codants et non codants (snoRNAs). L'objectif est d'essayer de comprendre la fonction de deux sous-domaines de cette protéine – le **2nd NTD** et les motifs **ZF** – dont l'organisation et la structure restent, pour l'instant, inconnues. Pour répondre à cette question, une analyse par **génétique inverse** chez *Saccharomyces cerevisiae* a été mise en place, par délétion et par mutation ponctuelle. Le second chapitre est basé sur des résultats préliminaires obtenus au laboratoire ayant permis de montrer qu'une **activité kinase** co-purifiait spécifiquement lors de la purification des facteurs du complexe de polyadénylation.

Dans l'introduction, après une partie détaillant chaque étape de synthèse par l'ARN polymérase II, trois parties sont développées : l'organisation de chaque sous-unité du complexe de polyadénylation chez la levure et chez les eucaryotes supérieurs, la terminaison de transcription des ARNs codants et non codants et enfin le rôle des modification post-traductionnelles sur l'activité et l'organisation du complexe. Chaque chapitre est ensuite traité indépendamment.

Introduction

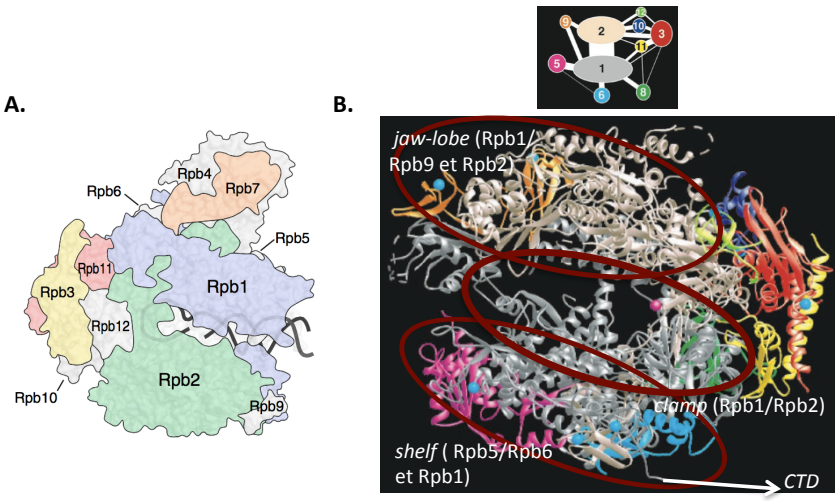


Figure 1.01 : Organisation de l'ARN polymérase II chez *S.cerevisiae*. **(A.)** Représentation schématique des douze sous-unités formant le *core enzyme* de l'ARN polymérase II. Rpb1 et Rpb2 sont les plus grandes sous-unités du complexe et forme le domaine catalytique. Rpb3, Rpb10, Rpb11 et Rpb12 sont des protéines d'architecture et permettent l'assemblage du complexe. Les sous-unités Rpb4 et Rpb7 ne sont pas nécessaires pour les activités de transcription. **(B.)** Structure des dix sous-unités nécessaires pour la transcription. Trois éléments mobiles peuvent être distingués : le module « *jaw-lobe* » (Rpb1/Rpb9 et Rpb2), le module « *shelf* » (Rpb5/Rpb6 et une partie de Rpb1) et le module « *clamp* » (Rpb1/Rpb2). Le CTD de Rpb1 s'étend vers l'extérieur au niveau de Rpb6. D'après Braberg *et al.*, 2013 et Cramer *et al.,* 2001.

INTRODUCTION

A. MATURATION DES PRECURSEURS DES ARNM

A.1. Généralités

A.1.1. ARN Polymérase II

Dans les cellules eucaryotes, le génome est transcrit par trois complexes protéiques : l'ARN polymérase Pol I (précurseur des ARNs ribosomiques), l'ARN polymérase Pol II (précurseur des ARNs codants et non codants) et l'ARN polymérase Pol III (précurseur des ARNs de transferts et des autres petits ARNs). L'ARN Polymérase II permet la synthèse des ARNs codants et des ARNs non codants : « *small nuclear RNAs* » (snRNAs), « *small nucleolar RNAs* » (snoRNAs), « *microRNAs* » (miRNA), « *cryptic unstable transcripts* » (CUTs) et « *stable unannotated transcripts* » (SUTs). L'ARN polymérase II est composée de 12 sous-unités codées par les gènes *RPB1* à *RPB12* chez la levure. Sa structure et l'organisation des douze sous-unités formant le « *core enzyme* » sont présentées dans la **Figure 1.01** (Braberg et al., 2013). Cette unité enzymatique de plus de 500kDa est remarquablement conservée dans tout le règne eucaryote. La plus grande des sous-unités, Rpb1 porte en partie l'activité catalytique et présente à son extrémité C-terminale une structure désordonnée, le CTD (« *Carboxy-terminal domain* »). L'activité de l'enzyme est modulée par des cycles dynamiques de phosphorylations/déphosphorylations au niveau de cette plate-forme, et ce tout au long du processus de transcription. L'ARN polymérase Pol II est au centre d'une machinerie complexe, dont la composition varie durant la transcription des gènes.

Traditionnellement, la synthèse par l'ARN polymérase Pol II peut être divisée en plusieurs étapes, que l'on sait étroitement liées physiquement et fonctionnellement. Dans un premier temps, durant l'initiation de la transcription, la Pol II est recrutée au niveau du promoteur en présence des facteurs généraux de transcription (GTFs) – TFIIA, TFIIB, TFIID, TFIIE, TFIIF et TFIIH. Le *core enzyme* et les facteurs généraux de transcription forment le complexe de pré-initiation (ou PIC – « *pre-initiation complex* »). Le « *mediator* » est un complexe de 25 sous-unités chez la levure qui transmet les signaux régulateurs à la polymérase II, permettant de stimuler l'activation de la transcription ainsi que la

phosphorylation du CTD de l'ARN polymérase II par l'une des sous-unité du complexe TFIIH (Boube et al., 2002). Le PIC permet l'ouverture, le déroulement des deux brins d'ADN et la synthèse des premiers nucléotides du pré-ARNm naissant. Lorsque la chaîne d'ARN atteint la taille de 10 nucléotides, la bulle de transcription se déplace, permet le démarrage de la transcription et la libération de la Pol II du complexe de pré-initiation (Holstege et al., 1997). La libération de la Pol II du PIC est favorisée par la phosphorylation du CTD sur la SERINE 5 ce qui permet de dissocier des GTFs, le « *mediator* » et le recrutement des facteurs d'élongation. Au cours de l'élongation, la nature des modifications du CTD varie – la séquentialité de ces modifications sont abordés dans le Paragraphe B.1.3. La bulle de transcription, qui correspond à l'ouverture de la molécule d'ADN sur 12 à 15 paires de bases, se déplace vers la région 3' UTR du gène. La Pol II synthétise alors 1500 à 2000 ribonucléotides par minute. La terminaison de transcription est caractérisée par une pause de la polymérase, le relargage du pré-ARNm et le recyclage de l'ARN polymérase Pol II (Proudfoot and O'Sullivan, 2002). Plusieurs mécanismes de terminaison de transcription peuvent être utilisés et seront décrits dans ce manuscrit.

Au cours de la synthèse, le pré-ARNm naissant subit différentes modifications permettant sa maturation. Il est modifié à son extrémité 5' par l'ajout d'une coiffe, les introns sont éliminés lors de l'épissage, et le pré-ARNm est modifié à son extrémité 3' par l'ajout d'une queue poly(A) (**Figure 1.02**). L'ARNm mature est ensuite exporté vers le cytoplasme. L'ensemble de ces réactions biochimiques est étroitement lié à la transcription. Le recrutement des facteurs impliqués est synchronisé par le CTD de l'ARN polymérase Pol II en fonction de son état de phosphorylation (Bartkowiak et al., 2011; Egloff and Murphy, 2008; Napolitano et al., 2014; Phatnani and Greenleaf, 2006). Un défaut de l'une de ces étapes de maturation entrave la progression des autres étapes, illustrant leur interdépendance (Nino et al., 2013).

A.1.2. Maturation de l'extrémité 5'

La première étape de maturation du pré-ARNm est l'ajout d'une coiffe à l'extrémité 5'. Cette étape est conservée dans le règne eucaryote et chez les virus, mais n'est pas retrouvée dans le règne des procaryotes – les enzymes impliquées n'étant pas présentes (Shuman, 2002). L'ajout de la coiffe, ou **m^7GpppN**, se fait dès la synthèse de 30 nucléotides par l'ARN polymérase II, par réaction séquentielle faisant intervenir trois enzymes. Dans un

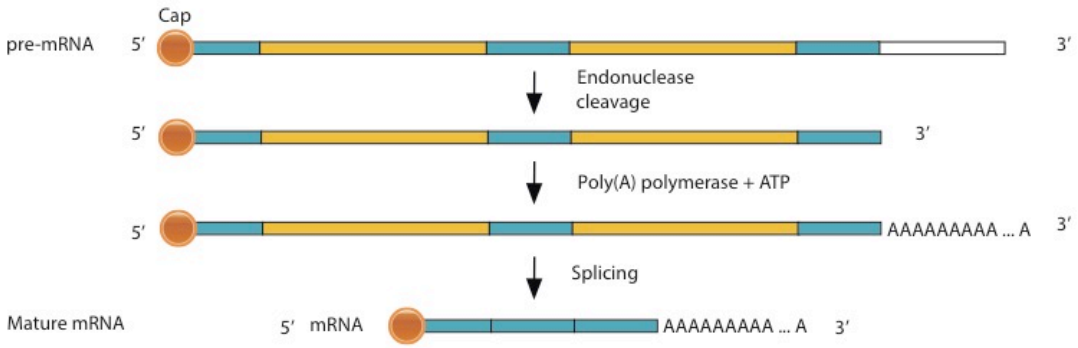


Figure 1.02 : Représentation des différentes étapes de maturation des pré-ARNm transcrits par l'ARN polymérase II. Le pré-ARNm synthétisé par l'ARN polymérase II subit trois modifications majeures afin d'être maturé. L'ajout de la **coiffe à l'extrémité 5'** prévient l'action des exonucléases. L'**épissage** correspond à l'élimination des séquences non codantes ou introns, et permet de « maturer » le message génétique. L'ajout de la **queue poly(A) à l'extrémité 3'** permet de moduler la stabilité de l'ARN et permet le recrutement des protéines permettant son export vers le cytoplasme

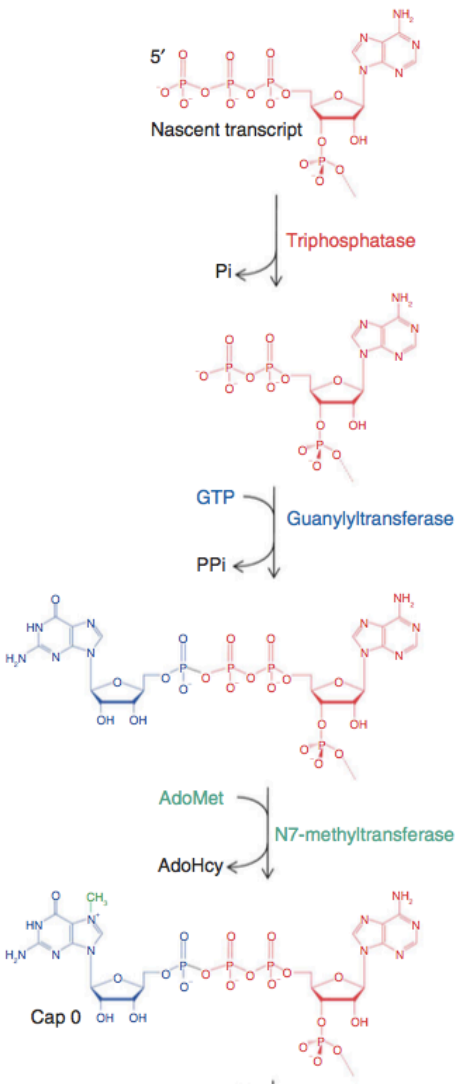


Figure 1.03 : L'ajout de la coiffe est une réaction séquentielle faisant intervenir trois activités enzymatiques. La maturation en 5' par l'ajout de m^7GpppN est une modification réversible. Dans un premier temps, **l'ARN triphosphatase** (en **rouge**) permet l'hydrolyse du phosphate γ du premier nucléotide en 5'. **L'ARN guanylyltransférase** (en **bleu**) permet ensuite le transfert de GMP sur le diphosphate de l'ARN à partir de GTP et la libération de PPi. La dernière étape est l'ajout d'un groupement méthyle sur la guanine en position N7, permise par **l'ARN guanine-N7-méthyltransférase** (en **vert**). Le groupement méthyle est fourni par le coenzyme, la S-adenosylméthionine ou AdoMet. La structure formée est une structure de base appelée « Cap 0 » ou coiffe 0. Chez certains eucaryotes, comme les trypanosomes, l'extrémité 5' peut subir d'autres étapes de méthylation – coiffe 1 à 4.

D'après Ghosh et Lima, 2010.

premier temps, le phosphate γ de l'extrémité 5' est hydrolysé par **l'ARN triphosphatase** (Cet1 chez *S. cerevisiae*). Une guanine monophosphate est ensuite transférée sur le 5' diphosphate de l'ARN par **l'ARN guanylyltransférase** (Ceg1 chez *S. cerevisiae*). Une dernière étape de méthylation de la guanine en position N7 par **l'ARN guanine-N7-méthyltransférase** (Abd1 chez *S. cerevisiae*) permet de terminer la réaction de maturation de l'extrémité 5' (**Figure 1.03**) (Ghosh and Lima, 2010; Li and Kiledjian, 2010). Ces étapes sont conservées, seule l'organisation génétique des enzymes et les mécanismes catalytiques varient dans le règne eucaryote.

Chez *S. cerevisiae*, l'**ARN triphosphatase** (Cet1) fait partie de la famille des TTM – « *Triphosphate Tunnel Métaloenzyme* » – et utilise le manganèse comme cofacteurs (Lima et al., 1999). L'**ARN guanylyltransférase** Ceg1 permet le transfert de GMP sur le diphosphate de l'ARN, en présence d'ion magnésium. Ceg1 est formée d'un domaine nucléotidyltransférase à l'extrémité N-terminale et d'un domaine OB (*Oligosaccharide Binding*) dans la région C-terminale. L'analyse par sédimentation a permis de montrer que Ceg1 et Cet1 fonctionnent sous la forme d'un complexe 1 : 2 via une région *linker* fournie par Cet1. Cela permet de stabiliser la fonction guanylyltransférase et le recrutement de Cet1 par interaction avec Ceg1. L'**ARN guanine-N7-méthyltransférase** Abd1 permet la méthylation de la guanine en position N7 et est dépendante de la présence d'un coenzyme, le S-adénosylméthionine (SAM).

Les enzymes nécessaires à l'ajout de la coiffe à l'extrémité 5' sont recrutées lors de l'initiation, par interaction avec la SER^5 phosphorylée du CTD de Rpb1, sous-unité de l'ARN polymérase II (**Figure 1.04**) (McCracken et al., 1997)

L'ajout de la coiffe présente un rôle majeur dans le devenir de l'ARN. Il permet tout d'abord de maintenir la stabilité de l'ARNm, en le protégeant de la dégradation par des exonucléases comme Xrn1 chez *S. cerevisiae*. La coiffe joue un rôle dans l'épissage des ARNm en participant au recrutement des snRNPs. L'ajout de la coiffe joue également un rôle important dans l'export des mRNPs et dans la traduction de l'ARNm, via le recrutement de la sous-unité 40S du ribosome (Varani, 1997).

La maturation 5' est une modification réversible, permettant de contrôler la demi-vie des ARNm. L'élimination de la coiffe ou « *decapping* » intervient dans les voies de dégradation 5' et 3' permettant de libérer un mono-phosphate à l'extrémité 5' (Gu and Lima, 2005). Ces enzymes de « decapping » interviennent après un raccourcissement de la queue

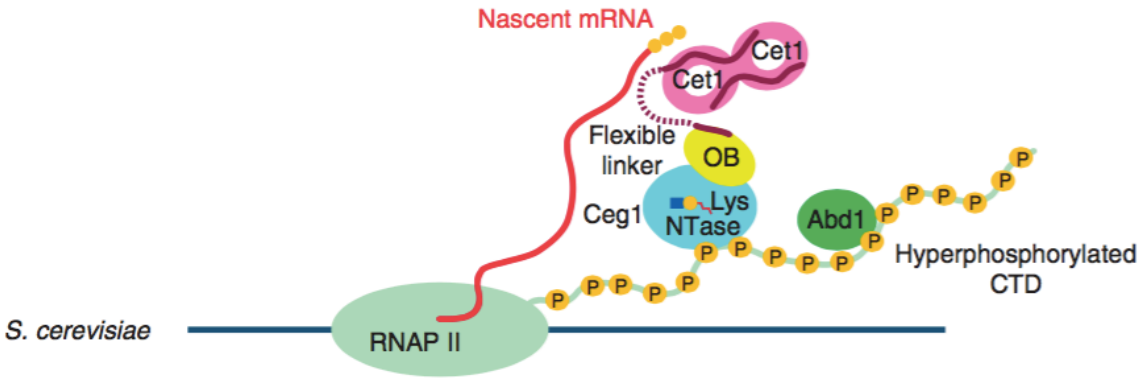


Figure 1.04 : Les enzymes impliquées dans l'ajout de la coiffe sont recrutées co-transcriptionnellement. Chez *S.cerevisiae*, **Cet1** porte l'activité d'ARN triphosphatase et est recrutée sous la forme d'un dimère. **Cet1** permet l'interaction avec l'ARN naissant. La région flexible de Cet1 (*flexible linker*) permet le recrutement de Ceg1, ARN guanylyltransférase. **Ceg1** est formée d'un domaine OB et d'un domaine catalytique dans la région N-terminale, permettant l'interaction avec le CTD de l'ARN polymérase II sous sa forme P-SER5. **Abd1**, l'ARN guanine-N7-léthyltransférase est recrutée indépendamment des autres activités enzymatiques, également par interaction avec la forme P-SER5 du CTD. D'après Ghosh et Lima, 2010.

poly(A). Dcp2 intervient dans la voie de dégradation en 5'. C'est une protéine liant l'ARN de la famille des hydrolases NUDIX, qui va cliver la coiffe en présence de manganèse ou de forte quantité d'ion magnésium. Dcp2 n'est actif que sur les ARNs de plus de 25 nucléotides. L'ARN monophosphate est ensuite dégradé par l'exonucléase Xrn1 (Sinturel et al., 2012). Dans la voie de dégradation en 3', DspS, nucléotide hydrolase de la famille des HIT (« *Histidine Triad* ») intervient en fin de réaction et permet la dégradation de la coiffe des ARNs de moins de 10 nucléotides.

A.1.3. Épissage

La réaction d'épissage est un événement co-transcriptionnel, réalisé en deux réactions de trans-estérification : l'excision de l'intron non codant et la liaison de deux exons (**Figure 1.05**) (Ruby and Abelson, 1991). Chez *S. cerevisiae*, seuls 5% des gènes contiennent des régions introniques (≈250 gènes) et la majorité d'entre eux ne contiennent qu'un seul intron, proche de l'extrémité 5' du gène (Lopez and Seraphin, 1999). *YRA1*, *RPL30* ou encore *ACT1* présente des régions introniques, généralement de grande taille. La majorité des gènes contenant un intron code pour des protéines ribosomiques chez *S. cerevisiae*. L'épissage reste donc un mécanisme relativement fréquent. Aucun épissage alternatif (jonction entre exons éloignés ou élimination d'un exon) n'est connu à ce jour chez *S. cerevisiae*. L'épissage de certaine protéine, comme Yra1, constitue un système de boucle de régulation négative – l'expression de *YRA1* est contrôlée par la présence d'un intron (Preker and Guthrie, 2006; Preker et al., 2002).

Tout d'abord, une attaque nucléophilique par le 2'-OH du ribose de l'adénosine du « point de branchement » (BP) a lieu sur le phosphate de la jonction exon-intron en 5'. Cela permet la formation d'une extrémité libre en 3' de l'exon 1 et d'une structure en boucle (« *lariat* ») correspondant à l'intron, toujours liée au 5' de l'exon 2. Après cette coupure, le 3'-OH libéré au niveau de l'exon 1 situé en amont établit une liaison avec le phosphate de la jonction intron-exon 2 en aval. Les produits obtenus sont les deux exons ligaturés et l'intron intermédiaire sous forme de lariat (**Figure 1.05A**) (Jurica and Moore, 2003; Ruby and Abelson, 1991).

De la même manière que pour la maturation de l'extrémité 3', cette réaction requiert des éléments positionnés en *cis* (« *cis-acting element* ») et des facteurs protéiques agissant en *trans* (« *trans-acting factors* ») (Will and Luhrmann, 2011). Chez *S. cerevisiae*, les éléments

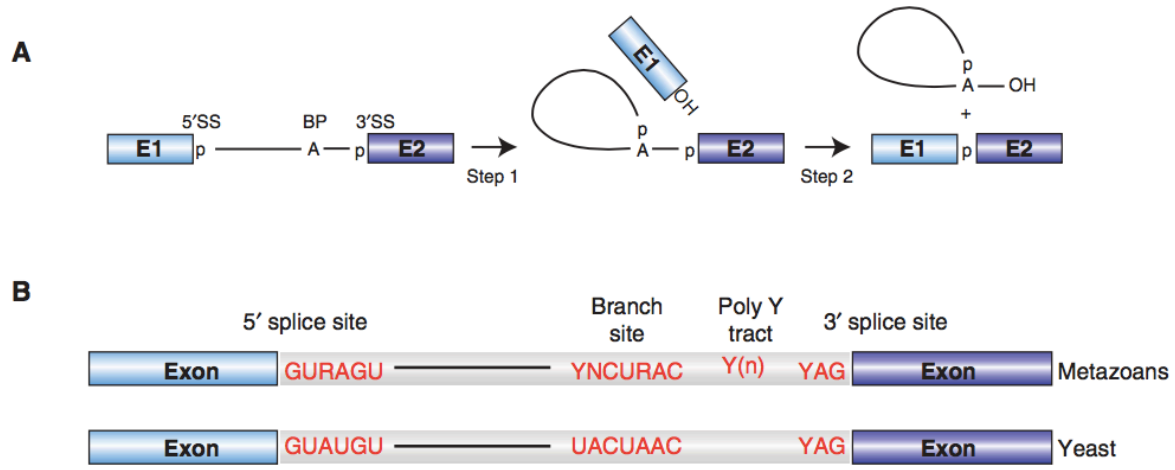


Figure 1.05 : L'élimination des introns se fait par deux étapes de trans-estérification. **(A.)** Représentation des deux étapes de trans-estérification. Le groupe OH de l'Adénosine de la boite de branchement (*BP*) attaque le phosphate de l'exon 1 en amont (*E1 5'SS*). Dans un second temps, le 3'OH libéré en 3' de l'exon 1 est lié au phosphate 5' de l'exon 2 en aval (*E2 3'SS*). Les produits obtenus sont les deux exons ligaturés et l'intron intermédiaire ou « *lariat* ». **(B.)** Les trois éléments *cis* impliqués dans l'épissage sont conservés entre les eucaryotes supérieurs et *S.cerevisiae*. La séquence de la boite de branchement est moins conservés chez les eucaryotes supérieurs. **Y** : résidus pyrimidines ; **R** : résidus purines. D'après Will et Luhrmann, 2011.

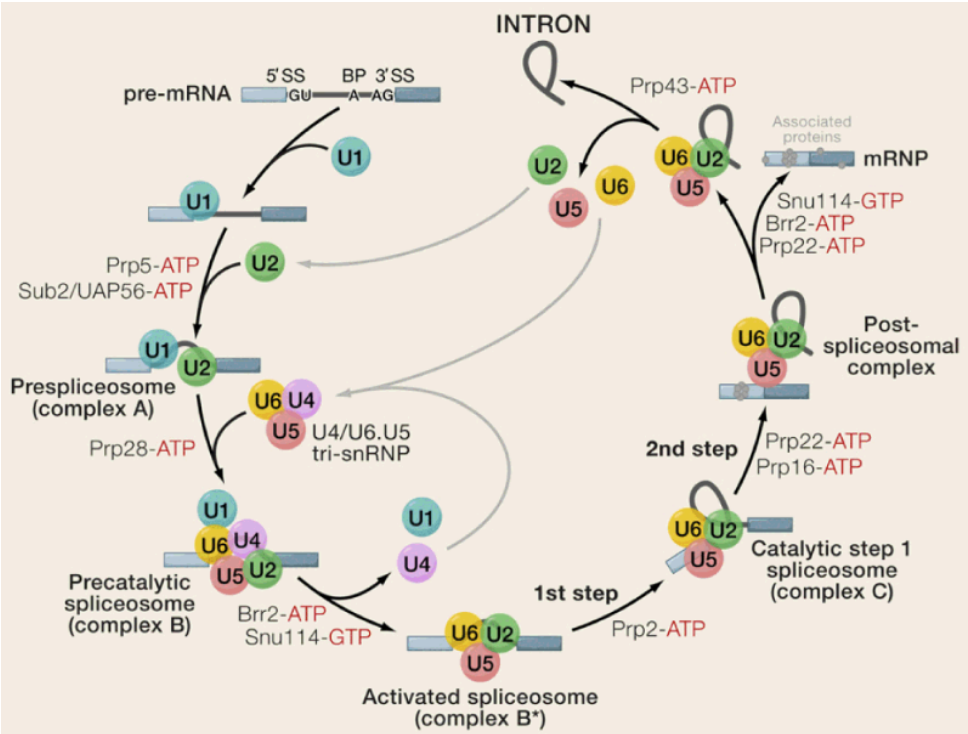


Figure 1.06 : L'épissage est permis par un complexe le « spliceosome ». Représentation de la réaction d'épissage de l'intron contenu dans un pré-ARNm contenant deux exons. Le spliceosome est formé de différentes snRNPs (U1, U2, U5 et U4/U6) dont le recrutement et l'assemblage est séquentiel, via les interactions avec les différents facteurs d'épissage, durant la réaction. Les snRNPs sont formées d'un snRNA et de différentes protéines accessoires (Sm et LSm). La réaction fait intervenir différentes ATPases et hélicases indiquées en gris. D'après Will et Luhrmann, 2011.

cis sont fortement conservés en comparaison du système mammifère. Le site d'épissage de l'exon 5' présente la séquence consensus GUAUGU. La séquence au niveau de la boîte de branchement est fortement conservée et correspond à la séquence UACUAAC – le A en position 6 correspond au A donneur lors de l'attaque nucléophile. La séquence d'épissage de l'exon en 3' est moins conservée et est définie par PyAG (Kuhn and Kaufer, 2003) (**Figure 1.05B**)**.** Les facteurs en *trans* sont regroupés dans un complexe de 1MDa, le « *spliceosome* ». Sa composition, son assemblage et son fonctionnement sont fortement conservés de la levure aux eucaryotes supérieurs et varient durant les différentes étapes de l'épissage (**Figure 1.06**). Contrairement au système mammifère, l'épissage est indépendant du CTD de l'ARN polymérase II chez *S. cerevisiae* (Leipe et al., 2002).

A.1.4. Export

Les étapes de synthèse et de maturation des ARNs messagers sont physiquement séparées de la traduction par la membrane nucléaire, régulant les échanges entre les deux compartiments. L'export des ARNm vers le cytoplasme est ATP dépendant mais ne requiert pas de protéines de type importine / karyophérine ß, contrairement aux ARNt ou au ARMmi reconnus respectivement par les exportines t et 5. Les mRNPs (« *mRNA RiboNucléoProteins Particules* ») (Batisse et al., 2009; Daneholt, 2001) formées de l'ARNm mature et des protéines accessoires, sont libérées du site de transcription puis prises en charge par les protéines, Mex67/Mtr2 chez *S. cerevisiae* (TAP/NXT1 chez l'homme). L'organisation de Mex67 est présentée dans la **Figure 1.07A**. Elles sont ensuite exportées vers le cytoplasme via les pores nucléaires. Les pores nucléaires (ou NPC) sont des structures complexes, fortement conservées, de 40 à 60MDa formées de 400 à 500 nucléoporines (ou nups) chez *S. cerevisiae* (Nino et al., 2013) (**Figure 1.08**). Les nucléoporines sont regroupées en fonction de leur localisation, et permettent d'ancrer la structure dans la membrane et s'ouvrent sous la forme d'un panier nucléaire et de filaments cytoplasmiques. Les nucléoporines-FG présentent une région répétée, riche en résidus phénylalanines et glycines, et sont situées au niveau du panier nucléaire, formant un maillage sélectif. Mex67 permet le cheminement des mRNPs à travers le pore nucléaire par des interactions séquentielles spécifiques avec les nucléoporines-FG (Zenklusen and Stutz, 2001). Mtr2 est requis pour une interaction optimale entre Mex67 et le pore nucléaire ; Mtr2 permettrait de maintenir la conformation du domaine d'interaction de Mex67 pour permettre l'interaction avec les nucléoporines-FG (Fribourg and Conti, 2003). Une fois au niveau des filaments cytoplasmiques, la conformation des ARNm est modifiée

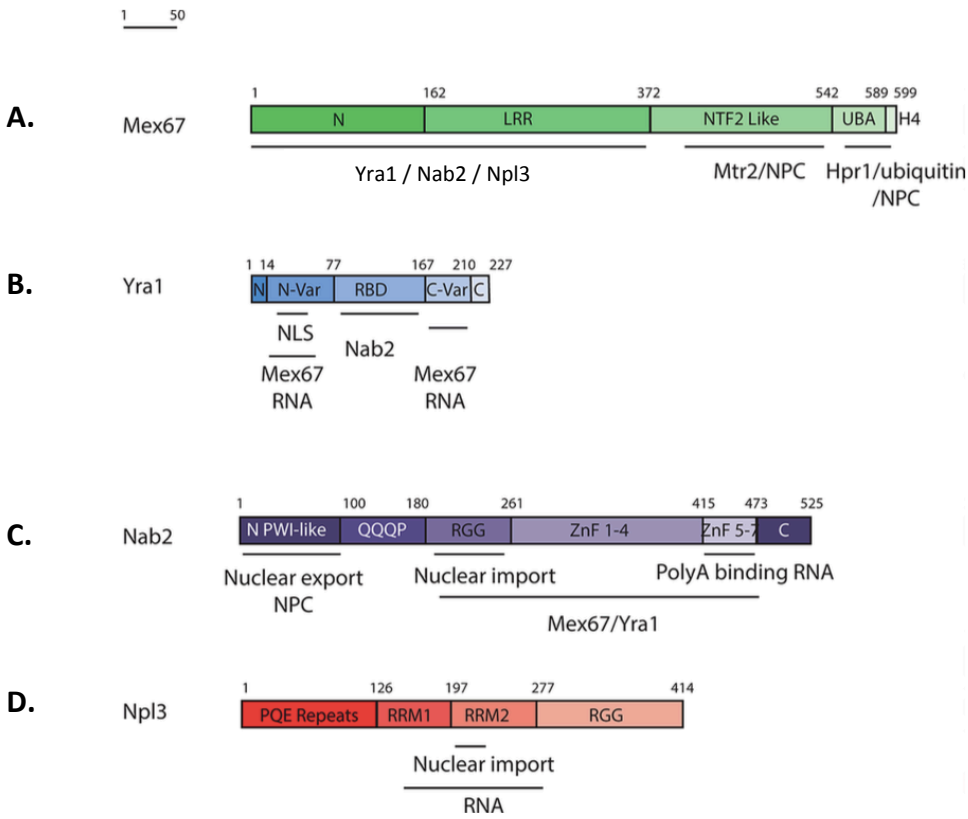


Figure 1.07 : Organisation des protéines impliquées dans l'export des mRNPs. Chaque protéine est représentée à l'échelle, en fonction de sa taille en aminoacides. Les régions impliquées dans les interactions avec d'autres facteurs d'export sont indiquées en dessous de chaque protéine. D'après Nino et al, 2013.

(A.) Mex67 est formée de quatre régions : une région N-terminal (1-162), un domaine riche en résidus Leucines (**LRR** – 162-372), un domaine d'organisation proche du récepteur d'import NTF-2 (NTF-2 like – 372-542), un domaine UBA (« *ubiquitin associated domain* » - 542-599). Le domaine UBA est formé de trois hélices α et d'une quatrième hélice α appelée H4 permettant de prévenir les interactions non spécifiques.

(B.) Yra1 est formée d'une région centrale (77-167), encadré par des régions N-terminale et C-terminale fortement conservées (N, C), et des régions variables caractérisées par la présence de résidus glycine et asparagine (N-var et C-var). L'interaction avec Pcf11 fait intervenir les résidus 1-66 et 167-226 (Johnson *et al.,* 2009).

(C.) Nab2 est formée cinq régions : la région N-terminale formée d'un domaine PWI-like (Proline-Tryptophane-Isoleucine – 1-100) suivi d'une région riche en glutamine (motif QQQP – 100-180), d'une région centrale formée d'un domaine RGG (riche en arginines et glycines – 180-261), de sept motifs en doigt de zinc de type CCCH (291-415 et 415-473) et d'une région C-terminale (473-525).

(D.) Npl3 est formée d'une région N-terminale contenant des répétitions PQE (proline, glutamine et acide glutamique – 1-126), de deux domaines centraux liant l'ARN (126-197 et 197-277) et d'une région C-terminale formée d'un domaine RGG (277-414).

par l'action de protéines contenues dans les mRNPs et de protéines liées au NPC – Dbp5 (ARN hélicase ATP dépendante), Nup159 (impliquée dans la localisation de Dbp5) et Gle1 (permettrait l'activation de Dbp5). Cela permet la dissociation ATP dépendante des mRNPs, de manière irréversible et la libération de l'ARNm dans le cytoplasme. Les protéines impliquées dans l'export sont réimportées vers le noyau.

Le dimère Mex67/Mtr2 présente une faible affinité pour l'ARNm (Braun et al., 1999). L'interaction est indirecte et nécessite le recrutement d'adaptateurs durant la transcription par l'ARN polymérase II (Lei et al., 2001). Les adaptateurs sont impliqués dans les différentes étapes de biogénèse des ARNm. Les adaptateurs THO/TREX et TREX2 permettent le recrutement de Mex67/Mtr2 et la formation de mRNPs compétents pour l'export mais ne sont pas impliqués dans le transport à travers les NPC.

Yra1 (Aly/REF chez l'homme) est l'adaptateur le mieux caractérisé. Cette protéine de 25kDa est codée par un gène essentiel chez *S. cerevisiae* et appartient à la famille des « *REF binding protein* » (« *RNA and Export Factor* »). L'organisation de Yra1 est présentée dans la **Figure 1.07B**. Yra1 est recrutée lors de l'élongation via son interaction avec Pcf11 (Johnson et al., 2011) et par interaction avec la forme P-SER^2/ P-SER^5 du CTD de l'ARN polymérase II (MacKellar and Greenleaf, 2011). Lorsque la polymérase atteint la région 3'UTR du gène, l'interaction entre Yra1 et Pcf11 serait abolie par l'interaction avec Clp1 (Johnson et al., 2009). Cela permettrait alors l'assemblage du complexe CF IA, le clivage, l'ajout de la queue poly(A) et le relargage de mRNPs nouvellement synthétisées. Le transfert de Yra1 au niveau des mRNPs, et ainsi l'interaction avec Mex67 est permise par interaction avec Sub2 (Johnson et al., 2009; Strasser and Hurt, 2001). Yra1 est dissociée des mRNPs avant leur translocation dans la partie cytoplasmique du NPC. Yra1 et Sub2 sont également recrutés au niveau de certains ARNs non codants (CUTs, SUTs, snoRNAs et ARNt) et pourraient jouer un rôle dans leur export vers le cytoplasme (Johnson et al., 2011).

Une surexpression de **Nab2** permet de compenser la délétion de *YRA1* (Iglesias et al., 2010). Nab2 (pour « *nuclear polyadenylated RNA-binding protein* ») permet, avec Yra1, le recrutement de Mex67 sous la forme d'un trimère. Contrairement à Yra1, Nab2 reste associée au mRNPs, lors de l'export et est relarguée au niveau de la face cytoplasmique du pore nucléaire. Des mutations de son homologue chez l'homme conduisent à des troubles neurologiques (Park et al., 2011). Nab2 est retrouvée impliquée dans la régulation de la taille de la queue poly(A) ; son organisation (**Figure 1.07C**) et sa fonction seront discutées en détails plus loin dans ce manuscrit.

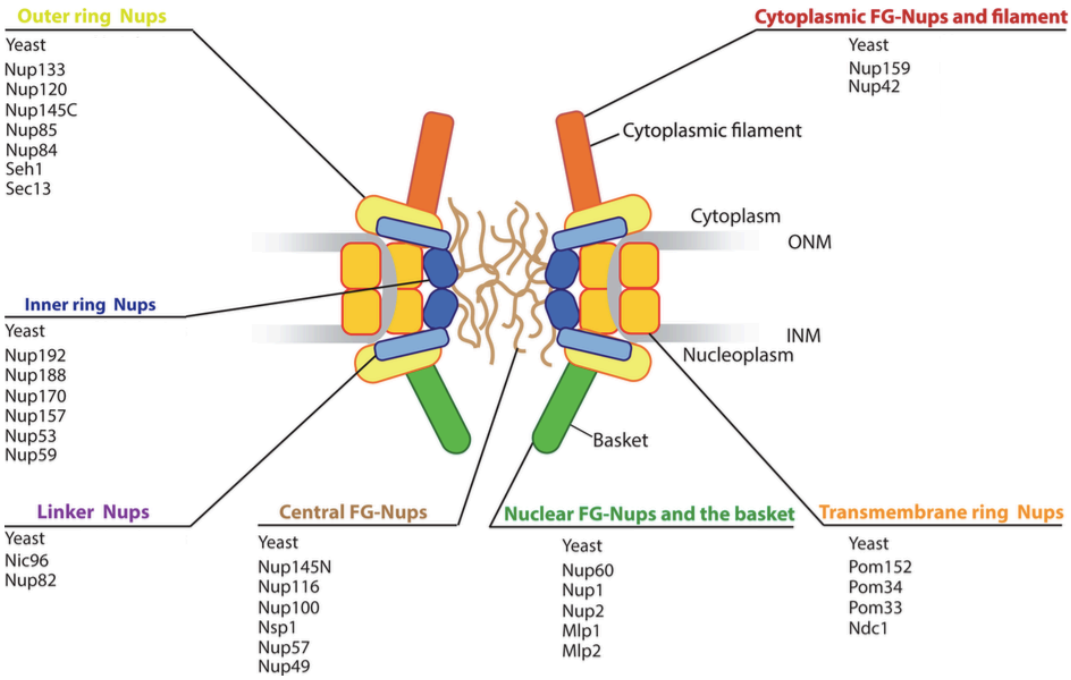


Figure 1.08 : Organisation et composition des pores nucléaires ou NPC chez *S. cerevisiae*. Les pores nucléaires sont des structures complexes, fortement conservées, de 40 à 60MDa formées de 400 à 500 nucléoporines ou Nups. Le pore nucléaire est formé de différentes nucléoporines d'architecture permettant l'ancrage à la membrane. Il s'ouvre dans le noyau sous la forme d'un panier nucléaire (en **vert** vers le bas) et vers le cytoplasme sous la forme de filaments cytoplasmiques (en **rouge** vers le haut) formés des FG-nups. Les FG-nups centrales (en **marron**) forment un maillage sélectif permettant l'interaction avec les différents facteurs d'export. D'après Nino *et al.*, 2013.

Le complexe **THO/TREX** a récemment été identifié comme impliqué dans l'export (Katahira, 2012). Le complexe THO est un complexe tétramérique, formé de Hpr1, Tho2, Thp2 et Mft1, faiblement conservé dans le règne eucaryote (Katahira, 2012). Tex1, Yra1 et son partenaire Sub2 interagissent avec le complexe THO formant le complexe TREX. Le complexe TREX est recruté durant les différentes étapes de synthèse par l'ARN polymérase II, interagi avec de nombreux cofacteurs (maturation de l'extrémité 5', épissage) et permet le recrutement des protéines d'export. En plus de permettre un recrutement précoce de Mex67 au niveau des mRNPs, THO/TREX jouerait un rôle dans le maintien de la stabilité du génome en prévenant les évènements de recombinaison (Chavez et al., 2000).

Le complexe **TREX2** a été identifié par recherche de létalité synthétique avec un mutant du gène *YRA1*, yra1ΔRRM (Fischer et al., 2002). Ce complexe est formé de cinq protéines : Sac3, Thp1, Sus1, Cdc31 et Sem1 et est localisé dans la partie nucléaire interne du NPC. De la même manière que le complexe TREX, TREX2 prévient la formation de duplexes ADN/ARN.

Npl3 est une protéine codée par un gène essentiel, liant l'ARN et de la famille des protéines SR (riche en Sérine / Arginine). L'organisation de Npl3 est présentée dans la **Figure 1.07D**. De la même manière que Nab2, Npl3 est une protéine navette entre le noyau et le cytoplasme. Son activité et sa localisation sont régulées par le biais de modifications post-traductionnelles au niveau de la région C-terminale. Npl3 sous sa forme phosphorylée, via la kinase Sky1, est adressée au noyau et est recrutée très tôt par interaction avec SER^2 du CTD (Gilbert et al., 2001; Lei et al., 2001). Dans le noyau, après libération des mRNPs et dissociation du complexe de polyadénylation, Npl3 est déphosphorylée par Glc7, permettant l'interaction avec les mRNPs et le recrutement de Mex67 (Gilbert and Guthrie, 2004; Gilbert et al., 2001). La phosphorylation de Npl3 diminue son affinité pour l'ARN et pour l'ARN polymérase II. Sous sa forme déphosphorylé, Npl3 est adressée vers le cytoplasme.

Ces adaptateurs permettent l'interaction entre l'ARNm et Mex67/Mtr2 chez *S. cerevisiae* et ainsi l'export vers le cytoplasme, via les NPC. Cette dernière étape d'export est fortement régulée, tant au niveau de la formation et de la « qualité » des mRNPs que du transport en lui-même. Un défaut dans les mécanismes d'export des ARNm – mutations des facteurs d'exports ou de maturation – peut conduire au développement de cancers (modifications du complexe THO/TREX) (Siddiqui and Borden, 2012) ainsi que des pathologies plus graves comme la maladie des os de verre ou ostéogénèse imparfaite de type 1 ou encore de pathologies neuro-dégénératives (Arthrogrypose ou « *Lethal congenital*

contracture syndrome 1 ») liées à la rétention d'ARNm aberrants dans le noyau (Hurt and Silver, 2008).

A.2. Maturation de l'extrémité 3'

À l'exception des gènes codants pour les histones chez les métazoaires (Dominski and Marzluff, 2007), tous les ARNs messagers sont maturés à l'extrémité 3' par un même mécanisme, permettant l'ajout d'une queue poly(A) à l'extrémité 3' de chaque transcrit. La queue poly(A) peut être de différentes tailles (60 à 250 nt) selon l'organisme et l'ARNm (Sachs and Wahle, 1993).

La maturation de l'extrémité 3' est nucléaire, et regroupe deux réactions indissociables *in vivo* : le clivage du pré-ARNm au niveau du site de clivage et l'ajout de résidus adénosine au niveau de la nouvelle extrémité 3' formée, correspondant à la queue poly(A) (Wahle and Kuhn, 1997). *In vitro*, ces réactions peuvent être découplées, pour l'analyse du clivage, de la polyadénylation de transcrits synthétiques, comme *CYC1* (Butler and Platt, 1988). La première étape de la formation de l'extrémité 3' est la reconnaissance des éléments *cis* présents sur le pré-ARNm et le clivage par l'endonucléase. Cette dernière, Ysh1 chez *S. cerevisiae*, est une hydrolase zinc-dépendante de la super famille des métallo-β-lactamases. La réaction correspond à l'attaque nucléophile d'un hydroxyde conservé sur un groupement phosphate du substrat ARN en interaction avec des atomes de zinc. Une fois le transcrit clivé, la poly(A) polymérase catalyse l'ajout de 70 à 90 résidus adénosines. La poly(A) polymérase, Pap1 chez *S. cerevisiae*, est une poly(A) polymérase matrice indépendante (Lingner et al., 1991), dont le site actif est fortement conservé contrairement à la surface de la protéine, en comparaison aux autres polymérases identifiées. Le domaine catalytique est formé d'un domaine « *helix turn* » caractéristique de la superfamille des ADN polymérases β. L'enchaînement entre les deux étapes reste à ce jour peu connu. Une fois synthétisée, la queue poly(A) est raccourcie par différentes dé-adénylases, comme les complexes CCR4/NOT ou PAN (« *Pab1-dependant poly(A) nuclease* »). Cela permet de mâturer la taille de la queue poly(A) entre 50 et 90 nucléotides. L'action du complexe PAN est nécessaire au bon devenir des ARNm. Une inactivation de *PAN2* et *PAN3*, sous-unités du complexe PAN chez *S. cerevisiae* conduit à l'accumulation de messagers anormalement longs - 90 à 200 nucléotides A (Brown and Sachs, 1998). Le complexe CCR4/NOT est la dé-adénylase majeure chez la levure et est formé de 9 sous-unités dont Ccr4 et Pop2/Caf1 qui portent les

activités enzymatiques. Ce complexe est requis pour la maturation de la taille de la queue poly(A) ainsi que pour le *turnover* des ARNm (Collart and Panasenko, 2012). Chez *S. cerevisiae*, la formation de la queue poly(A) et le contrôle de sa taille nécessitent la fixation de deux protéines Pab1 et/ou Nab2 (Amrani et al., 1997a; Dheur et al., 2005; Hector et al., 2002).

Malgré l'apparente simplicité de la réaction, la maturation de l'extrémité 3' fait intervenir un ensemble d'éléments positionnés en *cis* (« *cis-acting element* ») et des facteurs protéiques agissant en *trans* (« *trans-acting factors* »), regroupés en un complexe de plus de 1 MDa, le complexe de polyadénylation. Chez *S. cerevisiae*, ce dernier regroupe vingt protéines ; chez les métazoaires, le complexe de polyadénylation est formé d'environ 85 sous-unités (Shi et al., 2009). La machinerie est détaillée dans le paragraphe B.

La maturation de l'extrémité 3' est fortement connectée avec la transcription, puisqu'une terminaison de transcription correcte requiert un signal de polyadénylation fonctionnel (Proudfoot, 1989). Cette interdépendance sera abordée en détail dans le reste du manuscrit. La queue poly(A) et les protéines associées sont impliquées dans des événements postérieurs comme la traduction – la queue poly(A) interagit avec eIF-4G, facteurs d'initiation de la traduction – ou la dégradation de l'ARNm dans le cytoplasme. La maturation en 3' joue un rôle dans la terminaison de la transcription, confère sa stabilité à l'ARNm et facilite l'export.

Le rôle de la polyadénylation dans la stabilité de l'ARNm a été identifié en 1975, par Huez *et al.* (Huez et al., 1975). La queue poly(A) permet de protéger l'ARN de la dégradation. En effet, la dépolymérisation de 5 à 15 résidus conduit à l'activation de deux voies de dégradation des ARNm – 5'- 3' via l'activation de Dcp2 et de l'exonucléase Xrn1 ou 3'- 5' via l'activation de l'exosome (Mitchell et al., 1997) puis l'élimination de la coiffe par DcpS.

L'interconnexion entre la polyadénylation et l'export vers le cytoplasme des ARNm peut être illustrée par différents exemples, comme le recrutement de Nab2 ou de Yra1.

A.3. Les autres mécanismes de maturation de l'extrémité 3'

A.3.1. ARNm codants pour les histones

Tous les ARNm sont maturés par l'ajout d'une queue poly(A) à l'exception des ARNs codant pour les histones, chez les métazoaires. Les ARNm codants pour les histones sont

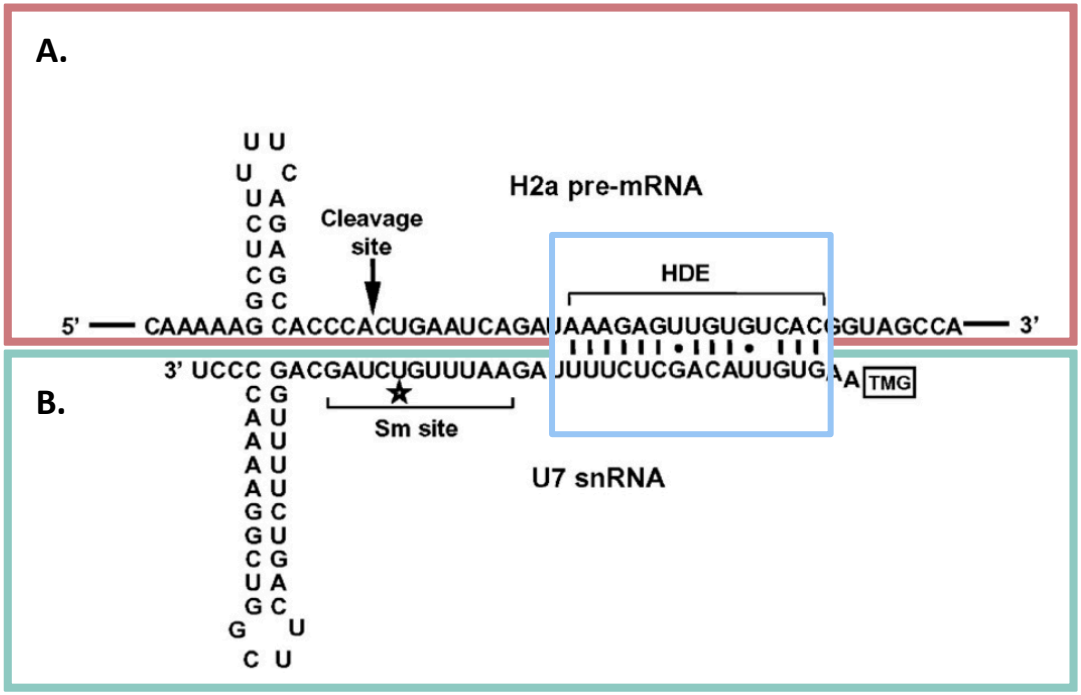


Figure 1.09 : Organisation des éléments *cis* des gènes codants pour les histones **(A.)** en présence du snRNA U7 **(B.)** – modèle murin. Les éléments *cis* de l'ARN (en **rouge**) nécessaire à la maturation des gènes codants pour les histones sont : la structure tige boucle (formée d'une séquence double brin de 6 nucléotides et d'une boucle de 4 nucléotides) le site de clivage représenté par une flèche (4 à 5 nucléotides de la structure tige boucle) et de l'élément HDE, complémentaire au snRNA U7 (en **bleu**). Le snRNA U7, maturé à l'extrémité 5' (en **vert**) permet le recrutement de protéines Sm au niveau du site Sm. D'après Dominski et Marzluff, 2007.

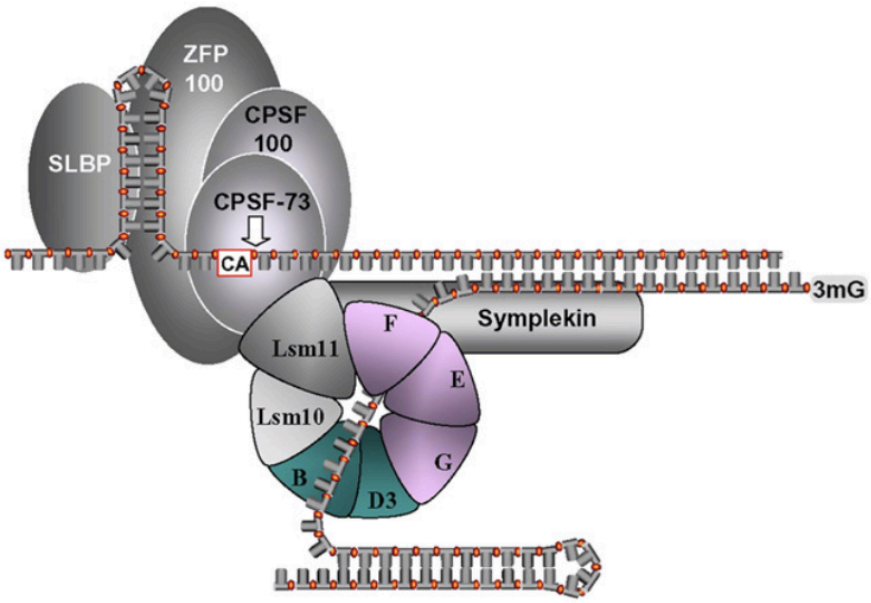


Figure 1.10 : Modèle d'organisation des éléments *cis* et des facteurs *trans* pour la maturation des ARN histones. Le site de clivage est représenté par le di-nucléotide CA. Le snRNA U7 permet le recrutement des autres facteurs, notamment les protéines Sm. CPSF-73 est l'endonucléase permettant le clivage. ZFP100 et SLBP permettent le maintien du complexe. L'organisation de CPSF-100 et la Simplekine sont hypothétiques. D'après Dominski et Marzluff, 2007.

clivés et polyadénylés chez *S. cerevisiae*. La majorité des protéines formant les histones sont codées par une série de gènes dont l'expression est réplication-dépendante ; ils ne sont exprimés que lors de la phase S du cycle cellulaire (Dominski and Marzluff, 2007). Certains variants, comme les sous-unités H3.3, H2aZ, H1, H3-cid et macroH2a sont codées par des ARNm polyadénylés. Les ARMm codant pour les cinq des protéines histones canoniques se terminent par une structure tige-boucle fortement conservée. Cette extrémité 3' est formée par le biais d'une machinerie spécialisée, qui reconnaît deux éléments *cis* présents sur le pré-ARNm : la **structure tige-boucle** elle-même et une séquence riche en résidus purine (A ou C), appelée **HDE** (« *Histone Downstream Element* ») (Dominski and Marzluff, 2007) (**Figure 1.09**). Le pré-ARNm est clivé entre ces deux éléments. Puisque les ARNm codants pour les histones ne présentent pas de séquences introniques, le clivage permet la formation d'ARNm mature. Ces ARNm sont ensuite exportés vers le cytoplasme pour y être traduits.

Le premier élément *cis* est la **structure tige-boucle**, fortement conservée, formée d'une séquence double brin de 6 nucléotides et d'une boucle de 4 nucléotides. Les ARNm ne présentant pas de structures tige-boucle à l'extrémité 3' sont dégradés. Le second élément *cis* est le **HDE**, dont la séquence est variable chez les eucaryotes, et qui est situé de 15 à 20 nucléotides en 3' de la structure tige-boucle. Chez *D. melanogaster* et *C. elegans*, la séquence consensus est ACCA ; ACCCA chez les vertébrés. La distance entre les deux éléments est importante puisqu'un décalage induit une diminution de la maturation (Scharl and Steitz, 1994). Le clivage a lieu entre ces deux éléments, en 3' du $4^{ième}$ ou $5^{ième}$ nucléotide chez les vertébrés. L'extrémité 3' de l'ARNm mature est formée d'une structure tige-boucle suivie d'une courte séquence simple brin.

L'un des premiers facteurs *trans* identifié comme nécessaire pour la maturation de ces ARNm est le **snRNA U7** (Dominski et al., 2003). Le snRNA U7 est le plus petit transcrit de l'ARN polymérase II, formé de 57 à 71 nucléotides en fonction des espèces (57 nucléotides chez les oursins, 63 nucléotides chez les mammifères et 71 nucléotides chez *D. melanogaster*). La région 5' est déstructurée et est complémentaire à la séquence HDE. Les mutations de la séquence HDE peuvent être compensées par les mutations du snRNA. La région 3' est formée d'une structure tige-boucle de séquence variable en fonction des organismes, et n'est pas nécessaire pour les activités de maturation (Gilmartin et al., 1988). Elle jouerait un rôle dans la stabilisation du complexe. La région centrale est formée d'une région d'interaction avec les **protéines Sm**, fortement conservée et de séquence

AAUUUGUCUAG. Le snRNA U7 permet le recrutement de sept protéines Sm – cinq (SmE, SmF, SmG, SmB et SmD3) appartenant au spliceosome et deux (Lsm10 et Lsm11) spécifiques de cette étape de maturation. Le snRNA U7 forme une plate-forme de recrutement pour le reste de la machinerie.

Le second facteur *trans* nécessaire pour la maturation des ARNm codants pour les histones est une protéine interagissant spécifiquement avec la structure tige-boucle de l'ARNm – **SLBP** ou ***Stem-Loop Binding Protein*** (Lanzotti et al., 2002). Les SLBP contiennent un domaine central d'interaction avec l'ARN. Cette interaction avec la structure tige boucle permet de stabiliser l'interaction du snRNA U7 avec le HDE (Dominski and Marzluff, 1999). Cette protéine reste associée à la structure tige boucle de l'ARNm mature. Cela permet de stimuler la traduction des ARNm codants pour les histones lors de la phase S du cycle cellulaire puis leur dégradation lors de la transition vers la phase G2.

ZFP100 est une protéine de 100kDa, possédant 18 motifs en doigt de zinc type **Cys_2His_2** ou chaque motif ont une fonction indépendante. **ZFP100** jouerait un rôle de protéine d'architecture, par interaction avec la SLBP, le snRNA U7 et les protéines Sm. L'endonucléase permettant le clivage est l'homologue de Ysh1/Brr5, **CPSF-73**. La réaction de clivage requiert le recrutement de **CPSF-100** (homologue de Ydh1/Cft2) et de **Simplekin** (homologue de Pta1). Ces deux protéines du complexe de clivage et polyadénylation permettraient (i) le recrutement de l'endonucléase, (ii) le maintien de l'organisation des différents facteurs *trans* et (iii) pourraient jouer un rôle de régulation, en régulant le recrutement de l'endonucléase.

Un modèle d'organisation des éléments *cis* et des facteurs *trans* a été proposé par Dominski et Marzluff (Dominski and Marzluff, 2007) (**Figure 1.10**).

A.3.2. Maturation des snoRNAs et snRNAs

L'ARN polymérase II est permet la synthèse de transcrits codants mais également de nombreux transcrits non codants comme les snoRNAs ou les snRNAs. Les snRNAs sont impliqués dans les réactions d'épissage et les 24 snRNAs chez *S. cerevisiae* sont codés par des gènes uniques (Bernstein and Toth, 2012). Les snoRNAs forment un groupe d'ARNs non codants, de taille variable (80 à 1000 nucléotides chez *S. cerevisiae*) impliqués dans la maturation des ARNr, transcrits par l'ARN polymérase I. Certains sont impliqués dans le clivage de l'ARN 35S mais la majorité d'entre eux permettent de guider les modifications

`

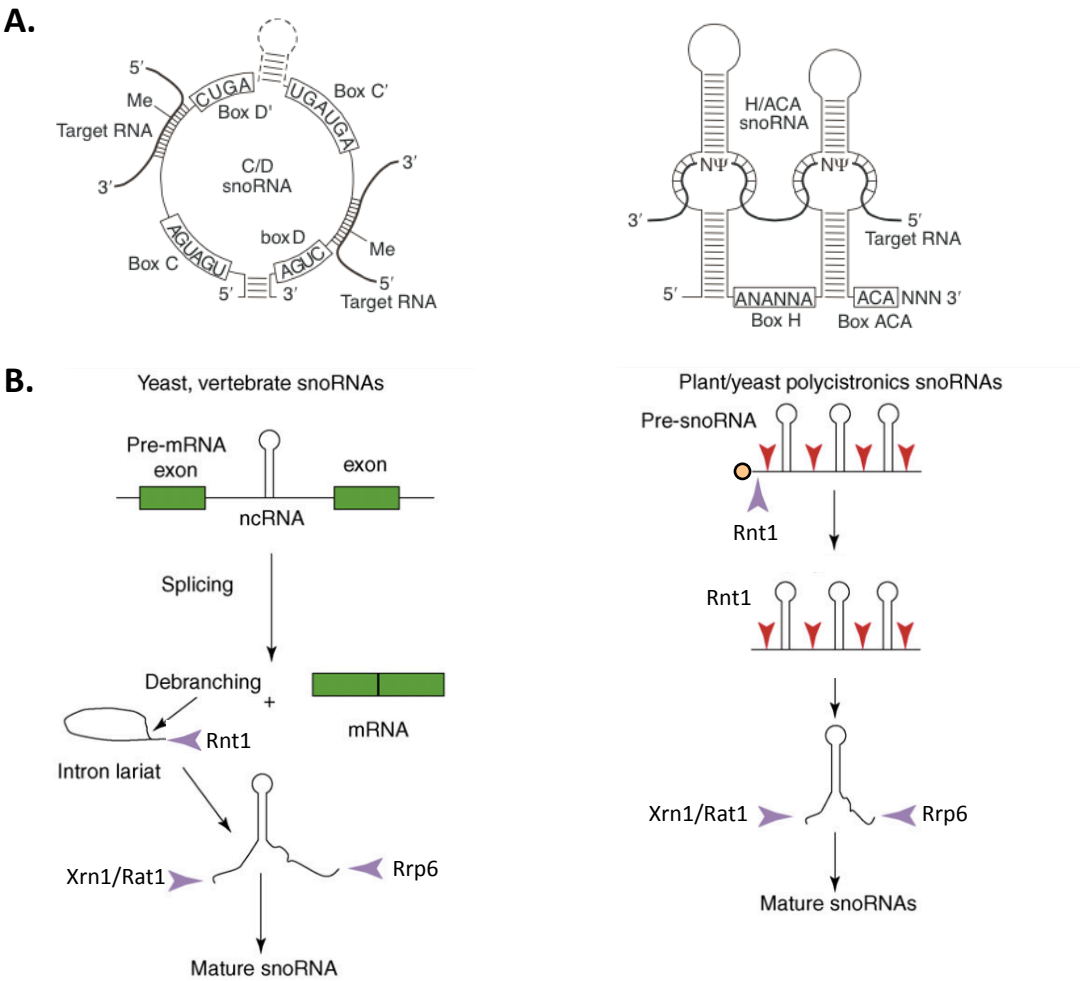


Figure 1.11 : **(A.)** Organisation des snoRNAs chez S. cerevisiae. Les snoRNAs à boîte C/D sont appariés avec deux ARNr cibles au niveau des séquences proches des boites D et D'. Les snoRNAs à boite H/ACA ne présentent pas de séquences consensus mais leur structure est conservée. Ils permettent l'appariement au niveau des structures tige-boucles avec un ARNr cible. Les modifications induites par les snoRNAs sont représentées (Met et NΨ).

(B) Maturation des snRNAs et des snoRNAs. *À Gauche*. Les sn/snoRNAs contenus dans les régions introniques sont maturés à partir des *lariats* formés lors de l'épissage. Les *lariats* sont ouverts par action de Rnt1 puis maturés par Rat1/Xrn1 et Rrp6. *À Droite.* Après transcription par l'ARN polymérase II, les sn/snoRNAs codée dans des unités transcriptionnelles sont clivés par Rnt1, ce qui permet l'élimination de la coiffe et confère un point d'entrée pour les exonucléases 5'-3' et 3'-5'. Rnt1 permet l'individualisation des sn/snoRNAs des unités polycistroniques. D'après Brown *et al*.. 2008.

post-transcriptionnelles des ARNr. Chez la levure, 80 snoRNAs ont été identifiés (Piekna-Przybylska et al., 2007) et la plupart proviennent d'unités transcriptionnelles mono ou polycistroniques (72/80) et une minorité est codée au sein des introns (8/80) (Kiss, 2002). Une étude menée par Kishore *et al.,* (Kishore et al., 2013) a montré par PAR-CLIP qu'il existerait un grand nombre de snoRNAs chez l'Homme, d'expression plus variable en comparaison à ceux identifiés. Les snoRNAs peuvent être séparés en deux classes, en fonction d'éléments de séquence conservés, dont l'organisation est présentée **Figure 1.11A** (Filipowicz and Pogacic, 2002). Chaque classe de snoRNAs possède une structure secondaire bien définie et s'associe très tôt avec un ensemble de protéines spécifiques pour former les snoRNPs correspondantes (Filipowicz and Pogacic, 2002).

- les snoRNAs à boite **C/D** sont requis pour les méthylation sur le 2'-O-ribose des ARNr. Ils présentent au minimum un motif C (RUGAUGA) et un motif D (UCUGA) et des motif moins conservés, C' et D'. Les snoRNAs à boite **C/D** interagissent avec Nop56, Nop58, Fibrillarin et une protéine de 15,5kDa.

- les snoRNAs à boite **H/ACA** sont requis pour la modification des uridines en pseudouridines des ARNr. Contrairement aux snoRNAs à boite C/D, ils ne présentent pas de motif consensus. Ils sont caractérisés par deux régions de séquences variables mais dont la position est conservée : H (ANANNA) et ACA. Leur structure secondaire – deux tiges-boucles – est conservée. Les snoRNAs à boite **A/ACA** interagissent avec Nhp2, Nop10, Gar1 et Dyskerin/Cbf5.

Bien que synthétisés par l'ARN polymérase II, les snoRNAs et snRNAs sont maturés par un mécanisme indépendant de la voie de polyadénylation utilisée pour les ARNs codants. Ils ne présentent pas d'éléments *cis* en aval des unités de transcription comme décrit précédemment. De plus, la présence d'une queue poly(A) interfère avec leur maturation (Fatica et al., 2000). Cette dernière fait intervenir plusieurs voies basées sur le même mode d'action : le précurseur contenant le snoRNA ou le snRNA est progressivement raccourci par des exocucléases (**Figure 1.11B**) (Brown et al., 2008).

Chez *S. cerevisiae*, les pré-snoRNAs sont tout d'abord clivés par Rnt1 - orthologue de la RNase III, possédant une activité d'endonucléase. Cela permet l'élimination de la coiffe à l'extrémité 5' et forme un point d'entrée pour d'autres exonucléases. Pour les pré-snoRNAs présents dans les régions introniques, on parle de « *debranching* ». Les précurseurs seraient ensuite individuellement dégradés par Rat1 et Xrn1, dans le sens 5'-3' (Tollervey, 2004) et

par Rrp6, sous-unité de l'exosome, dans le sens 3'-5' (Allmang et al., 1999). Les pré-snRNAs sont clivés par Rnt1 mais nécessitent l'action de l'hélicase Dob1 (Allmang et al., 1999) et sont ensuite dégradés par Rat1, Xrn1 et Rrp6. L'action des exonucléases est arrêtée lorsque celles-ci rencontrent les protéines formant les snoRNPs et snRNPs.

La maturation des snoRNAs et snRNAs est poly(A) indépendante mais fait intervenir un mécanisme similaire : la présence d'éléments *cis* et l'utilisation de facteurs *trans*. Le clivage avec Rnt1 nécessite la définition d'un site de clivage, qui se fait via trois éléments *cis* situés dans la région en amont du snoRNA entourant le site de clivage (région de 150 à 200 nucléotides). On distingue la séquence UGUAAAAU, une répétition d'éléments CUU (uniquement pour les précurseurs des boites C/D) et le site de clivage, défini par le dinucléotide AA ou AG. La coupure est réalisée en 3' du premier résidu A.

Plusieurs facteurs t*rans* sont également nécessaires pour la maturation des snRNAs et des snoRNAs. C'est le cas de Nrd1 (qui interagit avec la séquence UGUAAA), l'hélicase Sen1 qui interagit avec Rnt1 (Ursic et al., 2004), Nab3, la kinase Ctk1 et le CTD de l'ARN polymérase II (Steinmetz et al., 2001) ou encore des composants du complexe de polyadénylation pour la formation de l'extrémité 3' (Dheur et al., 2003; Morlando et al., 2002). En effet, de nombreux mutants des sous-unités - Pti1, Ref2, Pcf11, Clp1, Pta1, Swd2, Glc7 et Ssu72 (Nedea et al., 2008) - présentent des défauts de terminaison de transcription des snoRNAs. Nrd1 permettrait également de prévenir les interférences entre la transcription des snoRNAs/snRNAs et des ARNm au niveau des unités de transcription.

A.4. Polyadénylation alternative et sélection du site de polyA

Le choix du site de polyadénylation au niveau des ARNm est défini par l'affinité des éléments du complexe de polyadénylation avec les éléments *cis* de l'ARNm. Une modification des éléments *cis* ou une modification de l'expression des facteurs *trans* peut conduire à la sélection d'autres sites de polyadénylation. Des études récentes indiquent que la majorité des gènes présente plusieurs sites de polyadénylation (2,1 par gène), conduisant à la synthèse de plusieurs isoformes d'un même ARNm (Ozsolak et al., 2010). *HIS3* chez *S. cerevisiae* présenterait treize sites de polyadénylation potentiels (Mahadevan et al., 1997). Ces isoformes peuvent présenter une région codante différente et/ou une région 3'-UTR de taille variable (**Figure 1.12**). Cette polyadénylation alternative (**APA** pour « *Alternative PolyAdenylation* ») peut modifier le profil d'épissage et modifier l'accessibilité des séquences

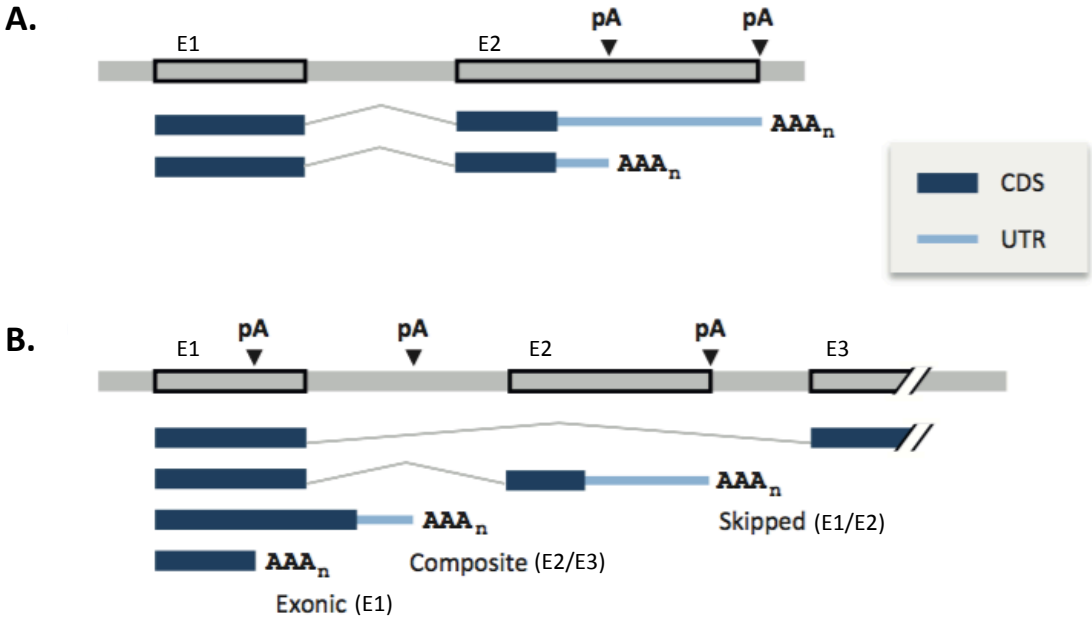


Figure 1.12 : Les isoformes formées par APA peuvent présenter une région codante différente et/ou une région 3'-UTR de taille variable. Les régions codantes des gènes sont représentées en gris. Les exons sont représentés en bleu. **(A.)** La sélection d'un site de polyadénylation (pA) alternative en aval ou en amont modifie la taille de la région 3'-UTR d'un même ARNm. **(B.)** Si les sites de polyadénylation alternative sont situés dans les régions introniques, cela permet de modifier la région codante. *Skipped* (E1/E2) : exclusion de l'exon final E3. *Composite* (E2/E3) : exclusion du premier exon. *Exonic* (E1) : le site de polyadénylation est situé dans l'exon. D'après Tian et Manley, 2013.

régulatrices, modifiant ainsi la stabilité, la localisation et l'expression d'un ARNm. Chez l'Homme, on estime que 50% des transcrits présentent une polyadénylation alternative, augmentant considérablement le potentiel codant du génome (Tian et al., 2005). Cette polyadénylation alternative est également très présente chez *D. melanogester* ou *C. elegans*, touchant plus de 50% des gènes. L'**APA** peut être tissu-spécifique, liée à un stade de différentiation ou stimulée par des signaux extracellulaires (Tian and Manley, 2013).

La **polyadénylation alternative**, en modifiant la taille de la région 3'-UTR, peut modifier la localisation et l'expression d'un ARNm. Les isoformes obtenues ont une séquence codante identique mais une région 3'-UTR de taille variable (**Figure 1.12A**). C'est le cas pour de nombreux ARNm impliqués dans le développement neuronal, comme le facteur neurotrophique issu du cerveau ou **BDNF**. L'isoforme courte est localisée dans le corps cellulaire des neurones, et l'isoforme longue est localisée au niveau des dendrites. Cette différence de localisation modifie l'efficacité de traduction, et module les fonctions synaptiques au niveau de l'hippocampe.

La région 3'-UTR est également la cible de nombreux facteurs protéiques liant l'ARNm et de nombreux ARNmi (50% des ARNmi chez l'Homme). Les cellules exprimant majoritairement l'isoforme court des gènes *IGF2BP1* (« *insulin-like growth factor 2 mRNA binding protein 1* ») ou *CCND2* (« *proto-oncogenes cyclin D2* ») ne possèdent pas la zone d'interaction pour miR15/16. Ces cellules sont plus sujettes à une prolifération non contrôlée et ont un potentiel oncogénique plus important que les cellules exprimant l'isoforme longue (Ueno et al., 1989). L'étude de la taille de la région 3'UTR dans certains cancers est utilisée comme marqueur de diagnostic.

L'**APA** peut également modifier la séquence codante d'un gène en permettant l'exclusion ou non d'un exon (**Figure 1.12B**). L'exemple le plus connu est le cas de chaînes lourdes d'immunoglobulines M ou IgM. Durant la différenciation des cellules B, l'utilisation d'un site de polyadénylation proximal au dépend du site distal conduit à la production de d'IgM libres au dépend d'IgM transmembranaires. Les protéines produites ont une séquence N-terminale commune et varient uniquement par l'exon C-terminal inclus : S pour les IgM libres et M1/M2 pour les IgM transmembranaires (Takagaki et al., 1996).

D'un point de vu mécanistique, l'**APA** reste très peu connue. Cependant, on sait que les facteurs du complexe de polyadénylation ont un rôle important dans ce mécanisme, comme PABPN1, CstF-64 ou hFip1 chez l'Homme. CstF-64 est une sous-unité du complexe

de polyadénylation qui interagit avec l'élément *cis* riche en GU. Lorsque CstF-64 est en quantité limitante, le site de polyadénylation proximal ne serait pas reconnu, et les IgM transmembranaires, incluant les exons M1/M2 seraient majoritairement synthétisées (Takagaki et al., 1996). Une expression plus importante de CstF-64 lors de la différenciation des cellules B permettrait la reconnaissance du site de polyadénylation proximal, l'inclusion de l'exon S et la production d'IgM libre (Takagaki and Manley, 1998). Une diminution de son expression est également observée lors de l'infection par des adénovirus, forçant l'expression des ARNs viraux. Récemment, hFip1 a été montrée comme nécessaire pour l'APA (Lackford et al., 2014). La polyadénylation alternative pourrait également être régulée par l'organisation de la chromatine, et les modifications post-traductionnelles des histones. L'homologue de Pcf11 chez *A. thaliana*, PCFS4 a été décrit comme impliqué dans la polyadénylation alternative de l'ARNm *FCA,* impliqué dans la transition entre état végétatif et développement reproductif (Xing et al., 2008).

Chez *S. cerevisiae*, il existe peu d'exemples de gènes dont l'expression est régulée par polyadénylation alternative. L'APA est observée en cas de changement de conditions de croissance – changement de sources de carbone – et formerait un mécanisme d'adaptation. C'est le cas de *CBP1*, *SIRI* ou encore *RNA14*, codant pour une sous-unité du complexe de polyadénylation (Sparks and Dieckmann, 1998). L'induction de la respiration conduirait à une augmentation de l'activité du complexe de polyadénylation, et une augmentation de la reconnaissance de sites de polyadénylation proximaux, plus faibles. Cela entraînerait la production d'isoformes de *RNA14* non traduits et une diminution du niveau général d'expression de *RNA14* – formant une boucle de régulation négative. L'activité du complexe de polyadénylation est diminuée, ce qui conduit à une diminution de la capacité de croissance des cellules et l'adaptation au milieu. Le choix du site de polyadénylation peut être également influencé par des facteurs *trans*, comme Hrp1 (Kim Guisbert et al., 2007), qui permet de « fixer » la machinerie au niveau d'un site de polyadénylation (Minvielle-Sebastia et al., 1998) ou encore Npl3. Il a été montré que Npl3 entrait en compétition avec Rna15 pour l'interaction avec l'ARN (Gross and Moore, 2001b). Dès lors, l'hypothèse proposée (Bucheli et al., 2007) est que Npl3 permet de réguler le choix du site de clivage, en masquant les sites de clivage plus faibles. Cela permet de moduler la réaction de maturation de l'extrémité 3' mais également la terminaison de transcription et le devenir de l'ARNm. De plus, Pta1 et Pcf11 ont été décrit comme impliqué dans la polyadénylation alternative de certains ARN messagers (Seoane et al., 2009).

A.5. La polyadénylation chez les procaryotes

Malgré la découverte de la première poly(A) polymérase (PAP) chez *E. coli* en 1962 (August et al., 1962), la polyadénylation des ARNs chez les organismes procaryotes reste très peu étudiée. De plus, la majorité des données obtenues proviennent d'études sur *E. coli* (Nakazato et al., 1975). Des études récentes s'intéressent à la polyadénylation chez *Bacillus subtilis* (*B. subtilis*) ou encore *Streptomyces coelicolor* (Mohanty and Kushner, 2011). Chez *B. subtilis*, il est estimé que 15 à 25% des ARNs totaux sont polyadénylés, contre 2% chez *E. coli*. Cependant, des études récentes indiquent que la polyadénylation serait un mécanisme plus fréquent – près de 90% des transcrits seraient modifiés à l'extrémité 3', par l'ajout de quelques résidus ou la formation de queue poly(A) de 15 à 30 nucléotides.

Deux enzymes ont été identifiées comme présentant des activités de polymérase (**Figure 1.13**). Contrairement au système eucaryote, ces enzymes ne sont pas nécessaires pour la viabilité cellulaire bien que la délétion du gène codant pour PAP I (*pcnB*) chez *E. coli* entraîne une diminution de la rapidité de croissance. La PAP I (pour « *Polyribonucleotide Adenylyl Transférase* ») est codée par le gène *pcnB* et permettrait la polyadénylation de près de 90% des ARNs chez *E. coli*. Elle permet l'ajout de 15 à 30 nucléotides en utilisant l'ATP comme substrat, sans spécificité quant à la nature de l'ARN (ARNm, ARNr). La PAP I permet également l'ajout d'une queue poly(A) sur les ARNt et ARNr mais d'une taille moins importante, de 6 à 8 résidus (August et al., 1962). Une surexpression de PAP I est toxique pour la bactérie (Cao and Sarkar, 1992). Elle conduirait à la polyadénylation d'ARNs qui ne sont pas les substrats naturels de PAP I comme des ARNt ou des snRNAs, modifiant leur conformation et leur activité. L'étude du mutant *pcnB* montre qu'il existe une autre activité de modification de l'extrémité 3' chez *E. coli.* Mohanty et Kushner (Mohanty and Kushner, 2000) ont permis d'attribuer cette activité à la PNPase. La PNPase est codée par le gène *pnp* et permet l'ajout d'une queue polynucléotidique poly(N), d'environ 40 résidus, contenant ≈ 50% de résidus adénosines. En présence de PAP I, la PNPase permet l'ajout de résidus préférentiellement C et U ; en absence de PAP I, les résidus guanines prévalent. Chez *B. subtilis*, l'analyse des ARNs totaux a permis de mettre en évidence des activités de PAP et également de PNPase. Cette dernière est codée avec le gène pnpA (Condon, 2003) ; le gène codant pour la PAP reste à ce jour inconnu. Une polyadénylation PNPase dépendante a également été mise en évidence dans l'embranchement des Cyanobactéries, ancêtre des chloroplastes (Rott et al., 2003).

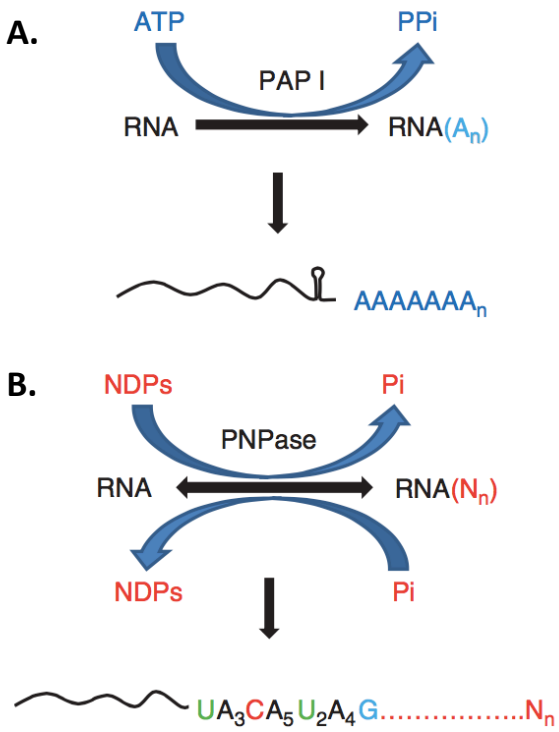


Figure 1.13 : Chez les procaryotes, deux enzymes ont été identifiées comme présentant des activité de polymérase.
(A.) *Haut*. La PAP1 permet l'ajout d'une queue poly(A) sur l'ARN cible. L'activité est dépendante de l'ATP et permet la libération de Ppi.
(B.) *Bas*. La PNPase permet l'ajout d'une queue poly(N) où N correspond à n'importe quel nucléotide, à partir de NDPs. L'activité de la PNPase permet la synthèse et la dégradation des queues poly(N), en fonction de la disponibilités de NDPs. Une forte concentration en NDPs et une faible concentration en Pi favorisent l'activité de synthèse de la PNPase.

D'après Mohanty et Kushner, en 2011.

Contrairement à la polyadénylation des ARNm chez les eucaryotes, il n'existe pas de signaux de polyadénylation à proprement parler. L'ajout de la queue poly(A), chez *E. coli*, dépendrait de la terminaison de transcription Rho-dépendante ou Rho indépendante (autrement appelée terminaison intrinsèque) (Yehudai-Resheff and Schuster, 2000). Dans la terminaison Rho-dépendante, le facteur Rho, hélicase ATP dépendante migre le long de la chaîne d'ARN permettant le décrochage de l'ARN polymérase. Dans la terminaison Rho-indépendante, l'ARN polymérase est décrochée par reconnaissance d'une séquence palindromique riche en AT ou terminateur intrinsèque. PAP I reconnaîtrait des séquences simples brin de 2 à 6 nucléotides laissées à l'extrémité 3' par une terminaison Rho-indépendante (72% des transcrits Rho-indépendant sont polyadénylés). Les transcrits Rho-dépendant présentent majoritairement des queues poly(N) ajoutées par la PNPase. De plus, il n'existerait pas de complexe de polyadénylation, comme chez les eucaryotes. En 2004, Mohanty *et al.,* (Mohanty et al., 2004) ont suggéré que la protéine Hfq, protéine liant l'ARN et apparentée à la famille des protéines Sm, jouerait le rôle d'adaptateur pour PAP I ; Hfq faciliterait l'interaction entre PAP I et son substrat en compétition avec la PNPase plus abondante.

La polyadénylation chez les procaryotes est faiblement utilisée et ne dépend pas de la quantité d'ARNs. De plus, la fréquence varie pour chaque famille de transcrits. Plusieurs paramètres sont à prendre en compte pour expliquer la faible polyadénylation des ARNs chez *E. coli*. PAP I est présent en faible quantité dans la cellule - 32 à 50 molécules par cellule – ce qui est dû à la présence d'un promoteur faible et d'un codon d'initiation non canonique faible. De plus, la localisation de PAP I est contrôlée durant le cycle cellulaire au niveau de la membrane. La fonction de ces modifications de l'extrémité 3' chez les organismes procaryotes reste peu étudiée. Cependant, des études suggèrent que l'ajout de queues poly(A) ou poly(N) permettrait de stimuler la dégradation des ARNs en déstabilisant les structures secondaires et en servant de contrôle qualité (O'Hara et al., 1995; Ryan et al., 2003). Selon le modèle proposé par Blum *et al.* (Blum et al., 1999), l'ajout d'une séquence nucléotidique simple brin à l'extrémité 3' permettrait d'initier l'action du dégradosome, dont fait partie la PNPase. Cependant, l'étude de la demi-vie de certains ARNm ne semble pas montrer de variation importante quant à leur stabilité dans les souches sauvages et *ΔpcnB*. De manière générale, l'ajout d'une queue poly(A) ou poly(N) permettrait de faciliter la dégradation et l'identification des ARNs a dégrader, permettant la fixation des PNPase, RNase II et RNase R, principales exoribonucléases chez *E. coli* ; on parle de « *docking plateforme* ».

A.6. TRAMP et Polyadénylation de dégradation

Dans le règne eucaryote, la polyadénylation est une étape indispensable permettant la maturation des ARNm. Les ARNm « défectueux » et les produits intermédiaires de réactions de maturation (régions introniques par exemple) sont dégradés par des machineries de dégradation, intervenant dans les voies de surveillance. Chez *S. cerevisiae*, deux grandes voies de dégradation sont décrites : la dégradation 5'-3', via Rat1 et la dégradation 3'-5', via l'exosome (Bernstein and Toth, 2012). Ce complexe est formé de dix sous-unités dans sa forme nucléaire (Mitchell et al., 1997) et d'un ensemble de cofacteurs permettant d'augmenter l'affinité de l'exosome pour l'ARN cible, comme Mpp6, Nop53, Nrd1/Nab3 ou Mtr4. Ces cofacteurs permettent le recrutement du cofacteur majeur de l'exosome, le complexe TRAMP. La dégradation par l'exosome fait intervenir une polyadénylation dite de dégradation, permise par les poly(A) polymérases non canonicales du complexe TRAMP. Le rôle de cette polyadénylation par TRAMP est donc opposé à la polyadénylation par Pap1, et est proche de la fonction supposée de polyadénylation dans le règne procaryote (Mohanty and Kushner, 2011).

Trf4 et Trf5 sont des protéines nucléaires, présentant une activité de poly(A)polymérase non canonicales (PAPnc). La délétion de *TRF4* peut être complémentée par la surexpression de *TRF5* (Castano et al., 1996). L'alignement de séquence en comparaison avec la Pap1 met en évidence des similarités de structures : le domaine catalytique est situé en N-terminal de la protéine, possédant trois résidus aspartates conservés et un domaine central conservé. *In vivo*, Trf4 et Trf5 font partie d'un complexe, capable de polyadényler les ARNs codants ou non codants (Vanacova et al., 2005). Cette polyadénylation, de 3 à 30 résidus, permet d'induire la dégradation des ARNs par l'exosome. Les PAPnc ont une faible affinité pour l'ARN, et acquièrent leur activité poly(A)polymérase en présence de Air1 et Air2, *zinc knuckle* protéines de type CCHC, qui interagissent avec l'ARN. Ce complexe, formé de Trf4/5, Air1/2, interagit physiquement et fonctionnellement avec Mtr4, hélicase ARN ATP-dépendante SF2 de la sous-famille des hélicases à boite DExH, décrite comme cofacteur de l'exosome. Mtr4 permettrait d'amener le complexe de dégradation 3'-5' au complexe initié par Trf4/5 et d'engager la dégradation (Losh et al., 2015). Mtr4 n'est pas fortement associé au complexe (dissociation en présence de 500mM NaCl (Bernstein and Toth, 2012)) ce qui suggère que cette protéine serait impliquée dans d'autres fonctions. Le complexe formé est appelé **TRAMP4** (Trf4, Air1/Air2, Mtr4) ou

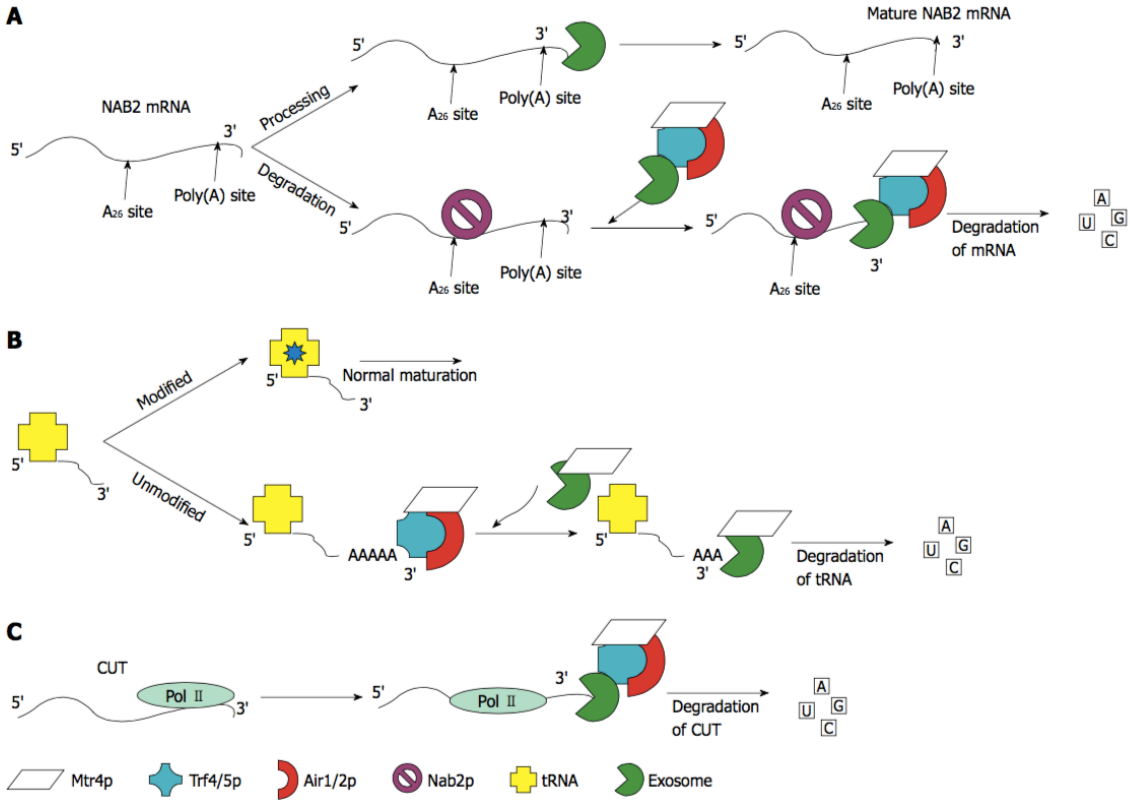


Figure 1.14 : Représentation des trois fonctions les mieux caractérisées du complexe TRAMP4/5. Le complexe TRAMP est formé de TRF4/5 (en **bleu**), Air1/Air2 (en **rouge**) et Mtr4 (en blanc) et permet une polyadénylation de dégradation par recrutement de l'exosome (en **vert**). D'après Bernstein et Toth, 2012.

(A.) Le complexe TRAMP permet la régulation de l'expression de *NAB2* (en **violet**). En condition normale, l'ARNm codant pour Nab2 est maturé et exprimé (« *processing* »). Lorsque Nab2 est en excès, Nab2 se fixe sur son propre ARNm, en amont du site du site de polyadénylation, au niveau d'une répétition de 26 résidus A (A_{26}). Cela permet le recrutement du complexe TRAMP, et la dégradation de l'ARN initiée par le complexe TRAMP (« *degradation* »).

(B.) Le complexe TRAMP est impliqué dans la surveillance des ARNt (en **jaune**). Certains ARNt nécessitent la méthylation d'un résidu pour acquérir une conformation correcte (« *modified* »). L'absence de méthylation (« *unmodified* ») et le défaut de conformation permettent le recrutement du complexe TRAMP, et la polyadénylation de l'ARNt. L'exosome est ensuite recruté et l'ARNt dégradé.

(C.) Le complexe TRAMP est impliqué dans la dégradation des ARN non codants comme les CUTs. Le recrutement et la dégradation se font co-transcriptionnellement lors de la synthèse par l'ARN polymérase II (en **vert**).

TRAMP5 (Trf5, Air1/Air2, Mtr4) ; le complexe TRAMP4 est supposé plus abondant que le complexe TRAMP5.

Le complexe TRAMP4/5 a été impliqué dans la polyadénylation de dégradation de nombreux ARNs : pré-ARNr, ARNt, ARNnc, snoRNAs, ainsi que les CUTs ou encore les ARNm (**Figure 1.14C**). Chaque complexe présente une spécificité d'action. Le TRAMP4 permet la polyadénylation des ARNm (absence du codon stop), des snoRNAs, des ARNt et des CUTs. Le complexe TRAMP5 permet la polyadénylation des pré-ARNr et des ARNnc (Porrua and Libri, 2015b). Le complexe TRAMP4 discrimine spécifiquement les ARNt hypo-méthylés et de conformation incorrecte, produit par l'ARN polymérase III (ARN_t^{Met}) (**Figure 1.14B**). Certains ARNt, comme ARN_t^{Met} nécessitent la méthylation d'un résidu (A58) pour acquérir une conformation native. L'absence de méthylation modifie l'architecture de l'ARNt, alors moins compacte, et les structures tige-boucles sont reconnues par les éléments du complexe capable de lier l'ARN, Air1 et Air2. Cette hypothèse, développée par Vanacova *et al.* (Vanacova et al., 2005), se vérifie en étudiant les ARN_t^{Ala} dont la conformation ne dépend pas de modifications post-transcriptionnelles et dont les formes non modifiées ou matures sont des substrats très faibles de Trf4. Le complexe TRAMP est également impliqué dans des fonctions indépendantes de son activité de poly(A) polymérase, comme la régulation de l'adaptateur d'export Nab2 (**Figure 1.14A**). Il interagit avec une répétition de 26 résidus A situés dans la région 3'-UTR de Nab2, en aval du site de clivage, permettant le recrutement de l'exosome et la dégradation des ARNm codants (Roth et al., 2009).

Les complexes TRAMP sont fortement régulés, pour éviter une utilisation excessive de l'ATP et une polyadénylation non spécifique, conduisant à la dégradation des transcrits. Une étude récente (Jia et al., 2011) montre que Mtr4 est un régulateur du complexe TRAMP via une activité de senseur des résidus adénosines. Lors des étapes initiales, l'ajout de résidus adénosine est rapide, puis la progression des PAPnc est ralentie après l'ajout de cinq résidus. Mtr4 « détecte » la taille de la queue poly(A) par interaction et permet de diminuer l'affinité de la PAPnc pour l'ATP et l'ARN, conduisant à un ralentissement de la synthèse de la queue poly(A). La rôle de régulateur de Mtr4 est indépendant de sa fonction d'hélicase.

A.7. Les maladies associées

La maturation de l'extrémité 3' est un processus biochimique complexe faisant intervenir de nombreux éléments de séquences et protéiques. Comme toute étape d'expression du génome, un défaut de maturation en 3' entrave la progression des autres étapes, et peut affecter la viabilité cellulaire, la croissance ou le développement.

La mutation d'un élément *cis* est notamment responsable de l'α-thalassémie. Chez l'Homme cette maladie est caractérisée par une faible production de globine α, indispensable pour l'assemblage de l'hémoglobine et l'érythropoïèse. La mutation du signal AAUAAA en AAUAAG conduisant à un défaut de reconnaissance, et un défaut de maturation en 3' ; les ARNm obtenus sont faiblement traduits (Higgs and Weatherall, 1983). La mutation des éléments *cis* peut également conduire à l'inhibition totale de l'expression d'un gène, comme c'est le cas pour gène GalA-α, conduisant à l'accumulation de sphingolipides et au développement de la maladie de Fabry (Danckwardt et al., 2008). À l'inverse, la thrombophilie est une prédisposition génétique au développement de thromboses, causée par une mutation au niveau du site de clivage de l'ARN codant pour un facteur de coagulation. Cela permet d'augmenter l'efficacité de la maturation en 3' et l'expression de ce gène. Enfin comme décrit dans le paragraphe A.4., une modification dans la sélection du site de polyadénylation – lors de la polyadénylation alternative – de facteurs oncogéniques peut modifier leur régulation, et ainsi favoriser la prolifération cellulaire et conduire du développement de cancers (Ueno et al., 1989).

Une dérégulation touchant les facteurs *trans* peut également être à l'origine de différentes pathologies. Dans les cellules humaines, une délétion de CstF-64 conduit à un arrêt du cycle cellulaire et l'apoptose de cellules (Takagaki and Manley, 1998). Une modification à la hausse de l'expression de la poly(A) polymérase PAP est également reliée au développement de cancers du sein, cancers du colon ou du pancréas (Pendurthi et al., 1997), ainsi que des leucémies. La surexpression de *PAP* est indirecte et serait due à la fixation du facteur TF (« *Tissue factor* ») ai niveau de récepteur cellulaire entrainant l'activation d'une cascade de signalisation activant les facteurs VIIa (« *clotting factor* ») et VIIIa (« *binding factor* »), qui induirait à leur tour une modification de l'expression de certains gènes dont *PAP*. Le taux de surexpression de l'enzyme est directement relié à l'agressivité de la pathologie et sert de bio-marqueurs de pronostiques. BARD1 interagit avec CstF-50, en réponse à une altération de l'ADN, permettant de bloquer la polyadénylation et la

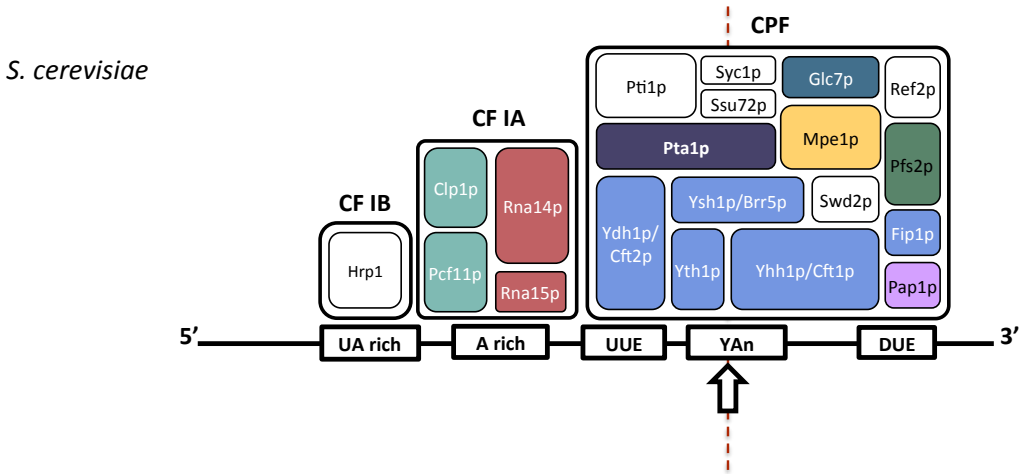


Figure 1.15 : Organisation du complexe de polyadénylation chez *S.cerevisiae*. Le complexe de polyadénylation fait intervenir des éléments *cis* au niveau de l'ARN (carrés noirs) et des facteurs *trans*. Ces derniers sont regroupés en trois sous-complexes : le complexe CF IA (« *Cleavage Factor IA* » - Rna14 ; Rna15, Clp1, Pcf11), le complexe CPF (« *Cleavage and Polyadenylation Factor* ») et le complexe CF IB (« *Cleavage Factor IB* » - Hrp1/Nab4). Le site de clivage/polyadénylation est défini par la séquence $Y(A)_n$. Les sous-unités présentant des homologues avec le système mammifère sont représentés en couleur. D'après Millevoi et Vagner, 2010.

maturation d'ARN de ce locus. La progression de certaines tumeurs est directement liée à la mutation du facteur BARD1. Cela permet l'expression de zones endommagées de l'ADN et conduit à une dérégulation du transcriptome de la cellule (Kleiman and Manley, 2001). Des maladies génétiques sont également imputées à des mutations des facteurs *trans* de la maturation de l'extrémité 3' comme la dystrophie musculaire oculo-pharyngée (OPMD), où la mutation de PABPN1 conduit à son auto-association et la formation d'agrégat dans les muscles squelettiques.

B. COMPOSITION DU COMPLEXE DE POLYADENYLATION

B.1. *Saccharomyces cerevisiae*.

Le complexe de polyadénylation, présentant les éléments *cis* et les facteurs *trans*, chez *S. cerevisiae* est présenté dans la **Figure 1.15** (Millevoi and Vagner, 2010).

B.1.1. Éléments *cis*

Les éléments *cis* sont des séquences retrouvées dans la région 3'-UTR du pré-ARNm, comprenant le site de clivage, permettant de guider et de positionner la machinerie (Proudfoot, 1976). Ces éléments sont retrouvés sur l'ensemble des pré-ARNm polyadénylés, à l'exception des pré-ARNm codants pour les histones. Chez *S. cerevisiae*, les éléments *cis* sont moins conservés en comparaison au système mammifère, et diffèrent par la séquence et l'organisation sur le pré-ARNm. L'organisation des éléments *cis* et la machinerie des deux systèmes sont comparées dans le paragraphe B2.2. On peut cependant distinguer des séquences consensus, ce qui traduit une pression de sélection importante (Proudfoot and Brownlee, 1976).

- L'« *efficiency element* » ou **EE** (ou *AU-rich*) est situé à une distance variable (24 à 40 nucléotides jusqu'à 75) en amont du site de clivage. Par analyse bioinformatique et par mutagénèse sur près de 1300 gènes, il a été permis de définir l'hexamère UAUAUA comme séquence consensus – 52% des cas (Guo and Sherman, 1996). Les séquences UACAUA et UAUGUA sont également fortement distribuées. Les U en positions 1 et 5 sont les plus conservées (Graber et al., 1999). L'*efficiency element* permet l'interaction avec Hrp1 et stimule la réaction.

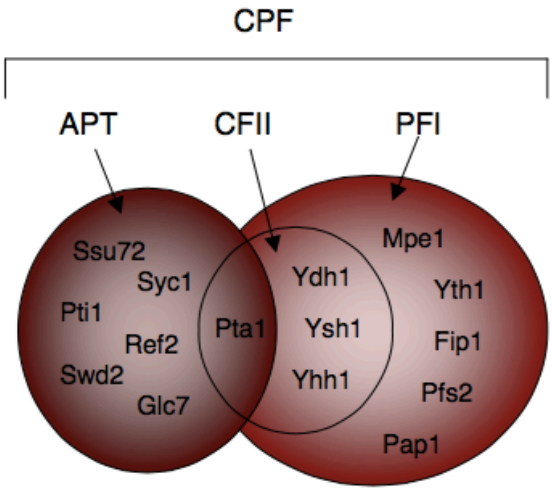


Figure 1.16 : Le sous-complexe CPF peut-être divisé en trois sous-parties en fonction de leurs interactions. **CF II** (« *Cleavage Factor II* ») porte l'activité de clivage - Ydh1/Cft2, Yhh1/Cft1, Ysh1/Brr5 et Pta1 (Zhao et al., 1997). **PF I** (« *Polyadenylation Factor I* ») porte l'activité de clivage, de poly(A) polymérase et les facteurs associés à sa régulation - Mpe1, Yth1, Fip1, Pfs2, Pap1 + CFII (Chen and Moore, 1992). **APT** (« *Associated with PTa1* ») regroupe les protéines interagissant avec Pta1, protéine d'architecture - Ssu72, Syc1, Pti1, Ref2, Swd2 et Glc7 (Nedea et al., 2003). Certains facteurs appartiennent aux complexes CF II ou PFI, qui forment une unité fonctionnelle *in vivo* (Ohnacker et al., 2000).

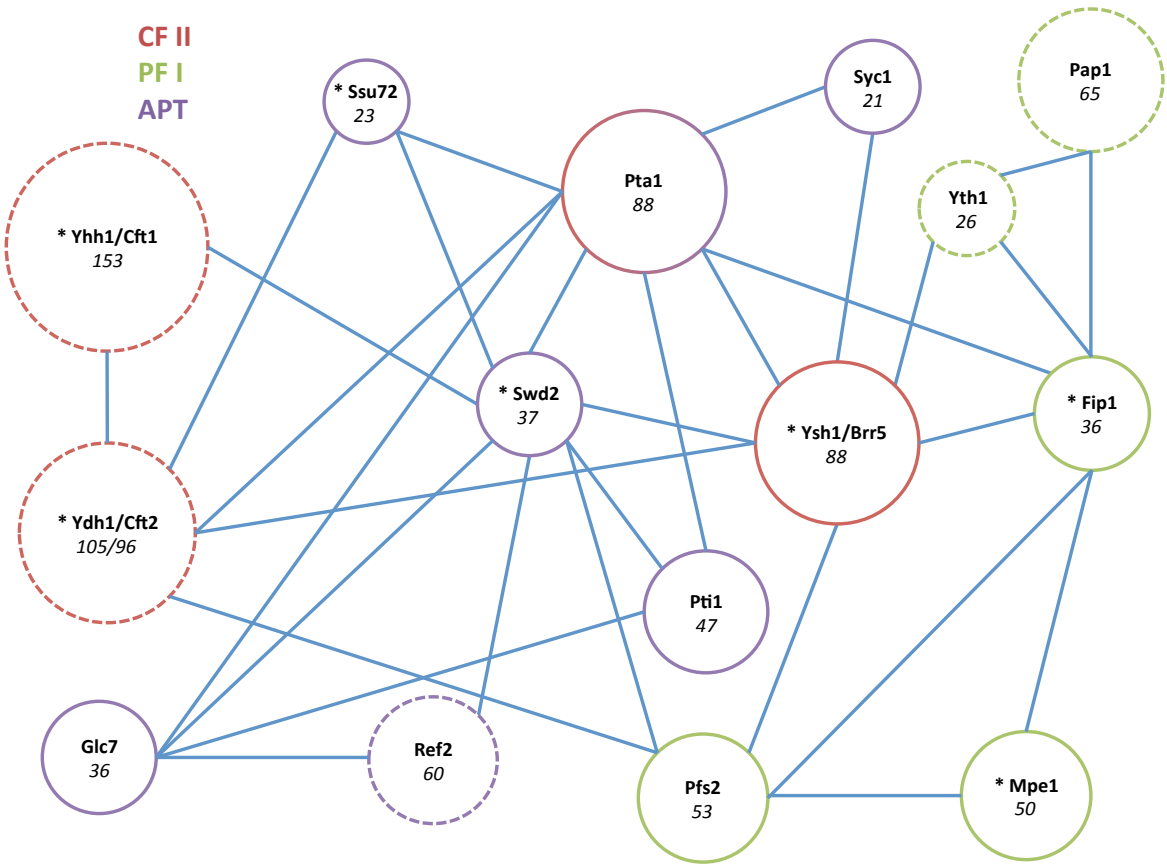


Figure 1.17 : Représentation des interactions entre les sous-unités du complexe CPF. Les interactions directes protéine-protéine, décrites dans la littérature sont représentées en **bleu**. L'appartenance pour chacune des sous-parties est représentée par un code couleur : **CF II**, **PF I** et **APT**. Le diamètre de chaque cercle est proportionnel à la taille des protéines, donnée en kDa. Les sous-unités interagissant avec l'ARN sont représentées en pointillées.

- Le « *positioning element* » ou **PE** est situé de 10 à 30 nucléotides en amont du site de clivage. C'est une séquence riche en A (Russo et al., 1991) ; les séquences UUAAGAAC, AAUAAA et AAAAAA sont les plus conservées (Graber et al., 1999). L'hexamère AAUAAA confère la meilleure efficacité de réaction *in vitro*. La majorité des mutations du PE ne modifie pas l'efficacité de la maturation en 3' mais restreint la localisation du site de clivage. Le PE permet l'interaction avec Rna15, sous-unité du sous-complexe CFIA (Gross and Moore, 2001b). Les PE et EE sont indissociables.
- le **site de clivage** est généralement défini par une séquence contenant un résidu pyrimidine Y suivi d'une répétition de A : $Y(A)_n$ (Graber et al., 1999). Le clivage par l'endonucléase Ysh1/Brr5 intervient en 3' d'un résidu adénosine.
- les éléments **riche en U** (**UUE** – *Upstream U-rich Element* - et **DUE** – *Downstream U-rich Element*) sont situés immédiatement (10 nucléotides ou moins) en amont et en aval du site de clivage. Ils participent à l'efficacité du clivage (Dichtl and Keller, 2001)

B.1.2. Facteurs *trans*

Les études génétiques et biochimiques, ainsi que l'étude à partir d'extraits nucléaires (Butler and Platt, 1988) suggèrent qu'un grand nombre de protéines sont nécessaires pour les activités de maturation en 3'. Chez *S. cerevisiae*, le complexe de polyadénylation regroupe vingt protéines recrutées en partie dès le début de la transcription (CF IA, CF IB et PFI), à l'extrémité 5' du gène (Licatalosi et al., 2002). Toutes les protéines sont nécessaires pour le clivage et la polyadénylation, mais peuvent avoir un effet sur d'autres mécanismes comme l'épissage (Ysh1/Brr5), l'initiation (Glc7, Ssu72) ou la terminaison de la transcription (Yhh1/Cft1, Swd2, Pta1, Ref2 …). La machinerie peut être divisée, chez la levure, en trois sous-complexes : **CPF** (« *Cleavage and Polyadenylation Factor* »), **CF IA** (« *Cleavage Factor IA* ») et **CF IB** (« *Cleavage Factor IB* »).

B.1.2.1. CPF

Le CPF regroupe la majorité des composants du complexe de polyadénylation et porte les activités de clivage (l'endonucléase Ysh1/Brr5) et de polyadénylation (Pap1). Le CPF peut être séparé en trois sous-complexes : **CF II** (« *Cleavage Factor II* »), **PF I** (« *Polyadenylation Factor I* ») et **APT** (« *Associated with PTa1* ») (Nedea et al., 2003) (**Figure 1.16**). À ce jour, les sous-unités Ref2, Swd2, Ssu72, Syc1 et Pti1 ne présentent pas d'homologues nécessaires aux activités du complexe de polyadénylation, chez les métazoaires. Un homologue

fonctionnel a pu être identifié pour les sous-unités Pfs2, Glc7 et Mpe1, mais leur fonction reste peu caractérisée (Shi et al., 2009). Les sous-unités du CPF interagissent spécifiquement formant un réseau d'interactions complexe résumé dans la **Figure 1.17**.

B.1.2.1.1. Yhh1/Cft1

Yhh1/Cft1 est une protéine de 153kDa, codée par un gène essentiel chez *S. cerevisiae*. Stumpf et Domdey ont identifié ce facteur en 1996 (Stumpf and Domdey, 1996) par homologie avec CPSF-160 chez les mammifères. **Yhh1/Cft1** présente 24% d'identité et 51% de similarité avec CPSF-160. Elle joue un rôle dans le clivage et la polyadénylation, puisque l'immuno-déplétion dans des extraits totaux abolit ces deux étapes (Stumpf and Domdey, 1996). Cependant l'analyse de la distribution globale de la taille de la queue poly(A) suggère que **Yhh1/Cft1** serait impliquée dans la définition de site poly(A) (Dichtl et al., 2002b), par son interaction avec les régions riches en U du pré-ARNm. **Yhh1/Cft1** permettrait de renforcer l'interaction du complexe de polyadénylation au niveau du site de clivage et ainsi la polyadénylation des ARNm présentant une reconnaissance moins spécifique du site de poly(A). **Yhh1/Cft1** serait également impliquée dans la terminaison de transcription par l'ARN polymérase II, via l'interaction avec le CTD de Rpb1. **Yhh1/Cft1** interagit avec des sous-unités de CPF (Ydh1 et Swd2) et également avec des sous-unités de CF IA (Rna14 et Pcf11) (Cheng et al., 2004; Kyburz et al., 2003). Récemment, il a été montré que la reconnaissance de la séquence AAUAAA était permise par WDR33 et non par CPSF-160, son homologue chez l'Homme (Chan et al., 2014; Schonemann et al., 2014).

B.1.2.1.2. Ysh1/Brr5

Ysh1/Brr5 est une protéine de 88kDa codée par un gène essentiel chez *S. cerevisiae* (Chanfreau et al., 1996). **Ysh1** est l'endonucléase responsable du clivage, métallo-β-lactamase de la famille des β-CASP – « metallo-β-lactamase associated CPSF Artemis SNM1/PSO2 » (Callebaut et al., 2002) – identifiée par homologie de séquence avec **CPSF-73** dans le système mammifère (Goffeau et al., 1996). La majorité des protéines β-CASP sont impliquées dans des interactions avec des acides nucléiques ou des étapes de maturation. La région C-terminale de la protéine présente une forte homologie de séquence (38%) avec une autre sous-unité du complexe CPF, Syc1, et est essentielle pour la viabilité cellulaire (Zhelkovsky et al., 2006). La plupart des données fonctionnelles sont obtenues par étude de son homologue **CPSF-73**. La structure de ce dernier, résolue par Mandel *et al*. (Mandel et al., 2006),

confirme l'appartenance à la famille des β-CASP et l'interaction de CPSF-73 avec le site de clivage, par interaction spécifique avec l'hexamère AAUAAA. L'analyse du site actif confirme l'interaction avec deux atomes de zinc, requis pour les réactions de clivage (Ryan et al., 2004). **Ysh1/Brr5** semble également impliquée dans l'épissage bien que les mécanismes ne soient pas connus (Chanfreau et al., 1996). Ysh1 interagit avec Clp1, Pcf11, Ydh1 (Kyburz et al., 2003), Pta1, Fip1, Syc1 (Zhelkovsky et al., 2006), Yth1 (Barabino et al., 2000) et Pfs2 (Ohnacker et al., 2000).

B.1.2.1.3. Ydh1/Cft2

Ydh1/Cft2 est une protéine de 105kDa, codée par un gène essentiel chez *S. cerevisiae* (Preker et al., 1997), caractérisée par Zhao *et al.*, en 1997 (Zhao et al., 1997). Les mutations conditionnelles de cette protéine (*ydh1-1*, *ydh1-2* et *ydh1-3*) modifient le clivage et l'ajout de la queue poly(A). **Ydh1** est requis pour les étapes de clivage et de polyadénylation *in vitro*, et permet également la reconnaissance du site de clivage du pré-ARNm (Dichtl and Keller, 2001; Kyburz et al., 2003). Ydh1/Cft1 possède 24,4% d'identité et 43% de similarité avec le facteur **CPSF-100**. La structure obtenue par Mandel *et al.,* (Mandel et al., 2006) présente des similarités avec l'endonucléase Ysh1/Brr5. Cependant, le motif β-métallo-lactamase ne présente pas de motifs liant le zinc ce qui ne permet pas une activité du même type que Ysh1/Brr5. **Ydh1/Cft2** peut s'auto-assembler via la région C-terminale et interagit avec Yhh1/Cft1, Ysh1/Brr5, Pta1, Pfs2, Ssu72, Pcf11 ainsi que le CTD de l'ARN polymérase II (Kyburz et al., 2003). Le profil de phosphorylation du CTD nécessaire pour l'interaction avec **Ydh1/Cft2** n'est pas connu. **Ydh1/Cft2** présente de nombreuses sérines/thréonines potentiellement phosphorylables, ainsi que deux séquences potentielles de localisation nucléaire (Zhao et al., 1997).

B.1.2.1.4. Yth1

Yth1 est une des plus petites sous-unités du CPF (26kDa) identifiée en 1997 par Barabino *et al.* (Barabino et al., 1997), codée par un gène essentiel et présentant 40% d'identité avec son homologue chez les mammifères, **CPSF-30**. Bien que certaines études (Murthy and Manley, 1992) aient suggéré que **CPSF-30** n'était pas requise pour la maturation de l'extrémité 3', des études récentes sur CPSF-30 (Chan et al., 2014) ainsi que l'analyse de mutants conditionnels (*yth1-7*, *yth1-8* et *yth1-4*) montrent que **Yth1** est requise pour le clivage et la polyadénylation (Barabino et al., 2000) et interagit préférentiellement avec les séquences

ARNs riches en U. **Yth1** contient cinq motifs liant le zinc de type CCCH impliqués dans l'interaction avec le pré-ARNm (ZF2 et ZF4) et dans des interactions protéines-protéines (Fip1 – ZF4 et ZF5, Pap1 et Ysh1 - ZF1 et ZF4). Les motifs ZF1 et ZF2 sont requis pour l'activité de clivage et les motifs ZF2, ZF4 et ZF5 sont nécessaires pour la polyadénylation (Barabino et al., 1997; Tacahashi et al., 2003). L'homologue de **Yth1** chez les eucaryotes supérieurs présente également un motif CCHC qui permettrait d'améliorer d'interaction avec l'ARN (Chan et al., 2014).

B.1.2.1.5. Fip1

Fip1 est une phosphoprotéine de 36kDa – cible de CK2 (Zielinski et al., 2006) – possédant 48% d'identité et 72% de similarité avec la sous-unité **hFip1** (66kDa – Kaufmann et al., 2004). Fip1 est plus petit que son homologue hFip1 de part l'absence du segment RD – alternance de résidus arginine et acide aspartique. Cette protéine a été identifiée en 1995, grâce à son interaction avec la polymérase **Pap1** et permettrait d'amener la polymérase au complexe CPF et au site de polyadénylation (Ezeokonkwo et al., 2011; Preker et al., 1995). L'hypothèse développée par Helming *et al.* (Helmling et al., 2001) est que l'activité de Pap1 est régulée par son interaction avec Fip1. La région C-terminale de **Fip1** (résidus 80 à 105) est complémentaire d'un domaine d'interaction à l'ARN de Pap1. L'interaction régulerait son affinité, permettant de réguler son activité (Zhelkovsky et al., 1998). La déplétion de Glc7 conduit à une accumulation de formes phosphorylées de **Fip1**, et à sa désolidarisation partielle du complexe CPF (He and Moore, 2005). **Fip1** interagit avec Pap1 et Rna14 (Preker et al., 1995) et avec Yth1, Pfs2, Ysh1/Brr5, Pta1 et Mpe1 (Barabino et al., 2000; Gavin et al., 2002; Ohnacker et al., 2000).

B.1.2.1.6. Pap1

Pap1 est une protéine de 65kDa, formée de trois domaines globulaires, requise uniquement pour l'étape de polyadénylation chez *S. cerevisiae* (Bard et al., 2000). Les 450 premiers résidus comprenant le domaine catalytique sont fortement conservés dans le règne eucaryote – 47% d'identité avec PAP – la région C-terminale n'est pas conservée (Martin and Keller, 1996). L'activité catalytique est portée par la région N-terminale de la protéine. La région C-terminale centrale interagit avec le pré-ARNm. **Pap1** catalyse l'ajout, à l'extrémité 3' de l'ARN clivé, de queues poly-adénosines sur la majorité des pré-ARNm. Son activité est améliorée lorsque l'ion Mg^{2+} est remplacé par un ion Mn^{2+} (Lingner et al., 1991).

Pap1 est active en absence des autres partenaires du complexe CPF (activité non spécifique). Son activité est régulée par l'interaction avec le domaine C-terminal de **Fip1**, qui permet de contrôler son accès au substrat ARN (acquisition de sa spécificité) et ainsi l'ajout de résidus adénosines – mode distributif, plus lent et moins efficace. La structure de **Pap1** en présence de MgATP et d'une séquence oligo(A) a permis de montrer que seul les trois derniers résidus du pré-ARNm interagissent avec l'enzyme. **Pap1** est phosphorylée et ubiquitinée durant le cycle cellulaire (Mizrahi and Moore, 2000). Pap1 interagit avec Yth1 et Fip1 (Preker et al., 1995).

B.1.2.1.7. Pta1

Pta1 est une protéine de 88kDa, codée par un gène essentiel, identifiée comme facteur essentiel pour l'épissage des ARNt ; la mutation *pta1-1* conduit à l'accumulation d'ARNt contenant des régions introniques (O'Connor and Peebles, 1992). **Pta1** présente une faible homologie avec la **Simplekin** chez les mammifères - 17% identité et 31% similarité (Takagaki and Manley, 2000). Elle présente cependant une fonction similaire dans l'architecture du complexe CPF. En présence de mutations conditionnelles de *PTA1*, Zhao *et al*., (Zhao et al., 1999b) décrivent un recrutement moins important des protéines de CF II (Ydh1/Cft2, Yhh1/Cft1 et Ysh1/Brr5) en présence des allèles *pta1-1* et *pta1-2*. **Pta1** interagit également avec Glc7, Pti1 (He and Moore, 2005), Syc1 (Zhelkovsky et al., 2006), Fip1, Ssu72 (Ghazy et al., 2009), Pcf11 et Clp1 (Zhelkovsky et al., 2006). La région C-terminale de la protéine est impliquée dans les activités de polyadénylation (Preker et al., 1997) et de clivage (Zhao et al., 1999b). Ces activités sont modulées par des cycles de phosphorylation / déphosphorylation par CK2 et Glc7 (He and Moore, 2005). En 2009, la région N-terminale de **Pta1** a été décrite comme nécessaire pour la terminaison de transcription des snoRNAs, et la déphosphorylation de la SERINE 5 du CTD par Ssu72. L'absence de Pta1 (« *degron strain* ») diminue la quantité de Ssu72 et conduit à une augmentation de la phosphorylation de SERINE 5 (Ghazy et al., 2009). Pta1, avec Pcf11 est impliquée dans la polyadénylation alternative de certains ARNs messagers (Seoane et al., 2009).

B.1.2.1.8. Glc7

Glc7 est une protéine de petite taille (36kDa), codée par un gène essentiel, impliquée dans la régulation de nombreux processus biologiques, de par son activité de sérine/thréonine phosphatase et sa faible spécificité (Feng et al., 1991; Stark, 1996). Une déplétion *in vivo* de

Glc7 conduit à une raccourcissement de la taille des queues poly(A) (He and Moore, 2005). Récemment, Schreieck *et al*. (Schreieck et al., 2014) ont montré que **Glc7** était responsable de la déphosphorylation de TYR[1] du CTD lorsque la Pol II atteint la région 3'-UTR du gène, permettant la transition entre élongation et terminaison (Nedea et al., 2008). L'absence de **Glc7** (« *degron strain* ») induit un défaut de terminaison de transcription des snoRNAs (snR13 – Nedea et al., 2008). **Glc7** est également impliquée dans l'export, par déphosphorylation de **Npl3**, adaptateur de la protéine Mex67. La déplétion de Glc7 conduit à l'accumulation de Npl3 dans le noyau et une accumulation de Rna15 sur l'ARNm polyadénylé (Gilbert and Guthrie, 2004). **Glc7** participe également à la régulation l'endocytose, la sporulation, la méiose, le maintien de l'intégrité de la paroi, le métabolisme du glycogène… Son activité est régulée par **Ypi1** et **Sds22** (Garcia-Gimeno et al., 2003; Pedelini et al., 2007). L'activité de son homologue chez l'Homme, une PP1 sérine/thréonine phosphatase a récemment été montrée comme nécessaire pour la maturation de l'extrémité 3' (Shi et al., 2009). En plus d'interagir spécifiquement avec Pta1 (He and Moore, 2005), **Glc7** interagit avec Pti1, Ref2 (Walsh et al., 2002), Swd2 (Nedea et al., 2008), Yhh1, Ydh1 et Ysh1 (Nedea et al., 2003). La délétion de **Glc7** conduit à la dissociation de Ref2 et Swd2 du complexe CPF et inversement (Nedea et al., 2008).

B.1.2.1.9. Mpe1

Mpe1 est une *zinc knuckle* protéine de 50kDa (Krishna et al., 2003), codée par un gène essentiel, et suppresseur des mutations thermosensibles de Pcf11 (Vo et al., 2001). **Mpe1** est formée de trois domaines, liés par des séquences non structurées : un domaine ubiquitine-like (ou UBL) en N-terminal, un domaine *zinc knuckle* dans la région centrale, et un domaine *RING finger* en C-terminal (Lee and Moore, 2014). Cette organisation laisse supposer que **Mpe1** est une protéine multifonctionnelle. L'interaction avec l'ARN est permise par les domaines *RING finger* et *zinc knuckle*. **Mpe1** est nécessaire pour le clivage et la polyadénylation, et permet une sélection du site de clivage. Récemment, Lee et Moore (Lee and Moore, 2014) ont proposé **Mpe1** comme médiateur de l'ubiquitination de Pap1. L'analyse du complexe de maturation de l'extrémité 3' chez l'homme (Shi et al., 2009) a permis d'identifier Rbbp6/PACT comme homologue de **Mpe1**, mais son rôle dans le complexe de polyadénylation reste inconnu. Rbbp6/PACT interagit avec p53 et pourrait permettre de lier maturation de l'extrémité 3' et régulation de la croissance cellulaire. **Mpe1** est retrouvé faiblement associé au complexe CPF, via son domaine *ubiquitine-like*, par

interaction avec Pfs2 et Fip1 (Gavin et al., 2002). **Mpe1** interagirait également avec le sous-complexe CF IA (Lee and Moore, 2014).

B.1.2.1.10. Pfs2

Pfs2 est une protéine de 53kDa qui, de la même manière que Swd2, présente sept répétitions WD-40 entre les résidus 90 à 380, caractéristiques des protéines retrouvées dans des complexes multimoléculaires (Ohnacker et al., 2000). **Pfs2** est une protéine essentielle (Ohnacker et al., 2000; Wang et al., 2005). En accord avec sa structure, **Pfs2** est retrouvée comme essentielle pour l'interaction entre les complexes CF IA et CPF – interaction avec Rna14, Clp1, Ssu72, Ysh1/Brr5, Ydh1/Cft2, Mpe1 et Fip1 (Ohnacker et al., 2000). De plus, **Pfs2** est nécessaire pour les activités de clivage et de polyadénylation (Wang et al., 2005). Son homologue dans le système mammifère, **WDR33** (Shi et al., 2009), a récemment été montré comme faisant partie du complexe CPSF. WDR33 interagit spécifiquement avec l'hexamère AAUAAA (Chan et al., 2014; Schonemann et al., 2014).

B.1.2.1.11. Swd2

Swd2 est une protéine de 37kDa, présentant sept répétions WD-40 et qui possède 35% d'identité avec WDR82, protéine de fonction inconnue. **Swd2** appartient indépendamment à deux complexes de fonctions distinctes : CPF et le complexe SET1C (ou COMPASS), permettant la méthylation de la lysine 4 de l'histone 3 (Nedea et al., 2008; Dichtl et al., 2004). Swd2 serait nécessaire pour la diméthylation et la triméthylation de la lysine 4 de l'histone 3 et permettrait l'assemblage et la stabilité du complexe SET1C. **Swd2** n'est pas essentielle pour les activités de clivage et de polyadénylation mais permettrait de lier les complexes CPF et CF IA - interaction avec Yhh1, Ysh1, Pta1, Pfs2, Ref2, Ssu72, Glc7, Pti1 et Pcf11 (Dichtl et al., 2004). **Swd2** est impliquée dans la terminaison de transcription des snoRNAs (Cheng et al., 2004) et l'export, participant à la régulation de l'ubiquitination de H2B (Dichtl et al., 2004; Vitaliano-Prunier et al., 2012).

B.1.2.1.12. Ssu72

Avec Glc7, **Ssu72** (23kDa) est la seconde phosphatase (motif CX_5RS) retrouvée associée au complexe CPF (Krishnamurthy et al., 2004). Elle permet la transition entre l'initiation et l'élongation de la transcription, par déphosphorylation du résidu SER^5 du CTD

de l'ARN polymérase II (Ghazy et al., 2009; Meinhart et al., 2003; Rosado-Lugo and Hampsey, 2014). Son activité serait augmentée en présence de Ess1/Pin1 impliqué dans l'isomérisation des résidus prolines du CTD. Initialement, **Ssu72** a été identifiée par son interaction génétique avec le facteur de transcription TFIIB et est impliquée dans la sélection du site d'initiation (Sun and Hampsey, 1996). Grâce à son interaction avec le facteur CPF, **Ssu72** est impliquée dans le clivage du pré-ARNm, mais ne l'est pas dans l'ajout de la queue poly(A) (He et al., 2003). L'allèle *ssu72-2* présente un défaut de terminaison des certains ARNm et snoRNAs (Dichtl et al., 2002a; Ganem et al., 2003). Indépendamment de sa fonction de phosphatase, **Ssu72** permet la maturation de l'extrémité 3' et le lien entre le facteur CPF, l'ARN polymérase II (via son interaction avec Rpb2 – (Dichtl et al., 2002a)) et TFIIB lors de la formation du PIC. **Ssu72** interagit avec Pta1, Pcf11, Pfs2, Ydh1Cft2 et Ysh1/Brr5 (He et al., 2003). Son homologue chez l'Homme, hSsu72 présente une activité de phosphatase et interagit avec TFIIB.

B.1.2.1.13. Syc1

Syc1 est la plus petite sous-unité du CPF (21kDa) et présente une forte homologie avec la région C-terminale de Ysh1/Brr5. **Syc1** n'est pas essentielle pour la maturation de l'extrémité 3' (Nedea et al., 2003). *In vitro*, la délétion de *SYC1* permet de restaurer partiellement le phénotype de thermosensibilité ainsi que le défaut de polyadénylation induit par la délétion de *YSH1/BRR5* et *PTA1*. **Syc1** est présentée par Zhelkovsky *et al*. (Zhelkovsky et al., 2006) comme un régulateur potentiel de l'endonucléase – par compétition avec l'extrémité C-terminale. **Syc1** est faiblement associée au CPF et interagit avec Ysh1/Brr5 et Pta1.

B.1.2.1.14. Ref2

Ref2 est une protéine de 60kDa, riche en lysine et sérine, nécessaire pour la viabilité cellulaire dans certains fonds génétiques (Dheur et al., 2003) et pour une croissance normale dans d'autres (Russnak et al., 1995). **Ref2** permet de stimuler le clivage des pré-ARNm présentant un faible signal de polyadénylation par interaction avec le pré-ARNm ; cela permet l'ancrage du complexe. **Ref2** est également impliquée dans la terminaison de transcription des snoRNAs. **Ref2** interagit avec Swd2 (Cheng et al., 2004) et Glc7 (Nedea et al., 2003). La délétion de **Ref2** (et/ou Swd2) peut être secourue par la surexpression de Sen1 (Nedea et al., 2008).

B.1.2.1.15. Pti1

Pti1 est une protéine de 47kDa, présentant une forte homologie de séquence avec Rna15 et identifiée comme suppresseur multicopie de mutations de Pcf11 (Dheur et al., 2003). Cette protéine présente un domaine d'interaction à l'ARN et plusieurs sites potentiels de phosphorylation. De la même manière que Ref2, **Pti1** n'est pas nécessaire aux activités de clivage et de polyadénylation, cependant une surexpression de Pti1 inhibe la polyadénylation (Dheur et al., 2003). **Pti1** serait impliquée dans la terminaison de transcription des snoRNAs, par inhibition de la polyadénylation. **Pti1** interagit génétiquement avec CTK1, l'une des protéines kinases ciblant le résidu SER^2 du CTD lors de l'élongation (Skaar and Greenleaf, 2002). **Pti1** interagit avec Pta1, Swd2 et Glc7 (Kyburz et al., 2003). Pti1 est supposée, avec Pta1, permettre le recrutement de Glc7. **Pti1** serait un régulateur négatif de Glc7 et Pta1 un régulateur positif (He and Moore, 2005).

B.1.2.2. Protéines liant l'ARN

La régulation de l'ajout de la queue poly(A) requiert, en plus de la machinerie de polyadénylation, le recrutement de protéines spécifiques, interagissant avec les séquences poly(A). Deux protéines sont proposées comme régulant la taille de la queue poly(A).

B.1.2.2.1. Pab1

Pab1 (64kDa) est impliquée dans (i) l'initiation de la traduction (interaction avec eIF4G et circularisation de l'ARN mature), (ii) l'élimination de la coiffe, et (iii) la diminution de la taille de la queue poly(A), par stimulation du complexe PAN (Caponigro and Parker, 1996; Dheur et al., 2005). **Pab1** est relativement abondante (190 000 copies par cellule) et est située dans le cytoplasme (Ghaemmaghami et al., 2003). Dans sa partie N-terminale, **Pab1** est formée de quatre domaines liant l'ARN (« *RNA Recognition Motifs* » ou « *RRMs* »). La région C-terminale est formée d'un domaine riche en glutamines et prolines ainsi que d'une série d'hélices α conservées. La région C-terminale permet des interactions protéines-protéines. **Pab1** est retrouvée associée aux ARNm matures dans le noyau et le cytoplasme et contribue au recrutement de la sous-unité 40S du ribosome (Setyono and Greenberg, 1981). Cependant, lors de la purification du complexe CF IA (Minvielle-Sebastia et al., 1997), **Pab1** est retrouvée associée au complexe, via son interaction avec Rna15 (Amrani et al., 1997a). Les analyses de mutants thermosensibles ou par immuno-déplétion indiquent que l'absence de **Pab1** conduit à la polyadénylation de transcrits de manière aberrante (110 à 200 résidus). À

l'inverse, un excès de **Pab1** peut conduire à une inhibition de la synthèse. **Pab1** est utilisée pour la reconstitution des activités de maturation de l'extrémité 3' *in vitro* (Minvielle-Sebastia et al., 1997; Amrani et al., 1997a).

B.1.2.2.2. Nab2

Nab2 (pour « *nuclear polyAdenylated RNA-binding protein* ») est une protéine de 58kDa, située dans le noyau et vingt fois moins abondante que Pab1 (Anderson et al., 1993; Ghaemmaghami et al., 2003). La partie N-terminale de la protéine est nécessaire pour son export, suivie d'une région riche en glutamine (répétition de onze glutamines et d'une répétition du motif QQQP) (Grant et al., 2008). La région centrale constitue le NLS et est formée d'un domaine RGG - riche en arginines et glycines. Dans la région C-terminale, se trouvent sept motifs en doigt de zinc de type CCCH, permettant l'interaction avec l'ARN (lors de l'export ou de l'interaction avec la séquence poly(A)). **Nab2** est un adaptateur de Mex67, protéine permettant l'export des mRNPs (Green et al., 2002) et est retrouvée associée au mRNPs dans le cytoplasme (Hanada et al., 2013). Contrairement à Pab1, **Nab2** n'est pas retrouvée associée aux complexes CF IA ou CPF. L'inactivation de l'expression de *NAB2* conduit à l'accumulation de transcrits hyperadénylés – supérieur à 170 résidus (Hector et al., 2002). La surexpression de Pab1 permet de restaurer le défaut d'export mais ne permet pas de restaurer la taille physiologique des queues poly(A). De la même manière que le complexe TREX2, **Nab2** aurait un effet sur la stabilité du génome (Jimeno et al., 2002).

B.1.2.3. CF IB – Hrp1

Le sous-complexe CF IB est formé d'une seule protéine, **Hrp1/Nab4** identifiée par Kessler *et al.*, en 1996, et qui ne présente pas d'homologue dans le système mammifère (Kessler et al., 1996). **Hrp1** est une protéine de 73kDa, présentant une structure proche des hnRNPs (« *heterogeneous nuclear ribonucleoprotein* ») de type A/B. **Hrp1** comporte deux domaines liant l'ARN (*RRMs*) dont la structure est résolue dans sa région centrale (**Figure 1.20E**) (Leeper et al., 2010; Perez-Canadillas, 2006). Bien que décrit comme nécessaire pour les activités de clivage et de polyadénylation (Kessler et al., 1997), Hrp1 n'est impliquée que dans l'étape de clivage et permet la reconnaissance spécifique de l'*efficiency element* (EE) (Minvielle-Sebastia et al., 1998; Perez-Canadillas, 2006). Chacun des RRMs permet de reconnaître l'hexamère riche en AU, permettant la sélection du site de clivage, en supprimant l'utilisation des sites alternatifs. L'interaction de **Hrp1** avec Rna15 par l'intermédiaire de

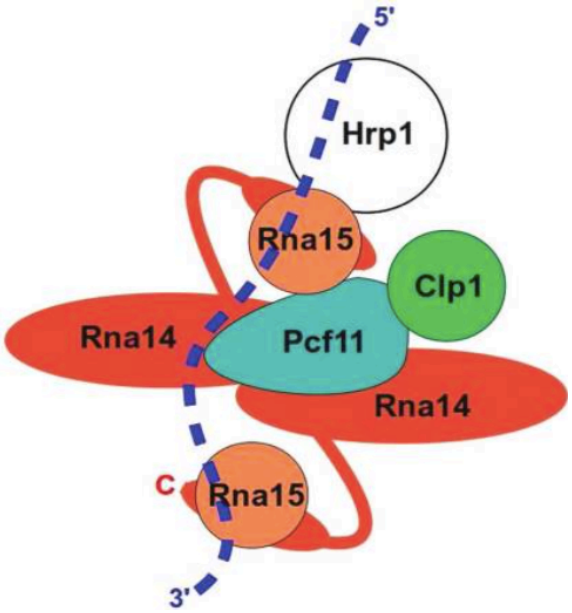


Figure 1.18 : Modèle d'organisation hétéro-hexamérique du complexe CF IA – CF IB.
Le complexe CF IA est asymétrique, formé d'un hétéro-dimère Pcf11/Clp1 (en **bleu**/**vert**) et d'un hétéro-tétramère Rna14/Rna15 (en **rouge**/**orange**). Hpr1 (en blanc) interagit avec une copie de l'hétéro-dimère Rna14/Rna15. Rna15 interagit avec la séquence UAUAUA et Hpr1 interagit avec le PE, AAUAAA. L'ARN est représenté en **bleu**. L'organisation du complexe a été obtenue à partir de protéines exprimées dans un système bactérien.
D'après Gordon *et al.*, 2011.

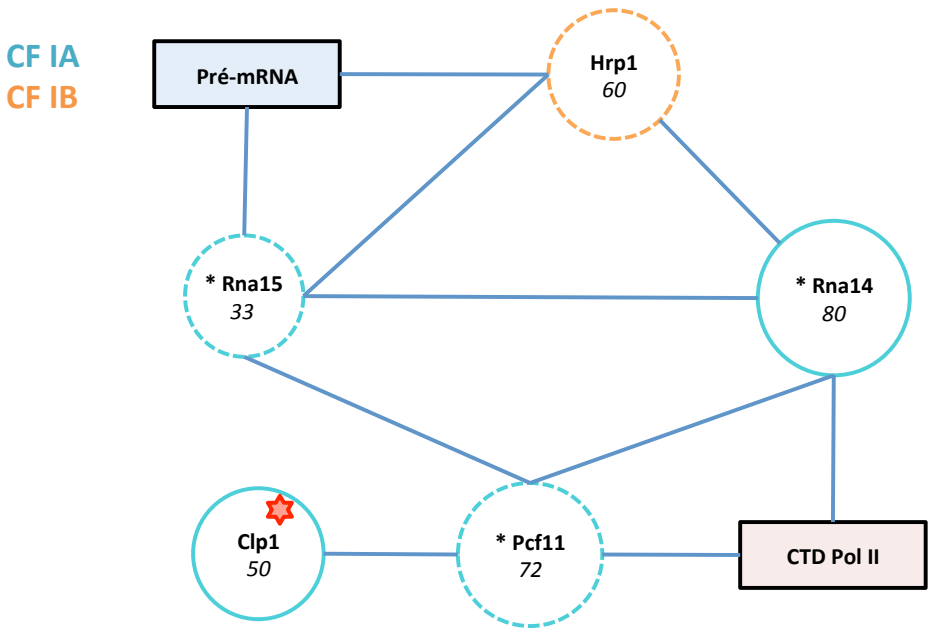


Figure 1.19 : Représentation des interactions entre les sous-unités des complexes CF IA et CF IB. Les interactions directes protéine-protéine, décrites dans la littérature sont représentées en **bleu**. L'appartenance pour chacune des sous-parties est représentée par un code couleur : **CF IA** et **CF IB**. Le diamètre de chaque cercle est proportionnel à la taille des protéines, donnée en kDa. Les sous-unités interagissant avec l'ARN sont représentées en pointillées. Pcf11 et Rna14 interagissent avec le **CTD** de l'ARN polymérase II. L'étoile rouge symbolise l'ATP interagissant au niveau du domaine « P-loop » Walker A de Clp1.

Rna14 augmente l'efficacité de reconnaissance de la séquence d'ARN (Barnwal et al., 2012; Gross and Moore, 2001b; Noble et al., 2004). **Hrp1** est également impliquée dans l'export des ARNm (Xu and Henry, 2004) et dans une de voie de dégradation des ARNm (Gonzalez et al., 2000).

B.1.2.4. CF IA

Le complexe CF IA est nécessaire aux activités de clivage et de polyadénylation et est composé de quatre protéines (Kessler et al., 1996) : Rna15, Rna14, Pcf11 et Clp1. Rna14 et Rna15 sont les premiers composants du CF IA identifiés, et sont impliqués dans la stabilité des ARNm et l'ajout de la queue poly(A) (Bonneaud et al., 1994; Minvielle-Sebastia et al., 1994; Minvielle-Sebastia et al., 1991). Dans le système mammifère, les homologues de Rna14 et Rna15 appartiennent au complexe CstF alors que les homologues de Pcf11 et Clp1 sont regroupés dans le complexe $CFII_m$ (de Vries et al., 2000). La stœchiométrie du complexe proposée par Gordon *et al.*, en 2011 (Gordon et al., 2011) est présentée dans la **Figure 1.18** – 2 :2 Rna14/Rna15 et 1 :1 Pcf11/Clp1. Les interactions entre les différentes sous-unités du complexe CF IA et CF IB sont représentées dans la **Figure 1.19**. Les complexes CF IA et CF IB reconnaissent les éléments PE (Rna15) et EE (Hrp1) du pré-ARNm (Leeper et al., 2010). Des études récentes suggèrent également que le complexe CF IA serait impliqué dans la ré-initiation de l'ARN polymérase II (Al Husini et al., 2013) ainsi que dans certaines voies de réparation de réparation de l'ADN (Gaillard and Aguilera, 2014).

B.1.2.4.1. Rna15

Rna15 est la plus petite sous-unité du complexe CF IA (36kDa), codée par un gène essentiel et présentant 43% d'identité avec CstF-64, homologue dans le système mammifère (Perez Canadillas and Varani, 2003). Les données structurales ont été obtenues par cristallisation des différents domaines isolés (**Figure 1.20C**). **Rna15** est la seule protéine du complexe CF IA décrite comme liant l'ARN (Noble et al., 2004) via un RRM situé dans la région N-terminale de la protéine. *In vivo*, **Rna15** est décrit comme interagissant avec l'élément PE (Gross and Moore, 2001b) bien que *in vitro*, son affinité soit plus importante pour des séquences riches en U. La spécificité d'interaction entre **Rna15** et la séquence PE du pré-ARNm est augmentée en présence de Hrp1 et de Rna14, par une modification du site reconnaissant l'ARN (Gross and Moore, 2001b; Leeper et al., 2010). Le domaine central de la protéine, ou région charnière « *hinge* » identifiée chez **CstF-64**, permet l'interaction avec

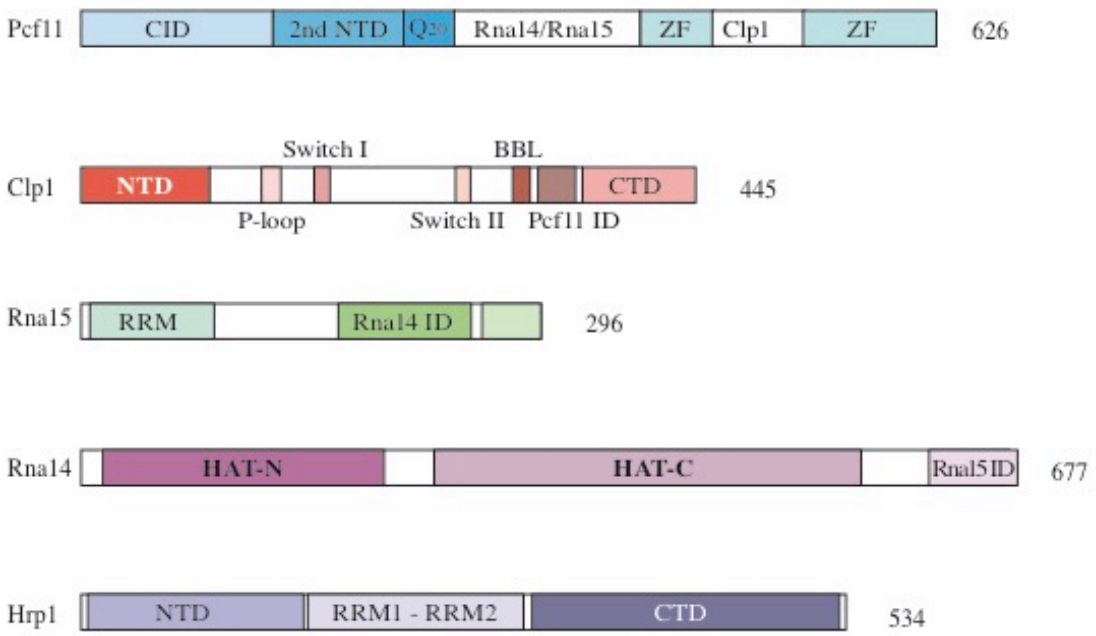


Figure 1.20 : Représentation schématique des différentes sous-unités du complexe CF IA et CF IB. Chaque protéine est représentée à l'échelle, en fonction de sa taille en aminoacides.
(A.) Pcf11 est formée de sept régions : le domaine d'interaction avec le CTD (CID – 1-141) et le 2nd NTD (141-233) dans la partie N-terminale de la protéine, une répétition de 20 glutamines (Q_{20}) et les domaines d'interaction avec Rna14/Rna15 et Clp1 (Rna15/Rna15 ID – 288-400 – et Clp1 ID – 460-499) dans la partie C-terminale. Le domaine d'interaction avec Clp1 est encadré de deux régions contenant des motifs en doigt de zinc (420-450 ; 560-610).
(B.) Clp1 est formée de trois régions : un domaine N-terminal (1-101), une région centrale (102-342) formée du domaine d'interaction avec Pcf11 (Pcf11 ID), d'un domaine « P-loop » Walker A d'interaction avec l'ATP et de deux motifs Switch (*switch* I – DxxQ – et *switch* II – DTPS). La région C-terminale (342-445) forment un domaine globulaire de feuillets β et d'hélices α.
(C.) Rna15 interagit avec l'ARN via un RRM situé dans la région N-terminale de la protéine (1-96). L'interaction avec Rna14 est permise par la région centrale (96-243) de la protéine (Rna14 ID – 127-232). La région C-terminale est fortement conservées et permet les activités de maturation de l'extrémité 3' et l'interaction avec Pcf11 (243-296). Elle est formée de deux hélices α (quatre tours) connectées par une hélice α plus petite (deux tours).
(D.) Rna14 est formée d'un domaine HAT-N (1-140 – motifs HATs 1 à 5) et d'un domaine HAT-C (241-549 – motifs HAT 6 à 12) impliqués dans sa dimérisation. Le domaine C-terminal (626-677) permet l'interaction avec Rna15 (Rna15 ID). La région de liaison entre le domaine HAT-C et le domaine C-terminal (546-626) est une région flexible, accessible aux protéases (Paulson et Ton, 2012)
(E.) Hpr1 est formée d'une région N-terminale (1-158), d'un domaine central d'interaction avec l'ARN ou RBD (160-317) formé de deux RRMs (« *RNA Recognition Motifs* ») et d'une région C-terminale (321-534).
D'après Barnwal *et al.,* 2012, Amrani *et al.*, 1997 et Holbein *et al.,* 2011.

Rna14 (Barnwal et al., 2012; Moreno-Morcillo et al., 2011; Noble et al., 2004). La région C-terminale - 50 acides aminés - est plus conservée et est nécessaire pour les activités de maturation de l'extrémité 3'. Elle est formée de trois hélices α permettant l'interaction avec une autre protéine du CF IA, Pcf11 (Gross and Moore, 2001a; Qu et al., 2007). En plus d'interagir avec les autres partenaires de CF IA et CF IB, **Rna15** interagit avec Pab1 (Amrani et al., 1997a). **Rna15** permet également le recrutement de Yra1, nécessaire pour l'export des mRNPs mais l'interaction est moins robuste qu'entre Pcf11 et Yra1 (Johnson et al., 2009; Lei and Silver, 2002).

B.1.2.4.2. Rna14

Rna14 est une protéine de 80kDa, codée par un gène essentiel et identifiée avec Rna15, pour son rôle dans les activités de clivage et de polyadénylation(Minvielle-Sebastia et al., 1994; Minvielle-Sebastia et al., 1991). Comme son homologue dans le système mammifère, **CstF-77**, **Rna14** présente des répétitions, à partir de l'extrémité N-terminale, de douze motifs HAT (séquence de 34 acides aminés - *Half a TRP*) impliqués dans des interactions protéine-protéine (Preker and Keller, 1998). Cette protéine est organisée en dimère, via la région HAT-C (HAT 6 à 12 – résidus 241 à 349) et formerait un hétérotétramère avec Rna15 via son domaine C-terminal (Noble et al., 2004) (**Figure 1.20D**). La structure de **Rna14** en présence de la région charnière de Rna15 a été décrite (Paulson and Tong, 2012). **Rna14** interagit avec différentes sous-unités des complexes CF IA, CPF et CF IB – Rna15 (Kessler et al., 1996; Minvielle-Sebastia et al., 1994), Hrp1 (Barnwal et al., 2012; Kessler et al., 1997), Pab1 (Amrani et al., 1997a), Fip1 (Preker et al., 1995), Pfs2 (Ohnacker et al., 2000), Yhh1 (Kyburz et al., 2003) et Pcf11 (Amrani et al., 1997b). **Rna14** interagit également avec le CTD de l'ARN polymérase II, préférentiellement lorsqu'il est phosphorylé (Sadowski et al., 2003). Cependant, la zone d'interaction n'est pas encore identifiée – Rna14 ne présente pas de domaine similaire aux CID connus. **Rna14** est considérée comme une protéine d'architecture.

B.1.2.4.3. Clp1

Clp1 est une protéine de 50kDa, codée par un gène essentiel, qui interagit exclusivement avec Pcf11 dans le complexe CF IA (Noble et al., 2007). Elle est l'homologue de **hClp1** – 26% d'identité - et présente un motif de liaison à l'ATP de type Walker A relativement conservé (de Vries et al., 2000). **Clp1** est la dernière sous-unité du complexe

CF IA à avoir été identifiée (Minvielle-Sebastia et al., 1997), mais est la sous-unité la mieux caractérisée d'un point de vue structural. **Clp1** contribue à maintenir l'intégrité du complexe CF IA. Elle est impliquée dans les étapes de maturation de l'extrémité 3' et dans la terminaison de transcription des ARNm et des snoRNAs (Haddad et al., 2012; Holbein et al., 2011). La structure de **Clp1**, en présence d'une molécule d'ATP et de la région minimale d'interaction avec Pcf11 a été décrite en 2007 (Noble et al., 2007). Son organisation, présentée dans la **Figure 1.20B**, montre une forte similitude en comparaison à d'autres ATPases. Chez la levure, des mutants perturbant la liaison à l'ATP sont délétères pour la maturation 3' (Ghazy et al., 2012; Haddad et al., 2012). La fonction de l'ATP ainsi que celle attribuée à son homologue **hClp1** seront abordées dans le Chapitre 2. **Clp1** interagit avec Yhh1/Cft1, Ysh1/Brr5, Ssu72, Pfs2 (Ghazy et al., 2012) et faiblement Ydh1/Cft2 (Holbein et al., 2011; Kyburz et al., 2003). Clp1 interagit également avec Rpb2, sous-unité de l'ARN polymérase II (Pearson and Moore, 2014).

B.1.2.4.4. Pcf11

Pcf11 est une protéine de 72kDa, codée par un gène essentiel, et identifiée en 1997, par Amrani *et al.* (Amrani et al., 1997b). **Pcf11** est une protéine multifonctionnelle ; elle possède un rôle structural, un rôle dans la maturation de l'extrémité 3', ainsi que dans la terminaison de transcription et l'export. **Pcf11** interagit spécifiquement avec Clp1 (Haddad et al., 2012) et l'hétérotétramère Rna14 et Rna15, permettant l'assemblage du complexe CF IA (Amrani et al., 1997b). **Pcf11** permet également de lier CF IA au complexe CPF, par interaction avec les sous-unités Pta1, Ydh1/Cft2, Yhh1/Cft1 (Kyburz et al., 2003), Mpe1 (Vo et al., 2001), Ysh1/Brr5 et Swd2 (Cheng et al., 2004). D'un point de vue fonctionnel, la région C-terminale de la protéine a été décrite comme nécessaire et suffisante pour la maturation de l'extrémité 3' – clivage et polyadénylation (Amrani et al., 1997b). La région N-terminale de la protéine a été décrite comme interagissant avec la forme P-SER[2] du CTD de l'ARN polymérase II (Barilla et al., 2001; Licatalosi et al., 2002) et nécessaire dans la terminaison de transcription par l'ARN polymérase II (Grzechnik et al., 2015; Sadowski et al., 2003). Le domaines d'interaction avec le CTD de l'ARN polymérase II (ou CID) de **Pcf11** permet également l'interaction avec le pré-ARNm (Hollingworth et al., 2006). Cette interaction a été décrite par Zhang *et al.*, en 2005 (Zhang et al., 2005), et a permis de proposer un rôle de **Pcf11** dans le désassemblage de la polymérase II en fin d'élongation. **Pcf11** est également impliqué dans l'export des mRNPs, via son interaction avec Yra1 (Johnson et al.,

2011). Son homologue chez l'Homme, hPcf11, est une protéine d'un poids moléculaire apparent de 130kDa à 200kDa (de Vries et al., 2000) impliquée dans la terminaison de transcription et la dégradation de l'ARN en 3' du site de clivage, par interaction avec Xrn2 (West and Proudfoot, 2008).

Les différentes fonctions de **Pcf11** ainsi que son organisation (**Figure 1.20A**) seront étudiées et discutées en détails dans le Chapitre 1.

B.1.3. CTD de l'ARN polymérase II

Depuis que la plupart des facteurs impliqués dans la polyadénylation des ARNm sont impliqués dans d'autres mécanismes – export, terminaison de transcription, épissage, ou encore traduction – leur recrutement tout au long de la transcription nécessite une coordination précise et spécifique. Ceci est permis par le domaine C-terminal de la plus grande sous-unité, Rpb1, de l'ARN polymérase II – ou CTD de l'ARN polymérase II. La délétion d'une partie du CTD conduit à l'accumulation d'ARMm instables, rapidement dégradés par Xrn1 chez *S. cerevisiae* (Licatalosi et al., 2002).

Le CTD de l'ARN polymérase II est une extension non structurée (Cramer et al., 2001), indépendante de la fonction catalytique *in vitro*, mais nécessaire pour la viabilité cellulaire *in vivo* – la délétion des 2/3 du CTD n'est pas viable (Zehring et al., 1988). Le CTD est formé de la répétition d'un heptapeptide conservé TYR[1]-SER[2]-PRO[3]-THR[4]-SER[5]-PRO[6]-SER[7] ; les répétitions varient de 26 chez *S. cerevisiae* à 52 chez l'Homme. L'augmentation du nombre de répétitions est corrélée avec des déviations dans la séquence consensus. Chez l'Homme, les 26 répétitions N-terminale sont strictement conservées alors que la sérine en position 7 est progressivement remplacée par des résidus arginine ou lysine vers l'extrémité C-terminale (Bartkowiak et al., 2011). Ces variations permettent de complexifier les modifications post-traductionnelles du CTD (méthylation du résidu R1810 par CARM1 chez l'Homme). Les positions 1 et 6 restent strictement conservées (Napolitano et al., 2014). Le CTD est une plateforme sélective et stable permettant de coupler la transcription avec les étapes de maturations, de manière temporelle (Hsin and Manley, 2012). Le recrutement est permis par une série de modifications post-traductionnelles des résidus du CTD, de manière indépendante et dynamique, tout au long de la transcription. Toutes les positions sont potentiellement modifiables par phosphorylation, méthylation, ubiquitination, glycosylation ou isomérisation. L'ensemble des possibilités de modifications forme le « code CTD » – 128^n

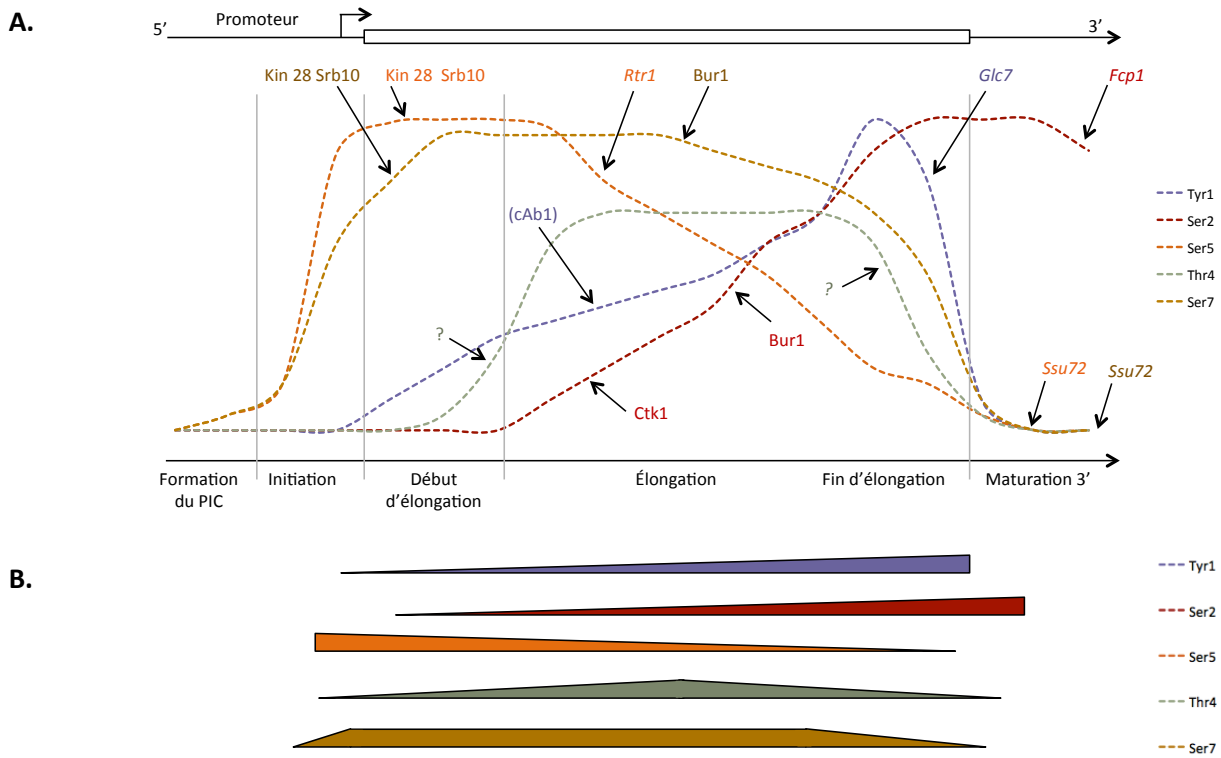


Figure 1.21 : Représentation des modifications du CTD durant la transcription d'un ARN codant par l'ARN polymérase II. **(A.)** et **(B.)**. L'état phosphorylation des différents résidus **Tyr1**, **Ser2**, **Thr4**, **Ser5** et **Ser7** d'une répétition du CTD est représenté en fonction de la distance par rapport au promoteur, pour un gène codant, chez *S.cerevisiae*. Les étapes de la transcription sont indiquées. Les kinases et les phosphatases (en *italique*) impliquées dans ces modifications post-traductionnelles sont indiquées pour chaque résidu. ? correspond aux enzymes non identifiées comme impliquées dans la modification du CTD. (cAb1) est l'homologue chez les mammifères. Cette figure est réalisée à partir des résultats de différentes études disponibles dans la littérature.

où n est le nombre de répétitions de l'heptapeptide. Le profil des modifications est variable en fonction des gènes transcrits (Kim et al., 2010). L'ensemble des modifications par phosphorylation est regroupé dans la **Figure 1.21** (Bartkowiak et al., 2011; Egloff and Murphy, 2008; Napolitano et al., 2014; Phatnani and Greenleaf, 2006). La glycosylation du CTD chez les eucaryotes supérieurs a lieu sur les résidus SER[2], THR[4], SER[5] et SER[7]. Seules les rôles des phosphorylation et isomérisation sont connues. La méthylation, la glycosylation et la phosphorylation semblent être mutuelles exclusives et pourrait former un mécanisme de régulation en particulier lors du recrutement de la polymérase.

Lors du recrutement de la polymérase au niveau de la séquence promotrice du gène, le CTD est sous sa forme hypo-phosphorylée. Cela permet la formation du PIC (« *Pre-Initiation Complex* ») et le recrutement des différents facteurs généraux de transcription et notamment TFIIH. TFIIH est un complexe formé de dix sous-unités, impliqué dans la transcription, le contrôle du cycle cellulaire et la réparation de l'ADN. Parmi ses dix sous-unités, se distingue Kin28/Ccl1 (CDK7/Cycline H chez l'homme – (Feaver et al., 1994)), protéine kinase, qui va phosphoryler les sérines des positions 5 et 7 du CTD – P-SER[5] et P-SER[7] (Komarnitsky et al., 2000). Cela favorise la libération de l'ARN pol II du promoteur et permettrait le début de l'élongation. Srb10 (CDK8) cible également les positions 5 et 7, mais son rôle reste moins étudié ; Kin28 est la kinase dominante.

P-SER[5] conduit au recrutement des enzymes d'addition de la coiffe (Ceg1 et Abd1 - McCracken et al., 1997), de modifications des histones (Set1 pour la triméthylation H3K4ou Dot1 pour la triméthylation H3K79) et le recrutement du facteur de terminaison de transcription Nrd1 (Gudipati et al., 2008). SER[5] est progressivement déphosphorylé par Rtr1 et Ssu72 jusqu'en fin d'élongation (Krishnamurthy et al., 2004; Mosley et al., 2009; Rosado-Lugo and Hampsey, 2014) (**Figure 1.21**). SER[5] peut être déphosphorylé par Fcp1 *in vitro*.

SER[7] est phosphorylé par Kin28 (kinase primaire) et Bur1 jusqu'en fin d'élongation puis est déphosphorylé par Ssu72 (RPABP2) lorsque l'ARN pol II atteint l'extrémité 3' du gène (Krishnamurthy et al., 2004). La déphosphorylation de SER[7] pourrait faire intervenir une autre phosphatase, puisque la déphosphorylation par Ssu72 est 4000 fois moins efficace que la déphosphorylation de P-SER[5] (Xiang et al., 2012). P-SER[7] joue un rôle important dans la transcription des snRNAs – non épissés et non polyadénylés – la mutation Ser > Ala affecte leur transcription chez l'Homme.

La phosphorylation de la tyrosine en position 1 (P-TYR[1]) est permise par c-Ab1 chez l'homme (Mayer et al., 2012a) et apparaît dès l'initiation de la transcription, puis chute avant

le site de polyadénylation par l'action de Glc7 (Schreieck et al., 2014) (**Figure 1.21**). P-TYR[1] contribue au recrutement du facteur d'élongation Spt6, au relargage de Nrd1 et prévient le recrutement des facteurs de terminaison de transcription Pcf11 et Rtt103 induit par le marquage P-SER[2]/P-SER[5] (Mayer et al., 2012a). La mutation de cette position est létale chez *S. pombe* (Schwer and Shuman, 2011).

Lors de l'élongation, le résidu SER[2] est progressivement phosphorylé par CTDK1 et Bur1/Bur2 considérée comme kinase « interne » qui « prime » le CTD pour CTDK1 (CDK9/Cycline T et CDK12/Cycline K), pour atteindre son maximum en fin de transcription (Ahn et al., 2004; Skaar and Greenleaf, 2002). La déphosphorylation de P-SER[2] par Fcp1 a lieu dès le passage de la polymérase au niveau du site de polyadénylation et contribue à la ré-initiation de la polymérase (**Figure 1.21**). La déphosphorylation de P-SER[2] est permise également permise par la dissociation de la kinase par dé-ubiquitination de certains histones. Le CTD reste parfois phosphorylé à cette position jusque lors de la traduction. P-SER[2] joue un rôle dans la maturation de l'extrémité 3' et la terminaison de transcription (Park et al., 2004). Les phosphorylations des positions 5 et 2 sont inversement proportionnelles et sont d'égale intensité à 450pb du TSS (« *transcription start signal* ») (Kim et al., 2010). La déphosphorylation de SER[5] serait un prérequis pour la phosphorylation de SER[2]. Le double marquage P-SER[2]/ P-SER[5] contribue au recrutement des facteurs d'épissage comme Prp40 (Bird et al., 2004) et des facteurs d'export comme Yra1 ou Npl3 (Bartkowiak et al., 2011; Nino et al., 2013). P-SER[2]/ P-SER5 permet également de recruter le facteur Set2, méthyltransférase lysine 36 de l'histone H3, et qui va inhiber la transcription au niveau de promoteur cryptique et la formation de CUTs (Gudipati et al., 2008).

Le résidu THR[4] est phosphorylé lors de l'élongation. Hsin *et al.*, en 2011 (Hsin et al., 2011) ont montré un enrichissement de P-THR[4] au niveau de la région codante du gène, puis une chute après le site de polyadénylation (**Figure 1.21**). À ce jour, les protéines kinases et phosphatases impliquées sont inconnues.

Le complexe de polyadénylation est recruté progressivement, par la déphosphorylation de **TYR[1]** et la phosphorylation de **SER[2]** (Lunde et al., 2010; Mayer et al., 2012a; Mayer et al., 2012b). Le CTD n'est pas essentiel pour les activités de maturation de l'extrémité 3' mais son absence conduit à une réduction importante des activités de clivage et de polyadénylation (Hirose and Manley, 1998; Licatalosi et al., 2002). Le CTD interagit avec Yhh1, Ydh1, Pcf11 et Rna14 (Barilla et al., 2001; Dichtl et al., 2002b; Kyburz et al., 2003; Licatalosi et al., 2002; Zhang and Gilmour, 2006).

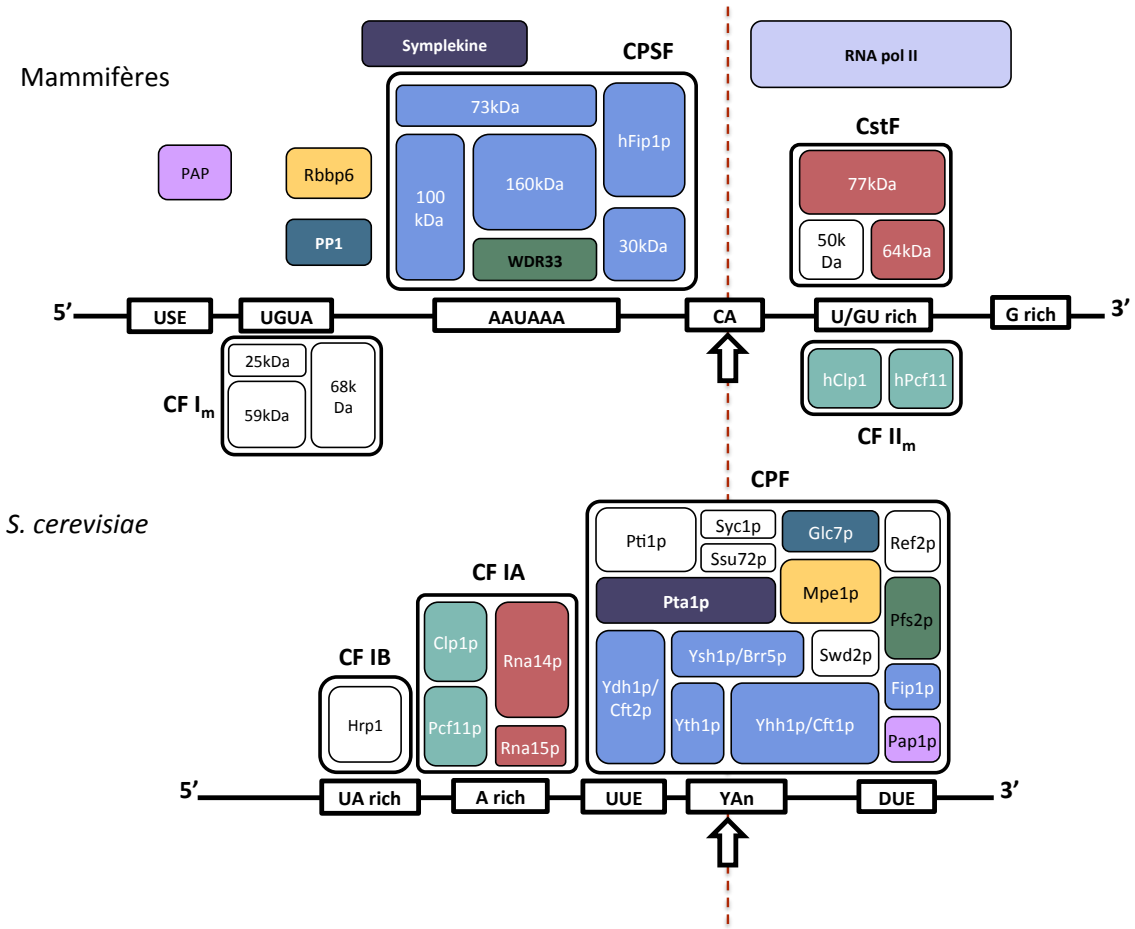


Figure 1.22 : Comparaison du complexe de polyadénylation chez *S.cerevisiae* et chez les mammifères. Représentation schématique de l'organisation des éléments *cis* et des facteurs *trans* impliqués dans la polyadénylation des ARNm chez *S.cerevisiae* (*en bas*) et chez les mammifères (*en haut*). Les facteurs décrits comme présentant des homologies sont représentés de la même couleur. Les facteurs spécifiques de la machinerie levure sont représentés en blanc. Les différents facteurs sont représentés sur l'ARN en fonction des données publiées – interactions ARN/protéine et protéine/protéine. Les homologues nouvellement identifiés (WDR33, Rrrpb6 et PP1) sont positionnés selon les données préliminaires disponibles. Les machineries sont alignés en fonction du site de clivage (ligne rouge - - - -). D'après Millevoi et Vagner, 2010.

Les résidus proline en position 3 et 6 - PRO^3 et PRO^6 - peuvent également être modifiées par isomérisation, via l'action de peptidyl-prolyl *cis/trans* isomérases comme Ess1 chez la levure (Myers et al., 2001) (Pin1 chez l'Homme). Cela permet d'induire une variabilité conformationnelle au sein du CTD (Noble et al., 2005). Ce changement de conformation ne modifierait pas le recrutement des différents facteurs impliqués dans la traduction mais permettrait de moduler le recrutement des protéines kinases du CTD (Egloff and Murphy, 2008). L'activité de Ssu72 serait modulée par la présence de Ess1/Pin1. Des mutations au niveau de Ess1 induisent un défaut de maturation en 3' (Hani et al., 1999). Des données suggèrent que Pcf11 reconnaitrait préférentiellement les conformations *trans* des résidus prolines (Noble et al., 2005).

B.2. Comparaison système levure et mammifères

B.2.1. Machinerie mammifères

De la même manière que chez *S. cerevisiae*, le clivage et la polyadénylation ainsi que le bon positionnement de la machinerie font intervenir des éléments *cis* et des facteurs *trans* (Millevoi and Vagner, 2010) (**Figure 1.22**).

B.2.1.1. Éléments cis

Les séquences *cis* regroupent trois éléments permettant de définir le site de clivage et deux éléments auxiliaires permettant de réguler l'interaction. Ces éléments sont généralement uniques et plus conservés en comparaison à ceux décrits chez *S. cerevisiae*.

- Le **PAS** ou signal de polyadénylation est la première séquence identifiée (Proudfoot, 1976). La séquence consensus AAUAAA est retrouvée dans près de 90% des pré-ARNm. L'hexamère AUUAAA est retrouvé dans près de 10% des pré-ARNm. Le PAS est retrouvé entre 10 et 30 nucléotides en amont du site de clivage. Un raccourcissement de cette séquence invalide le site de clivage. La mutation du dernier nucléotide de cet hexamère (A en G) peut conduire au développement de pathologies comme la thalassémie α ou β, caractérisée par un niveau bas d'ARMm (Higgs and Weatherall, 1983).
- Le **DSE** pour « *Downstream Sequence Element* » est une séquence moins conservée, située à 30 nucléotides en aval du site de clivage. Cet élément se caractérise par une richesse en U et en GU (Zhao et al., 1999a). Les mutations des nucléotides ont un impact moindre sur

le maturation 3', en comparaison à sa délétion. Il joue un rôle important dans la positionnement de la machinerie, par son interaction avec le sous-complexe CstF.

- Le **site de clivage** est positionné entre le PAS et le DSE. Le site de clivage est essentiellement constitué d'une séquence CA. Le second nucléotide peut également être un U, un C ou un G (A > U > C > G) (Zhao et al., 1999a). Le clivage a lieu entre les deux nucléotides. La séquence entourant le site de clivage n'est pas conservée.
- Les séquences auxiliaires regroupent le **USE** pour « *Upstream Sequence Element* » situé en amont du PAS – riche en U permettant de stimuler la maturation, de séquence UUUU, UGUA ou UAUA– et le **DSE**, séquence riche en G, moins bien définie et dont le rôle reste à déterminer.

B.2.1.2. Facteurs trans

Chez les métazoaires, le complexe de polyadénylation est formé de quatorze sous-unités et sont regroupés en quatre sous-complexes : **CPSF** (« *Cleavage and Polyadenylation Specificity Factor* »), **CstF** (« *Cleavage stimulation Factor* »), **CF I_m** (« *Cleavage Factor I* ») et **CF II_m** (« *Cleavage Factor II* ») (Mandel et al., 2008; Takagaki and Manley, 2000). La PAP, PABP et la Simplekin viennent compléter le complexe (**Figure 1.22** et *TABLEAU 1.1*). La purification du complexe chez l'homme, par Shi *et al*., en 2009 (Shi et al., 2009), suggère cependant que le complexe de polyadénylation interagirait avec plus de 50 protéines, impliquées dans des fonctions autour de la transcription par l'ARN polymérase II – épissage, traduction, réponse aux endommagements de l'ADN.

Le sous-complexe **CPSF** est formé de six sous-unités nécessaires pour les activités de clivage et de polyadénylation. On retrouve : CPSF-30, CPSF-100, CPSF-160, l'endonucléase CPSF-73, hFip1 et WDR33. CPSF-160, CPSF30, hFip1 et WDR33 sont nécessaires et suffisants pour la reconnaissance du sous-complexe **CPSF** avec l'hexamère AAUAAA et la polyadénylation (Schonemann et al., 2014). hFip1 et CPSF-160 permettent le recrutement de la poly(A) polymérase. Le sous-complexe **CstF** est formé de trois sous-unités : CstF-64, homologue de Rna15 qui interagit avec la région riche en GU de l'ARN, CstF-77 homologue de Rna14 et CstF-50 (Perez Canadillas and Varani, 2003; Takagaki et al., 1992). Le sous-complexe **CF I_m** regroupe trois sous-unités (25kDa, 59kDa et 68kDa) nécessaires pour le clivage et permettant la reconnaissance et le positionnement de la machinerie au niveau du signal de polyadénylation. Le complexe **CF I_m** est particulièrement important pour les ARNm ne présentant pas le PAS canonique. Le sous-complexe **CF II_m** est formé de deux protéines

		Metazoaires	Homologue Levure	Fonctions et Caracteristiques
CPSF	*Polyadenylation*	**CPSF-160**	*Yhh1/Cft1*	Interaction avec l'ARN
		CPSF-30	*Yth1*	Cinq motifs en doigts de zinc - interaction avec les éléments riche en U de l'ARN
		hFip1	*Fip1*	Interaction avec PAP - interaction avec les éléments riche en U de l'ARN – rôle dans la polyadénylation alternative
		WDR33	*Pfs2*	Activité de clivage et de polyadénylation – reconnaissance du PAS, AAUAAA
	Clivage	**CPSF-100**	*Ydh1/Cft2*	Métallo-β-lactamase modifiée, protéine d'architecture du complexe
		CPSF-73	*Ysh1/Brr5*	Métallo-β-lactamase – Activité **endonucléase**
CstF		**CstF-64**	*Rna15*	Interaction avec l'ARN au niveau de l'élément ruche en G/U (DES)
		CstF-77	*Rna14*	Dimérisation via les domaines HAT - Architecture du complexe CstF et CPSF
		CstF-50	-	Interaction avec le CTD de l'ARN polymérase II
CF I_M		**CF I_m-25kDa**	-	Reconnaissance du PAS et stabilise la PAP et PABP
		CF I_m-68kDa	-	Reconnaissance du PAS – Interaction avec le « *spliceosome* »
		CF I_m-59kDa	-	Reconnaissance du PAS
CF II_M		**hPcf11**	*Pcf11*	Interaction avec le CTD de l'ARN polymérase II – Protéine d'architecture du complexe
		hClp1	*Clp1*	Contient des motifs Walker A et Walker B, rôle dans la maturation des ARNt – Interaction avec les complexes CPSF et CF I_m
		PAP	*Pap1*	Poly(A) polymérase
		PABP N (nucléaire) ou **C** (cytoplasmiques)	*Pab1*	Stimule la polyadénylation et stabilise la queue poly(A)
		Simplekin	*Pta1*	Lien entre les complexes CPSF et CstF – régulation par des cycles phosphorylations
		PP1	*Glc7*	Sérine/thréonine phosphatase
		Rbbp6/PACT	*Mpe1*	?

Tableau 1.2 : Composition du complexe de polyadénylation dans le système mammifère

(hClp1 et hPcf11) identifiées par homologie avec les protéines chez *S. cerevisiae* (de Vries et al., 2000).

Récemment (Shi et al., 2009), la purification du complexe de maturation de l'extrémité 3' chez l'Homme a permis de mettre en évidence le rôle important de trois autres sous-unités : **WDR33**, **Rbbp6/PACT** et **PP1**.

- **WDR33** est l'homologue de Pfs2 (Ohnacker et al., 2000), contenant des régions répétées WD-40, et est nécessaire pour les activités de clivage et de polyadénylation. WDR33 co-purifie avec le complexe CPSF et migre à un poids moléculaire comparable à celui de CPSF-160. WDR33 interagit spécifiquement avec la séquence AAUAAA (Chan et al., 2014; Schonemann et al., 2014).
- **Rbbp6/PACT** est l'homologue de Mpe1 (Vo et al., 2001). Rbbp6 a une organisation proche de celle de Mpe1 – motif riche en R/S – et est impliquée dans la régulation de p53. Sa fonction dans la maturation de l'extrémité 3' reste pour le moment inconnue.
- **PP1** est une sérine/thréonine phosphatase, homologue de Glc7 (He and Moore, 2005) et serait impliquée dans la polyadénylation uniquement.

B.2.2. Comparaison

Les éléments *cis* et les facteurs *trans* sont relativement conservés entre la levure et les eucaryotes supérieurs, ce qui permet une comparaison entre les deux systèmes.

(i) Si on compare tout d'abord les éléments *cis* chez *S. cerevisiae* et chez les mammifère, on remarque que le signal de polyadénylation minimal est formé de trois éléments composés d'une séquence riche en A, un élément riche en U et un site de clivage, commun à tous les eucaryotes. *(ii)* Cependant, l'élément **PAS** est unique, hautement conservé chez les mammifères et strictement nécessaire pour la maturation en 3' ; chez *S. cerevisiae*, le **PE** est nécessaire au positionnement de la machinerie, est moins conservé et fréquemment dupliqué. *(iii)* La machinerie de polyadénylation est positionnée par interaction avec le **PE** (riche en A) chez la levure, et par interaction avec **DSE** (riche en U) chez les mammifères. *(iv)* Chez les mammifères, les séquences riches en U sont situées en aval du site de clivage, rarement dupliquées et essentielles ; chez la levure, ces éléments sont situés amont du site de clivage et sont fréquemment dupliqués. *(v)* Le clivage par l'endonucléase a lieu préférentiellement sur un site CA chez les mammifère et sur un site $Y(A)_n$ chez les levures.

De manière générale, de par leur conservation moins importante, les éléments *cis* semblent plus complexes chez *S. cerevisiae* que chez les mammifères où ils sont hautement conservés.

Bien que très conservées, les différentes sous-unités du complexe ne présentent pas la même organisation chez les mammifères et les levures. Par exemple, *(i)* chez la levure, le complexe **CF IA** est formé de quatre sous-unités (Rna14, Rna15, Pcf11 et Clp1) et est impliqué dans le clivage et la polyadénylation, via la reconnaissance du PE. Les homologues chez les mammifères sont répartis dans deux sous-complexes : **CstF** (CstF-77 et CstF-64) et **CF II_m** (hPcf11 et hClp1) et reconnaissent l'élément DSE en aval du site de clivage. *(ii)* La séquence **PE** est reconnue par le complexe **CF IA** (Rna15) chez la levure et par le complexe **CPSF** (WDR33) chez les mammifères. *(iii)* De même, le complexe **CF IB** permet chez la levure, la reconnaissance de l'élément riche en U. Chez les mammifères, cette reconnaissance est permise par le complexe **CF I_m**, qui regroupe trois protéines. *(iv)* Les complexes **CstF**, **CF I_m** et **CF II_m** regroupent des protéines impliquées dans la reconnaissance des éléments *cis* et la stabilisation du complexe, et qui ne sont impliquées que dans le clivage. Chez la levure, la majorité des composants du complexe de polyadénylation sont impliqués dans les deux étapes de maturation de l'extrémité 3'. *(v)* La machinerie de polyadénylation chez levure regroupe un plus grand nombre de protéines, étroitement liées ; seul Ref2, Mpe1 et Syc1 n'interagissent qu'avec deux sous-unités du complexe. Le complexe de polyadénylation chez la levure présente une organisation plus compacte. La machinerie chez les mammifères est plus relâchée, ce qui permettrait des interactions plus importantes avec d'autres facteurs impliqués dans d'autres processus cellulaire, comme le suggère les résultats obtenus par l'étude du complexe chez l'homme – où le complexe de polyadénylation est retrouvé comme interagissant avec plus de 50 « autres » protéines (Shi et al., 2009).

C. TERMINAISON TRANSCRIPTION

Le contrôle de l'initiation de la transcription par l'ARN polymérase II est faible dans la cellule. La transcription a donc lieu dans tout le génome, sans tenir compte des régions annotées et des régions codantes (sens, anti-sens, à l'intérieur ou à l'extérieur des gènes…). On parle de « ***pervasive transcription* ».** En considérant la taille du génome, le nombre

d'ARNs potentiellement produits est important. Cependant, ces évènements ne sont pas ou peu détectables, les produits ARNs étant rapidement dégradés. Cela entraîne la production de différentes espèces d'ARNs, comme les CUTs, les XUTs ou les SUTs, dont la fonction reste inconnue mais qui pourrait jouer un rôle dans des mécanismes de régulation. La ***transcription pervasive*** est régulée post-transcriptionnellement, via la terminaison de transcription et la dégradation des ARNs produits.

La terminaison de transcription est la dernière étape de synthèse par la polymérase. Elle est définie par la dissociation de l'ARN et de l'ARN polymérase II après un événement de transcription. Cela permet de cloisonner et de définir les limites des unités de transcription. Une terminaison trop tôt ou trop tard peut modifier le devenir de l'ARN et moduler l'expression du gène correspondant. De plus, la terminaison de transcription peut interférer avec l'expression de gènes codants – « *transcritpional collision* » (Prescott and Proudfoot, 2002). Les acteurs de la terminaison de transcription sont recrutés tout au long de l'élongation, par interaction avec le CTD de la Pol II (McCracken et al., 1997). On distingue deux voies : la voie dépendante du complexe de clivage et de polyadénylation permettant la terminaison des ARNs codants et la voie dépendante du complexe NNS, permettant la terminaison des ARNs non codants (Kim et al., 2006). Des facteurs, comme Pcf11, sont impliqués dans les deux voies de terminaison. D'autres voies sont également décrites chez *S. cerevisiae*, permettant le strict maintien des limites des unités de transcription.

La mise en place de ces différentes voies de contrôle présente un coup énergétique pour la cellule. La plupart des protéines impliquées dans ces voies sont donc impliqués dans d'autres mécanismes.

C.1. Terminaison des ARNs messagers chez la levure

C.1.1. Mécanisme général

La terminaison de transcription des ARNs codants permet la libération d'un ARNm mature qui sera ensuite exporté vers le cytoplasme. Cette voie de terminaison est utilisée pour les ARNs codants et pour les ARNs non codants de plus de 1kB. La terminaison de transcription dépend du complexe de polyadénylation puisque de nombreuses mutations des différentes sous-unités affectent la terminaison de transcription (Birse et al., 1998). Des signaux de polyadénylation effectifs sont requis pour la terminaison (Proudfoot, 1989). Le recrutement de ces facteurs permet la libération de l'ARN naissant qui est clivé et polyadénylé. L'ARN polymérase II continue la synthèse d'environ 150 nucléotides avant

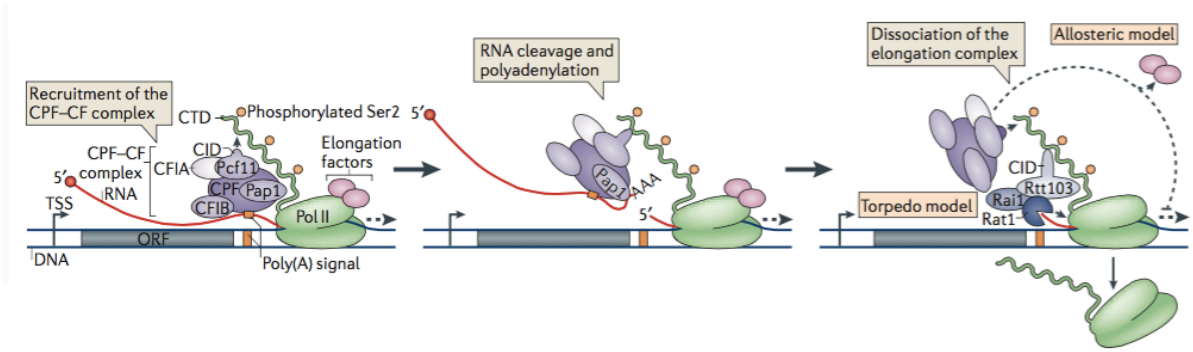


Figure 1.23 : Terminaison des ARNs codants par la voie dépendante du complexe de polyadénylation. La terminaison de transcription des ARNs codants est dépendante du complexe de polyadénylation (CPF, CF IA et CF IB), recruté progressivement jusqu'en fin d'élongation. Les différentes sous-unités permettent la reconnaissance des signaux de terminaison sur l'ARNm naissant et l'interaction avec la polymérase. Pcf11, sous-unité du complexe CF IA, interagit via son CID avec le CTD de l'ARN polymérase II. Le clivage et la polyadénylation permettent la formation d'un ARNm mature. Deux modèles ont été décrits comme permettant le décrochage de la polymérase. Dans le modèle **allostérique** (ou **clivage-indépendant**), le recrutement des facteurs de terminaison et le décrochage des facteurs d'élongation conduisent à un changement de conformation de la polymérase et son décrochage. Dans le modèle **torpédo** (ou **clivage-dépendant**), le clivage par l'endonucléase Ysh1/Brr5 confère un point d'entrée à Rat1/Rai1, recrutée via Rtt103, qui va dégradé l'ARN naissant et permettre le décrochage de l'enzyme. D'après Porrua et Libri, 2015.

d'être décrochée par un mécanisme discuté plus loin dans ce paragraphe. Les différentes sous-unités des sous-complexes CF IA, CF IB et CPF sont impliquées, notamment dans la reconnaissance des signaux de terminaison sur l'ARNm naissant ou l'interaction avec la polymérase - Rna15, Yhh1/Cft1, Ydh1/Cft2, Yth1, Mpe1 ou encore Hrp1 (Minvielle-Sebastia et al., 1997). Clp1 et Ssu72 interagissent avec la sous-unité Rpb2 de l'ARN polymérase II (Pearson and Moore, 2014). Pcf11, sous-unité du complexe CF IA est fortement impliquée dans la terminaison de transcription des ARNm, via son domaine d'interaction avec le CTD de la Pol II (Barilla et al., 2001; Mann et al., 2002; Sadowski et al., 2003).

Le clivage est une étape nécessaire pour la terminaison de transcription. Une étude globale de la localisation de la Pol II sur le génome par PAR-CLIP révèle un défaut de terminaison des ARNm après délétion de l'endonucléase (Schaughency et al., 2014). Deux modèles peuvent être proposés pour la voie dépendante du complexe de clivage et de polyadénylation. La différence entre ces deux modèles est que le clivage n'est pas requis pour le décrochage de l'ARN polymérase dans le cas du modèle ***allostérique.*** Dans le cas du modèle ***torpedo*** c'est une étape obligatoire pour la terminaison de transcription (**Figure 1.23**). Enfin, un modèle couplant la terminaison dépendante du complexe de polyadénylation et le recrutement d'une exonucléase peut être proposé ; on parle d'un modèle ***allostérique/torpédo.***

C.1.2. Allostérique – clivage indépendant

Dans le cas du modèle ***allostérique*** (ou anti-terminateur), la transcription des signaux de polyadénylation induit une modification du complexe d'élongation de la polymérase. Cela permet une pause de la synthèse par la Pol II (Park et al., 2004), le décrochage des facteurs d'élongation et/ou facteurs d'anti-terminaison et le recrutement de facteurs de terminaison, comme Pcf11. À terme, cela conduit au décrochage de l'ARN polymérase II. Le clivage n'est donc pas requis pour la terminaison. De nombreux parallèles peuvent être établis entre le modèle allostérique et la terminaison de transcription Rho-dépendante décrite chez *E. coli*. Le facteur Rho est une ATPase ARN dépendante, présentant une activité ADN : ARN hélicase. L'hypothèse développée par Epshtein *et al.*, en 2010 (Epshtein et al., 2010) suggère que la polymérase est décrochée de l'ARN, par un changement de conformation induit par l'interaction avec le facteur Rho (Washburn and Gottesman, 2015).

Plusieurs études illustrent ce modèle. (i) Tout d'abord, l'indépendance de la terminaison vis-à-vis des facteurs de maturation en 3' est suggérée par l'étude de mutants, présentant un défaut de terminaison de transcription mais pas de maturation (*pcf11-13*,

(Sadowski et al., 2003) ou Yhh1 (Dichtl et al., 2002b)) et inversement : un défaut de maturation 3' n'induit pas un défaut de terminaison. (ii) Kim *et al.,* en 2004 (Kim et al., 2004a) ont montré, par ChIP que la plupart des facteurs d'élongation étaient dissociés après le passage du site de polyadénylation (complexe Paf, Rad26…) et que les sous-unités du complexe de polyadénylation ou facteurs de terminaison étaient recrutées en fin d'élongation (Fip1, Pta1…), notamment Rna14, Rna15 et Pcf11. (iii) Zhang *et al.*, en 2005 (Zhang et al., 2005) ont montré que cette dernière sous-unité était capable de déstabiliser le complexe de polyadénylation *in vitro* – systèmes *S. cerevisiae* et *D. melanogaster*. Pcf11 seul interagit avec le CTD de la Pol II et l'ARN, conduisant à un changement de conformation et son décrochage. Ce modèle d'action de Pcf11 est proche de celui observé chez les procaryotes (terminaison rho-dépendante) mais ne requiert pas l'hydrolyse d'ATP. (iv) De plus, l'activité des facteurs impliqués dans la terminaison peut être modulée par interaction avec des facteurs d'élongation. C'est le cas de Rna15 et le co-activateur, Sub1 (Aranda and Proudfoot, 2001; Calvo and Manley, 2003). Sub1 est considérée comme un anti-terminateur, qui interagit avec le domaine C-terminal de Rna15, permettant une inhibition de son activité. En fin d'élongation, Sub1 est décrochée, permettant l'action de Rna15 et la terminaison de transcription. (v) Enfin, une des sous-unités du complexe CPF, Ssu72, joue un rôle dans la modulation de la capacité d'élongation de la polymérase. Ssu72 permet la pause de la polymérase en aval de signaux de polyadénylation – accumulation de Pol II dans le mutant *ssu72-2* (Dichtl et al., 2002a).

Récemment, Zhang *et al.* (Zhang et al., 2015) ont décrit un argument supplémentaire en faveur du modèle indépendant du clivage, dans un système mammifère. Ils ont analysé la localisation de l'ARN polymérase avant et après le site de clivage, en utilisant deux substrats ADN, sauvage et mutant au niveau du site de clivage (AATAAA et AAGTAC). Le décrochage de la polymérase se fait, *in vitro*, indépendamment du site de clivage et de la réaction de clivage.

C.1.3. Torpedo – clivage dépendant

Le modèle ***torpedo*** est un modèle dépendant de la maturation en 3' (Connelly and Manley, 1988). Le clivage par l'endonucléase confère un point d'entrée à une exonucléase 5'-3' - Rat1 chez *S. cerevisiae* et Xrn2 chez l'Homme. La rencontre de l'exonucléase avec la polymérase II induit sa déstabilisation et le décrochage de l'enzyme. Ce modèle expliquerait les résultats obtenus pour les différents mutants où l'inhibition du clivage et de la

polyadénylation est corrélée avec un défaut de terminaison de transcription (Barilla et al., 2001; Birse et al., 1998).

De la même manière que pour le modèle ***allostérique***, plusieurs études sont en faveur de cette hypothèse. Tout d'abord, Kim *et al.*, en 2004 (Kim et al., 2004b) ont montré que l'inactivation de Rat1 (*rat1-1*) conduisait à une accumulation de transcrits plus long, illustrant un défaut de terminaison. Cependant, la dégradation par Rat1 ou par Xrn1 (homologue de Rat1, exonucléase 5'-3' située principalement dans le cytoplasme) ne permet pas le décrochage de l'enzyme (Luo et al., 2006). Rat1, en présence de son cofacteur Rai1, est recrutée par le biais de Rtt103, protéine interagissant avec le CTD de l'ARN polymérase II. Rat1 est une exonucléase hautement processive, qui va dégrader l'ARN situé en aval du site de clivage, et va « rattraper » la polymérase permettant son décrochage (Tollervey, 2004). Teixeira *et al.*, en 2004 (Teixeira et al., 2004) et West *et al.*, en 2004 (West et al., 2004) ont décrit un système similaire chez l'Homme, via Xrn2, argumentant en faveur de ce modèle.

C.1.4. Allostérique/Torpedo

Ce modèle est basé sur le recrutement coopératif, mutuel des deux facteurs de terminaison, Pcf11 et Rtt103 (Lunde et al., 2010). Pcf11 et Rtt103 sont deux protéines interagissant avec le CTD de l'ARN polymérase II, principalement la forme P-SER^2 du CTD, caractéristique de la fin d'élongation. Cela permet leur recrutement uniquement quand la polymérase atteint le site de polyadénylation et prévient d'une terminaison précoce. La fixation simultanée de Pcf11 et Rtt103 contribue au recrutement du reste de la machinerie de polyadénylation et permet le recrutement de Rat1/Rai1 et le décrochage de la polymérase, par changement de conformation (Lunde et al., 2010). La dégradation par Rat1 (ou Xrn1) n'est pas systématiquement induite par la terminaison de transcription et n'est pas suffisante pour initier le décrochage de l'enzyme. Les analyses par ChIP montrent que le recrutement de Rat1 est fortement réduit dans le mutant *pcf11-2* et inversement – Pcf11 et Rna15 sont moins efficacement recrutées dans le mutant *rat1-1* (Luo et al., 2006).

C.2. Terminaison des ARNs non codants chez la levure

L'ARN polymérase permet la synthèse d'un grand nombre d'ARNs non codants. On distingue les snRNAs (impliqués dans l'épissage des pré-ARNm), les snoRNA (impliqués dans la maturation des ARNr), ou encore des ARNs non codants, produits de la **transcription pervasive**, détectables lorsque les voies de dégradation sont inactivées. Cela regroupe :

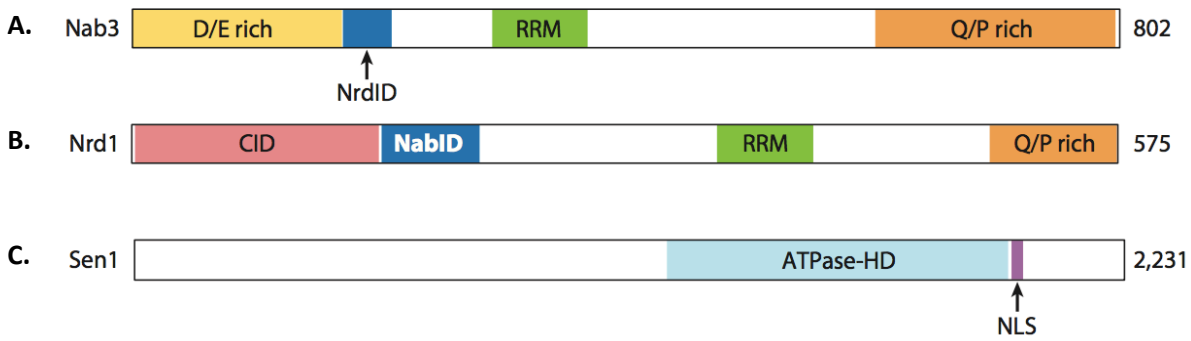


Figure 1.24 : Représentation schématique de l'organisation des sous-unités du complexe NNS. Les protéines ne sont pas représentées à l'échelle ; leur taille est indiquée en aminoacides. D'après Arndt et Reines, 2015.

(A.) Nab3 est formée d'un domaine N-terminal (1-204) riche en acides aminés D et E (acides aspartiques et glutamiques), d'un domaine d'interaction avec Nrd1 (Nrd1 ID – 204-248) et d'un domaine d'interaction avec l'ARN (RRM – 327-403). La région C-terminale est formée d'une séquence riche en P et Q (Q/P rich – prolines et glutamines – 565-794).

(B.) La région N-terminale de **Nrd1** permet l'interaction avec le CTD de l'ARN polymérase II (CID, huit hélices α – 6-151) et avec le NIM (« *Nrd1 Interaction Motif* ») de Trf4. Le CID est immédiatement suivi d'un domaine d'interaction avec Nab3 (NabID, 151-214). La région C-terminale de la protéine est formée d'un domaine d'interaction avec l'ARN (RRM – 332-398) et d'un domaine riche en P et Q (Q/P rich – 516-575).

(C.) Sen1 est une hélicase ADN/ARN de 253kDa, dont deux domaines ont été caractérisés : le domaine ATPase/hélicase (ATPase-HD – 1146-1869) et une séquence de localisation nucléaire (NLS – 1907-1929). La partie N-terminale de la protéine n'est pas nécessaire pour la croissance et interagit avec l'ARN polymérase II, Rnt1 et Rad2 (endonucléase de la voie NER). La partie C-terminale de la protéine contient le domaine hélicase, le NLS et la domaine d'interaction avec Glc7.

les CUTs pour « *cryptic unstable transcripts* » (potentiellement impliqués dans des mécanismes de régulation et de « *silencing* »), les SUTs pour « *stable unannotated transcripts* » (fonction inconnues pouvant être proche de celles des CUTs) ou encore les XUTs pour « *Xrn1-dependant unstable transcripts* » (van Dijk et al., 2011) et les RUTs « *Reb1-dependent unstable transcripts* ».

C.2.1. Mécanisme dépendant du NNS

Chez *S. cerevisiae*, le complexe NNS permet la terminaison de la transcription des snoRNAs, des snRNAs et des CUTs (Arigo et al., 2006). Certains ARNm de moins de 1kB, comme *CYC1* sont également terminés par cette voie indépendante du complexe de polyadénylation (Steinmetz et al., 2006).

Le complexe NNS est recruté par interaction avec la forme P-SER^5 du CTD et est formé par : Nrd1, Nab3 et Sen1 dont l'organisation est présentée dans la **Figure 1.24** . Nrd1 est une protéine liant l'ARN présentant un CID en N-terminal. Le domaine CID de Nrd1 interagit également avec le motif NIM (« *Nrd1 Interaction Motif* ») de la PAPnc Trf4 du complexe TRAMP (Tudek et al., 2014). Comme Nrd1, Nab3 est une protéine interagissant avec l'ARN, et qui interagit génétiquement avec Ctk1 – phosphorylation des SER^2 du CTD (Conrad et al., 2000). Sen1 est une protéine de 253kDa, codée par un gène essentiel et présente en faible quantité dans le noyau (≈125 copies/cellule) (Ghaemmaghami et al., 2003). Sen1 possède une activité ADN/ARN hélicase de la superfamille 1 capable d'initier seule la terminaison de transcription *in vitro*, et dont l'activité requiert la présence d'ATP et la reconnaissance du transcrit (Martin-Tumasz and Brow, 2015; Porrua and Libri, 2013). L'activité de Sen1 est proche de celle observée pour le facteur Rho chez les procaryotes. L'analyse par PAR-CLIP a permis de définir les séquences consensus d'interaction de Nrd1 et Nab3 sur l'ARN – UGUAG/A et GNUCUUGU respectivement (Steinmetz et al., 2006). Nrd1 est recrutée très tôt lors de l'initiation de la transcription et son recrutement est impliqué dans la terminaison précoce de certains ARNs, par reconnaissance de l'état de phosphorylation du CTD (Lunde et al., 2010). Cette terminaison précoce est utilisée en cas de réponse au stress, comme la floculation (Singh et al., 2015).

Le modèle général proposé est que Nrd1 et Nab3 jouent le rôle d'adaptateurs, et reconnaissent les séquences terminatrices sur l'ARN et permettent le recrutement et la translocation de Sen1 (Steinmetz et al., 2001). Sen1 va dérouler les hybrides ADN/ARN

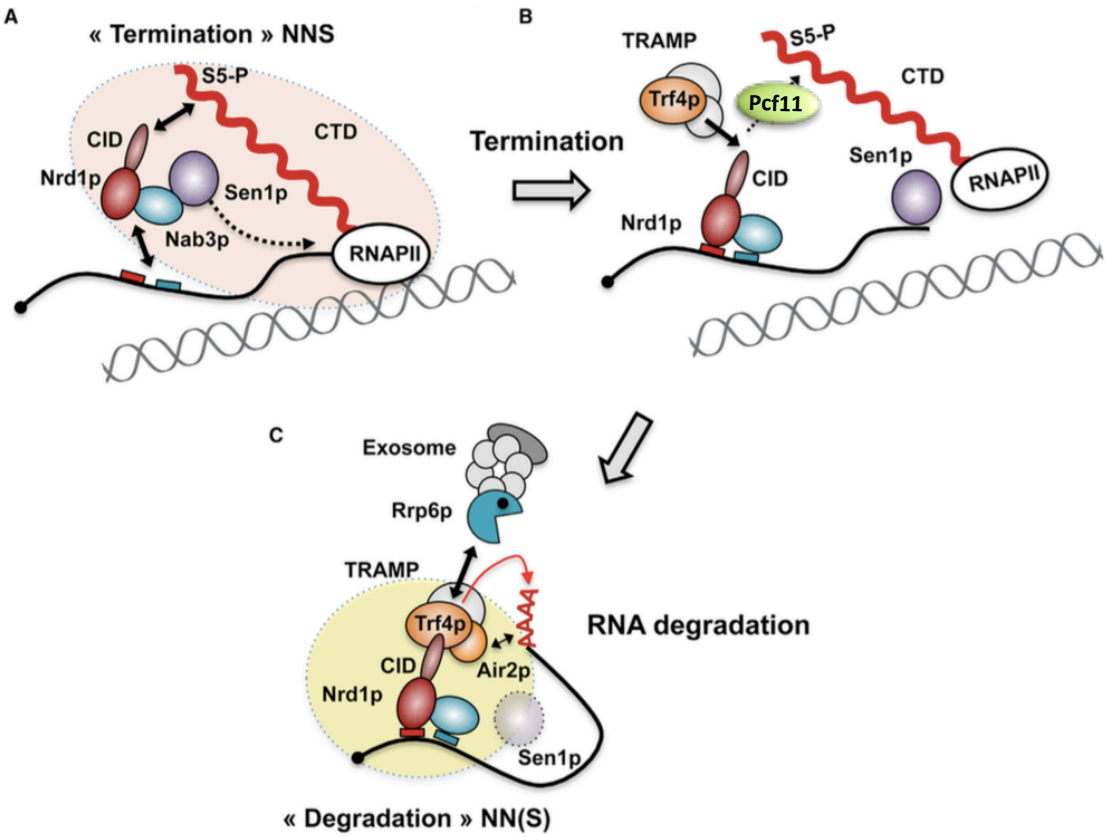


Figure 1.25 : Deux types de complexe NNS sont impliqués dans la terminaison de transcription des ARNs non codants. **(A.)** Le complexe **NNS de terminaison** (Nrd1, Nab3 et Sen1) est recruté par interaction avec la forme P-Ser5 du CTD de l'ARN polymérase II (RNAPII). L'activité hélicase de Sen1 permet le décrochage de l'ARN pol II. **(B.)** L'interaction entre le CTD et le CID de Nrd1 est modifiée par le recrutement de Pcf11. Le CID de Nrd1 est libéré et interagit avec Trf4 (domaine NIM) permettant le recrutement du complexe TRAMP. On parle de complexe **NNS de dégradation**. **(C.)** Le complexe TRAMP permet la polyadénylation de l'ARN non codant et le recrutement de l'exosome, et ainsi la dégradation de l'ARN non codant. D'après Tudek *et al.*, 2014.

formés entre l'ARN naissant et le brin d'ADN complémentaire, ce qui va déstabiliser la polymérase par interaction avec Rpb1. L'action de Sen1 requiert la phosphorylation des résidus SER^2 du CTD (Chinchilla et al., 2012) Cependant, les études récentes menées contribuent à complexifier le modèle proposé, en fonction des ARNs cibles.

Les CUTs n'ont pas de fonctions biologiques connues. Ils sont rapidement dégradés par les différentes voies cellulaires après leur terminaison par le complexe NNS, via le complexe TRAMP et l'exosome. Cela suggère une association étroite entre le complexe NNS et le complexe TRAMP/Exosome. Tudek et, en 2014 (Tudek et al., 2014) ont décrit un modèle comprenant deux formes du complexe NNS, en fonction du partenaire du CID de Nrd1 (**Figure 1.25**). Ces complexes sont formés de manière séquentielle et permettent le décrochage de l'enzyme et la dégradation des transcrits produits. Un premier, comprenant Nrd1/Nab3/Sen1 où CID-Nrd1 interagit avec le CTD de l'ARN polymérase II, permet le décrochage de l'ARN polymérase II par action de Sen1. Un second, formé de Nrd1/Nab3/Sen1 où CID-Nrd1 interagit avec le complexe TRAMP, via Trf4, permet la polyadénylation de dégradation des transcrits et leur élimination par l'exosome.

De nombreuses études ont précédemment décrit la terminaison de transcription des snoRNAs comme dépendante de Pcf11 (Sadowski et al., 2003). Une étude récente (Grzechnik et al., 2015) a défini la sous-unité Pcf11 comme nécessaire pour la terminaison des snoRNAs, des snRNAs, des CUTs et de certains SUTs/XUTs. Nrd1 est recrutée en début de transcription via la forme P-SER^5 du CTD. Pcf11 est recrutée en fin de transcription, via la forme P-SER^2 du CTD. L'interaction de Nrd1 stimule le recrutement de Pcf11 et inversement ; cela permet d'abolir l'interaction entre Nrd1 et le CTD. Ce mécanisme d'interactions séquentielles permet de prévenir l'élongation et d'induire la pause de la polymérase II. Le recrutement de Pcf11 permettrait de restreindre le recrutement de Nrd1 dans les régions plus en amont. Son recrutement permet également d'améliorer la phosphorylation du P-SER^2 du CTD, indispensable pour l'action de Sen1. Sen1 peut ensuite dérouler les hybrides ADN/ARN et déstabiliser la polymérase. Enfin, Pcf11 permettrait le désassemblage de la polymérase en fin de réaction (Zhang and Gilmour, 2006).

Un second modèle général peut ainsi être proposé, plus détaillé. Le complexe NNS est recruté lors de l'initiation de la transcription, par interaction avec la forme P-SER^5 du CTD. Lors de la phosphorylation du résidu SER^2 du CTD, Pcf11 est recrutée, permettant la libération du CID de Nrd1. L'interaction du CTD avec Pcf11 permet de protéger le CTD de

l'action de Fcp1, P-SER[2] phosphatase. Cela conduit : (i) au recrutement du complexe TRAMP/Exosome, via le CID de Nrd1 et (ii) l'action hélicase de Sen1 sur les hybrides ADN/ARN et le décrochage de la polymérase. Le recrutement du complexe TRAMP/Exosome permet la dégradation des ARNnc de types CUTs et la maturation des snoRNAs et des snRNAs.

C.2.2. Autres mécanismes de terminaison des ARNnc chez la levure

L'analyse globale décrite par Grzechnik *et al.*, en 2015 (Grzechnik et al., 2015) a permis de montrer que la terminaison de transcription de la majorité des ARNs non codants était réalisée par le complexe NNS – Pcf11 dépendant ou Nrd1 dépendant - snoRNAs (74% et 85%), CUTs (66% et 69%), SUTs (70% et 32%) et XUTs (72% et 33%).

La terminaison de transcription de SUTs et XUTs a été décrite comme dépendant du complexe de polyadénylation (Marquardt et al., 2011; Tudek et al., 2015; Tudek et al., 2014).

C.3. Autres mécanismes de terminaison de transcription chez *S. cerevisiae*

Deux autres mécanismes de terminaison de transcription des ARNs codants et non codants ont été décrits chez *S. cerevisiae.* Ces voies sont mises en jeu en cas de défauts de la voie dépendante du complexe de polyadénylation.

La première voie Rnt1 dépendante est une variation du modèle ***torpedo*** et permet de prévenir un défaut de celle-ci – site de clivage faible ou mutations de Rat1 (Rondon et al., 2009). La voie dépendante de Rnt1 permet également la terminaison « naturelle » de certains gènes, comme *NPL3* et la régulation de son expression. Rnt1 (homologue de la RNAse III, impliquée dans la maturation des snoRNAs et des snRNAs) joue le rôle de l'endonucléase Ysh1/Brr5 et permet de former un point d'entrée pour Rat1. Rnt1 interagit avec Sen1, qui faciliterait la terminaison dépendante de Rnt1, en s'assurant que l'ARN naissant est correctement replié (Ursic et al., 2004). Le devenir de l'ARN terminé par la voie dépendante de Rnt1 dépend de la présence d'un signal de polyadénylation. En présence d'un site de clivage faible, l'ARN est polyadénylé puis exporté vers le cytoplasme. L'absence de sites de clivage conduit à sa dégradation par la voie TRAMP.

Récemment, une seconde voie de terminaison fait intervenir Reb1 chez *S. cerevisiae*, protéine liant le double brin d'ADN, a été décrite (Colin et al., 2014), selon un modèle proche de celui décrit pour la terminaison de transcription de l'ARN polymérase I. Reb1 reconnaît la séquence GCCGGGGTAA, et permet de bloquer la progression de l'ARN polymérase II dans

les zones NFRs (« *Nucleosome Free Regions* ») (Reeder and Lang, 1994). Le décrochage de la polymérase est permis par l'action de deux ubiquitine-ligases, Rsp5 et Elc1/Cul3. Les ARNs transcrits par cette voie sont dégradés via le complexe TRAMP et l'ARN polymérase II est dégradée par le protéasome.

C.4. Interaction fonctionnelle entre les deux mécanismes

Afin de permettre une terminaison de transcription correcte, et le maintien des limites des unités de transcription, les voies de terminaisons principales – dépendante du complexe de polyadénylation et dépendante du complexe NNS – doivent être correctement définies et connectées. L'une assurant la terminaison en cas de défaut de l'autre (Rondon et al., 2009).

La spécificité de chacune des voies est définie par la reconnaissance des éléments *cis* (UGUAG/A et GNUCUUGU pour la voie NNS et les éléments *cis* de polyadénylation pour la voie dépendante du complexe de polyadénylation) et par le recrutement des facteurs *trans,* via le CTD de la polymérase II. Leur recrutement est strictement dépendant de l'état de phosphorylation du CTD (Gudipati et al., 2008). Les facteurs de la terminaison dépendante du NNS sont recrutés par reconnaissance de la forme P-SER^5 du CTD. Cela permet la terminaison d'ARNs codants ou non codants de moins de 1kB. Les facteurs de la terminaison dépendante du complexe de polyadénylation sont recrutés progressivement durant l'élongation, majoritairement par reconnaissance de la forme P-SER^2 du CTD. Le recrutement des facteurs de terminaison est inhibé par la phosphorylation du résidu TYR^1. Cela permet la terminaison d'ARNs de plus de 1kB. La phosphorylation de SER^2 et la distance par rapport au site d'initiation apparaissent comme cruciales pour le choix entre les différentes voies.
P-SER^5/P-SER^2 sont d'égales intensités à 450pB du TSS et cette zone est considérée comme la zone de transition entre les deux voies principales. De plus, l'inhibition de la phosphorylation de SER^2 permet la terminaison dépendante du complexe NNS d'ARNs de plus de 1kB (Gudipati et al., 2008). Lors de l'initiation et durant l'élongation, la transcription peut être à tout moment terminée par la voie dépendante du NNS, mais dès la fin de l'élongation, la voie dépendante du complexe de polyadénylation prédomine. Chez *S. cerevisiae*, les snoRNAs présentent des éléments *cis* pour la voie NNS et les éléments *cis* de polyadénylation. Leur transcription est terminée majoritairement par la voie dépendante du complexe NNS (Grzechnik et al., 2015). En cas de défaut de cette voie, la transcription continue jusqu'au site

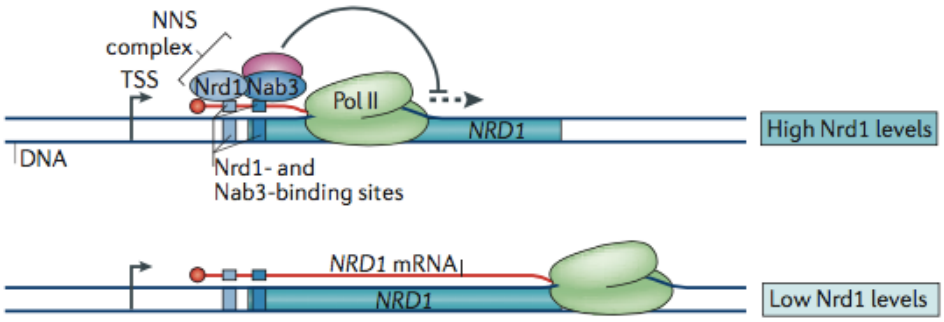


Figure 1.26 : L'expression de *NRD1* est autorégulée par terminaison précoce induite par le complexe NNS. Lorsque Nrd1 est en excès dans la cellule (**A**. High Nrd1 levels), Nrd1 et Nab3 se fixent au niveau de séquence spécifique dans la région 5' du gène. Cela permet le recrutement du complexe NNS et la terminaison de transcription. *NRD1* n'est pas exprimé. Une faible quantité de Nrd1 dans la cellule (**B**. Low Nrd1 levels) ne permet pas la fixation de Nrd1 et Nab3. La terminaison de transcription est terminée par la voie dépendante du complexe de polyadénylation. L'ARNm produit est exporté et traduit. D'après Porrua et Libri, 2015.

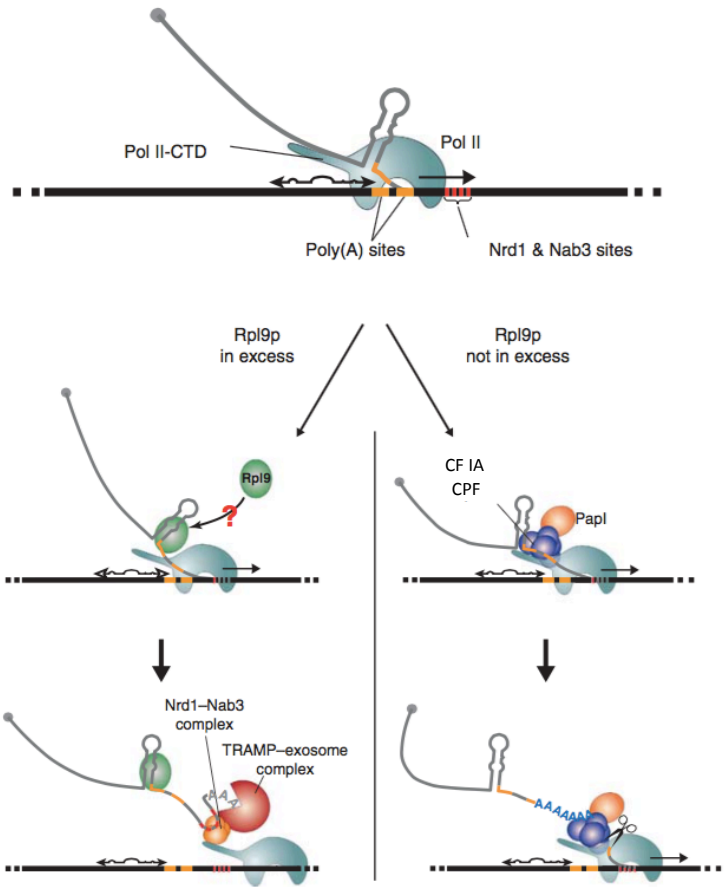


Figure 1.27 : L'accessibilité de site de polyadénylation permet de réguler l'expression de *RPL9*.

La région 3'-UTR de *RPL9* est formée de séquences terminatrices pour la voie dépendante du complexe de polyadénylation (*poly(A) site*) et pour la voie dépendante du NNS (*Nrd1 et Nab3 sites*). Ces séquences sont précédées d'une structure tige boucle. Lorsque Rpl9 est en faible quantité dans la cellule (*à droite*), les sites de polyadénylation sont accessibles et la terminaison est terminée par la voie dépendante du complexe de polyadénylation. Lorsque Rpl9 est en excès dans la cellule (*à gauche*) la fixation d'une protéine sur la structure tige boucle bloque l'accès au site poly(A) et la transcription est terminée par la voie dépendante du NNS. L'ARN naissant est dégradé par recrutement du complexe TRAMP/exosome. **?** indique que l'interaction de la protéine interagissant avec la structure tige boucle n'est pas connue – son interaction serait permise par Rpl9.

D'après Gudipati et al, 2012.

de clivage permettant d'activer la voie dépendante du complexe de polyadénylation (Kim et al., 2006).

Certains facteurs *trans* peuvent être impliqués dans les deux voies. C'est le cas de Ssu72 qui a été montré comme impliquée dans la voie de terminaison dépendante du complexe de polyadénylation et dans celle dépendante du complexe NNS (Steinmetz and Brow, 2003). Pcf11 est également un exemple de l'interdépendance des voies, puisque un défaut de l'interaction avec le CTD conduit à un défaut de terminaison des ARNs codants et des ARNs non codants de type snoRNAs.

Nrd1 et Nab3 ont été montrés comme jouant le rôle de régulateur de l'expression canoniques et cryptiques. C'est le cas par exemple de l'expression de *NRD1* (**Figure 1.26**). Lorsque Nrd1 est présente en forte quantité dans la cellule, le complexe NNS se fixe au niveau des éléments *cis* présent dans la région promotrice du gène. Cela conduit à une terminaison précoce et une baisse de l'expression de *NRD1*. En cas de faible quantité de Nrd1 dans la cellule, le complexe NNS n'est pas retrouvé comme interagissant avec la région 5' et l'expression n'est pas inhibée.

Le choix entre chacune des voies de terminaison permet également de contrôler l'expression de certains gènes chez *S. cerevisiae* comme la protéine ribosomique RPL9B (Gudipati et al., 2012). L'ARN codant est terminé par la voie dépendante du complexe de polyadénylation, ce qui permet la formation d'un ARNm fonctionnel, exporté puis traduit. Lorsque RPL9 est en excès, la structure tige-boucle situé sur l'ARN naissant à proximité du site de polyadénylation – 150 à 200 nucléotides de la fin de l'ORF – permet la fixation de RPL9. Le site de clivage est inaccessible et le complexe NNS peut terminer la transcription (**Figure 1.27**). L'ARN produit n'est pas fonctionnel et est dégradé par les complexes TRAMP/Exosome. L'expression de NPL3 est également régulée par une sélection des voies de terminaison. NPL3 est naturellement terminé par la voie dépendante du complexe de polyadénylation. La présence de Npl3 en forte quantité dans la cellule, conduit à sa fixation au niveau du site de clivage, le rendant inaccessible. La transcription est terminée par la voie Rnt1-dépendante (Rondon et al., 2009).

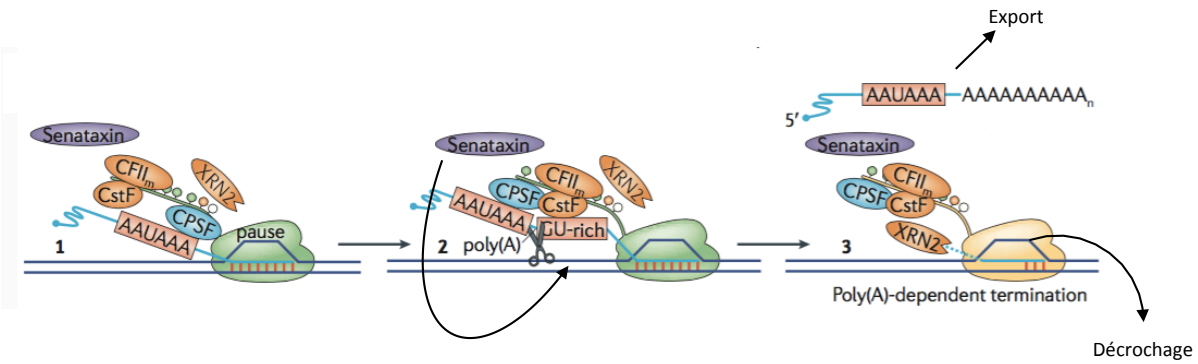


Figure 1.28 : Modèle de terminaison **torpedo** chez l'Homme. **(1.)** L'ARN polymérase II (en **vert**) ralentie par recrutement du complexe de polyadénylation et interaction du sous-complexe CPSF – recrutés par interaction avec l'hexamère AAUAAA et interaction la forme P-Ser2 du CTD – avec l'enzyme. **(2.)** La transcription de la séquence riche en GU (*GU-rich*) permet l'assemblage du complexe de polyadénylation, le clivage et la polyadénylation. L'ARN est exporté vers le cytoplasme. La Senataxine (SETX), homologie de Sen1, déroule les hybrides ARN/ADN. **(3.)** SETX permet le passage de Xrn2 qui va dégrader l'ARN naissant et permettre le décrochage de l'ARN polymérase II. D'après Kuehner et al, 2011.

C.5. Terminaison de la transcription chez les métazoaires

De la même manière que chez *S. cerevisiae*, plusieurs voies de terminaison de transcription sont décrites, mais ne sont pas pleinement connues. La terminaison est associée à la pause de l'ARN polymérase en fin d'élongation (Gromak et al., 2006).

Pour les ARNs codants, comme chez *S. cerevisiae* deux modèles prédominent, dépendant du complexe de polyadénylation : un modèle ***allostérique*** (le clivage n'est pas requis pour la terminaison, et la polymérase est décrochée par un changement de conformation) et un modèle ***torpedo*** (le clivage fourni un point d'entrée pour l'exonucléase Xrn2 qui va déstabiliser l'ARN polymérase). Dans le modèle ***torpedo***, l'action de Xrn2 est dépendante du recrutement de l'homologue de Sen1, la senataxine (SETX), qui déstabilise les hybrides ADN/ARN (Kim et al., 2004; Kuehner et al., 2011) (**Figure 1.28**). Zhang *et al.* (Zhang et al., 2015) suggèrent que le choix entre ces deux voies dépend de la « force » du site de clivage ; un signal plus faible conduisant à une terminaison selon le modèle ***torpedo***. Le cas particulier des ARNs codants pour les histones, réplication-dépendant, est décrit dans le Paragraphe A.3.2.

Il n'existe pas d'homologue de Nrd1 et Nab3 chez les métazoaires et la transcription des snRNAs est terminée par un modèle proche du modèle ***torpedo***. Lors de l'élongation des snRNAs, la phosphorylation de SER^7 prédomine, permettant le recrutement du complexe « *Integrator* ». Les snRNAs sont clivés par action de deux sous-unités homologues de CPSF73 et CPSF100. La polymérase est décrochée par l'action de l'homologue de Sen1, la senataxine et de Xrn2 (Skourti-Stathaki et al., 2011). Le sous-complexe de polyadénylation CF II_m et le complexe CBC/ARS2 sont également impliqués (Hallais et al., 2013). Les snoRNAs sont majoritairement présents dans les régions introniques et sont maturés après l'épissage.

C.6. Terminaison de transcription et réparation de l'ADN

Une cellule eucaryote subit près de 10^5 évènements de lésions spontanés de l'ADN par jour. Ces altérations peuvent être provoquées par la réplication de l'ADN lors du cycle cellulaire (mauvaise incorporation ou dé-purination) ou par des agents environnementaux (chimiques ou physiques). Pour maintenir l'intégrité du génome et permettre la transmission fidèle du message génétique, une voie de signalisation est mise en place, la « *DNA Damage Response* » ou DDR. Chaque type de lésions est réparé par une voie spécifique (Ciccia and Elledge, 2010). Les lésions simples, mauvaise incorporation ou altération chimique, sont

réparées par les voies MMR (« *MisMatch Repair* ») et BER (« *Base Excision Repair* »). Les lésions plus complexes produites lors de la réplication et la transcription de l'ADN sont réparées par la voie NER (« *Nucleotide Excision Repair* »). Cette voie permet l'élimination d'un oligonucléotide de 30 pb contenant la région endommagée. On distingue deux voies : la voie GG-NER (« *Global Genome Repair* ») qui permet l'élimination des lésions sur n'importe quelle zone du génome et la voie TC-NER (« *Transcription-Coupled Repair* ») qui permet l'élimination de lésions au niveau du brin transcrit d'une région active. La réparation des coupures simples brins et double brins sont permises par respectivement les voies SSBR (« *Single-Strand Break Repair* ») et NHEJ/HR (« *NonHomologous end joining* » / « *Homologous Recombination* »). Un défaut de l'une de ces voies peut conduire à une augmentation de la sensibilité à des agents chimiques. Un défaut dans la synthèse de l'ARN, l'assemblage des mRNPs ou dans l'export vers le cytoplasme (complexe THO/TREX et TREX2) conduit à une augmentation de l'instabilité du génome.

Gaillard et Aguilera, en 2014 (Gaillard and Aguilera, 2014) ont montré que les sous-complexes CF IA et CF IB sont impliqués dans la DDR. Ces deux sous-complexes sont impliqués dans la polyadénylation des ARNm et également dans le décrochage de la polymérase et de l'ARN naissant lors de la terminaison de transcription. Rna15 et Hrp1 seraient impliquées dans la voie GG-NER. Rna14 serait impliquée dans les voies de réparation NER, NHEJ, HR et MMR. Cette étude suggère un rôle important des sous-complexes CF IA et CF IB dans la coordination de la terminaison de transcription, la réplication et la réponse aux altérations de l'ADN.

Rtt103 est un facteur de terminaison de transcription permettant le recrutement du dimère Rat1/Rai1. Srividya *et al.,* en 2012 (Srividya et al., 2012) ont montré que Rtt103 serait associée aux lésions de l'ADN (cinq fois plus présente au niveau des lésions) et ainsi serait impliquée dans la réparation des coupures double brins de l'ADN, via la voie HR. Rtt103 ne serait cependant pas directement impliquée dans la cascade de signalisation HR. Son rôle dans cette voie de réparation reste à ce jour mal caractérisé.

De la même manière, Sen1 a été décrit comme impliquée dans le maintien de l'intégrité du génome (Mischo et al., 2011; Ursic et al., 2004). Sen1, comme son homologue chez l'homme SETX, prévient les évènements de recombinaison pouvant conduire à une instabilité du génome, en limitant la formation des hybrides ADN/ARN (ou « *R-loop* ») entre l'ADN de la boucle de réplication et le transcrit naissant, lors de la transcription (Alzu et al., 2012). La formation de « *R-loop* » interfère avec la réplication de l'ADN et conduit à des

coupures simples brins. Sen1 interagirait avec Rad2, endonucléase impliquée dans la voie NER (Ursic et al., 2004). Récemment, Hatchi *et al.*, en 2015 (Hatchi et al., 2015) ont montré que SETX était recrutée au niveau du site de pause de l'ARN polymérase II, via BRCA1. Cela préviendrait la formation de lésions de l'ADN lié à la formation de « *R-loop* ». La mutation de BRCA1 est associée à l'augmentation des risques de développement de cancers du sein et des ovaires. Cette mutation conduit à une augmentation des évènements d'insertions/délétions au niveau de sites de terminaison normalement régulé par BRCA1.

D. MODIFICATIONS POST-TRADUCTIONNELLES DE LA MACHINERIE DE POLYADENYLATION

La modification de l'extrémité 3' est une étape conservée, permettant la formation de tous les ARNm matures, à l'exception des ARNs codant pour les histones chez les métazoaires maturés par l'action endo/exonucléases. Cette étape est fortement régulée par la sélection et la reconnaissance des éléments *cis*, le recrutement et la fixation des facteurs *trans* et également par le biais de modifications post-traductionnelles (ou *PTMs*). Ces PTMs contribuent à moduler l'activité, la localisation intracellulaire, la capacité d'interaction et/ou la stabilité des facteurs et régulateurs de la polyadénylation. Plusieurs types de modifications post-traductionnelles (Ryan and Bauer, 2008) sont décrits dans la littérature, chez *S. cerevisiae* ou dans le système mammifère, et contribuent à la régulation de la polyadénylation des ARNm. La majorité des mécanismes de régulation décrits ciblent, directement ou indirectement, les poly(A)polymérases Pap1 et PAP.

- la phosphorylation de résidus Sérines/Thréonines/Tyrosines
- la méthylation
- l'acétylation de résidus Lysines
- la sumoylation de résidus Lysines
- l'ubiquitination

D.1. Phosphorylation

La phosphorylation est la modification post-traductionnelle la plus commune dans les cellules eucaryotes (30% du protéome) ; c'est donc naturellement la plus étudiée. La phosphorylation est une modification réversible faisant intervenir deux types enzymes : kinases et phosphatases (Stark, 1996). Les kinases sont des enzymes du groupe des

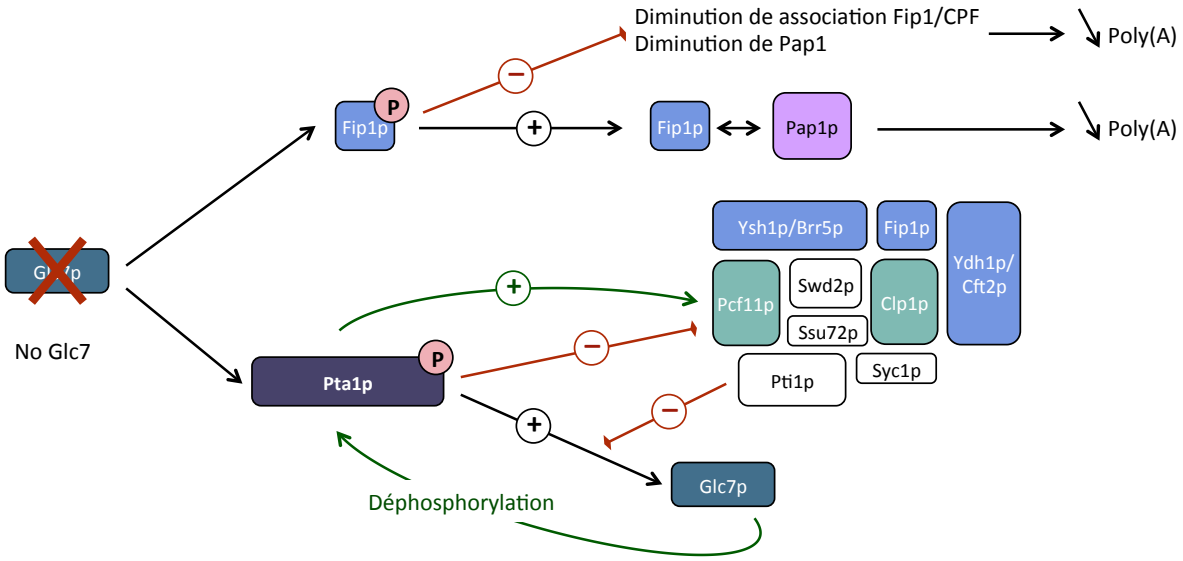


Figure 1.29 : Glc7 est un acteur clef de la régulation de la maturation de l'extrémité 3'. L'absence de **Glc7** conduit à l'accumulation de formes phosphorylées de **Pta1** et **Fip1**. La phosphorylation de Pta1 modifie son interaction avec les différents partenaires des complexes CF IA et CPF (flèche **rouge**) et permet la recrutement de Glc7. Cela permet la déphosphorylation de Pta1 et l'association de Pta1 avec le complexe CF IA et CPF (flèches **vertes**). Pti1 inhibe le recrutement de Glc7. La phosphorylation de Fip1 permet d'augmenter son interaction avec **Pap1**, et diminue sa processivité. La phosphorylation de Fip1 permet également de diminuer son association avec le complexe CPF et l'association de Pap1 (flèche **rouge**). L'étape de polyadénylation est inhibée.

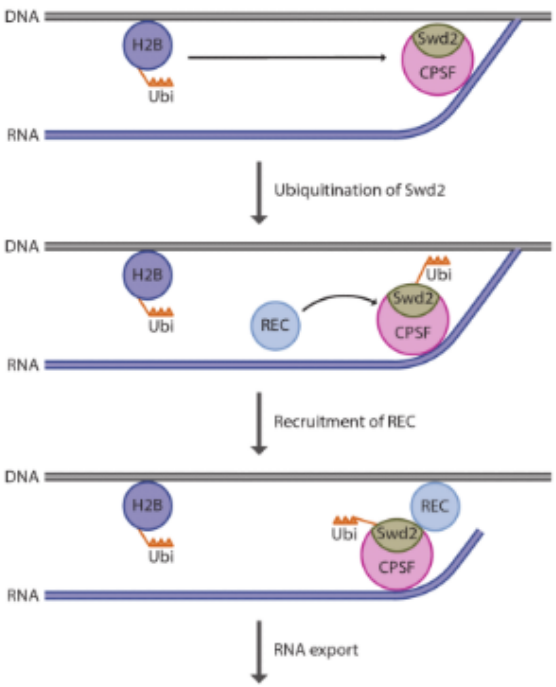


Figure 1.30 : L'ubiquitination de Swd2 permet l'export des ARNm. L'ubiquitination (**Ubi**) de l'histone H2B au niveau de l'ADN (en **gris**) permet l'ubiquitination de **Swd2**, sous-unité du complexe CPF situé au niveau de l'ARN (en **violet**). Swd2-Ubi permet le recrutement de complexe d'export des ARNm (**REC** – Mex67/Mtr2 et Yra1/Nab2) et l'export des mRNPs formées.

transférases qui catalysent l'addition d'un groupement phosphate sur un substrat. Plusieurs mécanismes de régulation de la maturation en 3' sont décrits dans les systèmes levures et mammifères.

L'exemple de la régulation de la capacité d'interaction par phosphorylation le plus connu et le mieux décrit est le cycle de phosphorylation/déphosphorylation du CTD durant la transcription ou « code CTD ». Cela permet le recrutement et le décrochage des différentes protéines nécessaires aux réactions. Il a également été montré que l'assemblage du complexe CPF chez *S. cerevisiae* pouvait être modulé par phosphorylation (**Figure 1.29**). La déplétion de Glc7 conduit à l'accumulation de formes phosphorylées de Pta1, protéine d'architecture, modifiant les interactions avec ses différents partenaires. Pta1, sous sa forme phosphorylée, permettrait le recrutement de Glc7, et ainsi sa déphosphorylation. Pta1 déphosphorylée interagit avec ses partenaires du complexe CPF. Pti1 est proposée comme régulateur négatif du recrutement de Glc7 par interaction protéine-protéine (He and Moore, 2005; Kyburz et al., 2003).

De nombreuses études ont montré que la phosphorylation permettait de réguler l'activité de polyadénylation, portée par Pap1 chez *S. cerevisiae*. Plusieurs mécanismes entrent en jeu. Tout d'abord, Pap1 elle-même est la cible de phosphorylation durant le cycle cellulaire (Mizrahi and Moore, 2000). La phosphorylation d'une partie du *pool* de Pap1 en phase S et G2 conduit à une diminution l'activité globale de maturation. Cela permettrait la régulation de l'expression de gènes, non requis en phase de synthèse – gènes codants pour les histones. Ensuite, l'activité de Pap1 peut être régulée indirectement, par phosphorylation de ces partenaires du complexe CPF (**Figure 1.29**). Pap1 est amenée au complexe CPF par interaction avec Fip1. L'interaction de Fip1 avec Pap1 a un effet inhibiteur sur son activité. La phosphorylation de Fip1 par CK2 augmente son affinité pour Pap1, ce qui va diminuer/inhiber son activité (Zielinski et al., 2006). He et Moore (He and Moore, 2005) ont également montré qu'une accumulation de la forme phosphorylée de Fip1 – par déplétion de Glc7 – diminue son association avec le complexe CPF et conduit à une diminution de la polyadénylation. Glc7 régule indirectement la polyadénylation, en permettant l'association de Pap1 avec le complexe CPF et en régulant son activité, par déphosphorylation de Fip1. La régulation de l'activité de Pap1 pourrait former un point de contrôle entre le clivage et la polyadénylation.

La phosphorylation peut également influencer la localisation intracellulaire, comme c'est le cas pour le facteur d'export Npl3. Sous sa forme phosphorylée, Npl3 est adressée au

noyau. La déphosphorylation par Glc7 permet son export vers le cytoplasme pour un nouveau cycle d'export. Une accumulation de la forme phosphorylée de Npl3 conduit à l'accumulation de Rna15 sur le pré-ARNm, pouvant modifier l'une ou l'autre des étapes de maturation en 3' (He and Moore, 2005).

Différentes études ont également illustré le rôle de la phosphorylation dans le système mammifère. De la même manière que Pap1, la poly(A)polymérase PAP est régulée par des cycles de phosphorylation/déphosphorylation pendant le cycle cellulaire. En phase de mitose, PAP est hyper-phosphorylée par action de la kinase cdc2/cycline B ; cela permet l'inhibition de son activité. Cette régulation est supposée faciliter la condensation des chromosomes et leur ségrégation (Colgan and Manley, 1997). PABPN1, protéine interagissant avec la queue poly(A) est également une cible de phosphorylation (Amanchy et al., 2005). L'interaction entre CF I_m-25 et PAP est également régulée par phosphorylation de CF I_m-25 au niveau de quatre résidus ce qui va conduire à un changement de conformation de cette protéine (Ryan and Bauer, 2008).

D.2. Ubiquitination

L'ubiquitination fait intervenir trois activités enzymatiques différentes (portées par trois familles d'enzymes E1, E2 et E3) et correspond à la fixation covalente d'une ou plusieurs protéines d'ubiquitines (8,5kDa) sur des résidus Lysines. L'ubiquitination est une modification réversible, par action des DUBs - déubiquitinases. Généralement, cette PTM permet la dégradation de la protéine par le protéasome et la régulation de la demi-vie. L'ubiquitination des protéines peut aussi conduire à des procèdes indépendants de la protéolyse. Elle permet également la régulation de la localisation intracellulaire, de l'activité d'une protéine, de l'interaction d'une protéine avec ses différents partenaires, des voies de signalisation, ou encore le remodelage de la chromatine. Plusieurs études ont décrit des mécanismes de régulation de l'export par ubiquitination (Iglesias et al., 2010).

Plusieurs sous-unités du complexe de polyadénylation ont été impliquées dans des mécanismes d'ubiquitination. Chez *S. cerevisiae*, l'activité de Pap1 est directement régulée lors du cycle cellulaire par phosphorylation mais également par ubiquitination, en phase S et G2 (Mizrahi and Moore, 2000). Cela permettrait de diminuer la quantité de Pap1 présente dans la cellule, pour réduire les activités de maturation. Pap1 peut être dégradée par le

protéasome ou l'activité peut être modulée par des protéines liant spécifiquement ces PTMs, comme Ufd1 (del Olmo et al., 1997). Récemment, Mpe1 a été montrée comme impliquée indirectement dans les processus d'ubiquitination – la mutation de Mpe1 induit une diminution de l'association d'ubiquitine avec Pap1 (Lee and Moore, 2014).

Une étude, publiée en 2012 (Vitaliano-Prunier et al., 2012) a permis de montrer que l'export et de maturation de l'extrémité 3' était interconnectés, notamment par ubiquitination (Gwizdek et al., 2006). Il a été montré que l'ubiquitination de Swd2 par Bre1, induite par ubiquitination de H2B, faciliterait le recrutement de Yra1 et Nab2 et l'assemblage des mRNPs (**Figure 1.30**).

Il est également important de noter que Rpb1 est ubiquitinée par l'action de Rsp5. Cela permet la régulation des complexes ARN polymérases pouvant présenter des défauts d'assemblage et permet de contrôler le niveau d'expression intracellulaire.

D.3. Sumoylation

La sumoylation correspond à la liaison covalente d'une ou plusieurs protéines SUMO – pour « *small ubiquitin-like modifier* », protéine de 12kDa, structuralement proche de l'ubiquitine – sur un résidu Lysine. De la même manière que l'ubiquitination, cette PTM fait intervenir trois types d'enzymes, E1-SUMO, E2-SUMO et E3-SUMO. C'est une modification réversible, par l'action de protéases de la famille des SENP. La sumoylation permet de modifier les propriétés biochimiques d'une protéine.

Dans la maturation en 3', CPSF-73 ou la Simplekin ont été décrits comme cibles pour la sumoylation, par SUMO-2/3 (Vethantham et al., 2007). Cette étude montre que la sumoylation de la Simplekin est essentielle pour assurer la viabilité cellulaire – l'ajout de protéines non sumolyées ne permet pas de complémenter la délétion de la Simplekin. Les mécanismes sous-jacents ne sont pas identifiés, mais la sumoylation de ces deux protéines est impliquée dans les activités de maturation 3' *in vitro.* Vethantham *et al.,* (Vethantham et al., 2008) ont décrit un rôle de la sumoylation dans la régulation de l'activité de PAP, par SUMO-2/3. *In vitro*, l'activité de PAP est diminuée par ajout de protéines SUMO. La sumoylation de PAP permet également d'augmenter la stabilité de la protéine et de réguler sa localisation. La mutation des sites de sumoylation ou la surexpression de SENP1 conduit à une rétention de PAP dans le cytoplasme. La localisation nucléaire de PAP est SUMO-dépendante.

Chez *S. cerevisiae*, de nombreuses protéines comme Ref2 ou Ydh1/Cft2 (Panse et al., 2004) ont été identifiées comme substrat potentiel de sumoylation. Pap1 a été montrée comme interagissant avec Uba2, E1-SUMO (del Olmo et al., 1997). Cependant aucune régulation de la polyadénylation, chez la levure, par sumoylation n'est pour l'instant décrite.

D.4. Acétylation / Méthylation

Chez les métazoaires, l'acétylation (ajout d'un groupement $–CO-CH_3$) est généralement impliquée dans la régulation de la transcription, en modifiant l'état de condensation de la chromatine. L'acétylation est étroitement liée à l'ubiquitination, pouvant présenter des fonctions antagonistes ou coopérer pour assurer une même fonction cellulaire.

Dans le cas de la maturation en 3', des études ont montré que l'acétylation de la PAP modifie son interaction avec les importines α/β et conduit à sa rétention nucléaire. Cela modifie également son interaction avec CF I_m-25, impliquée dans sa stabilisation au niveau du complexe de polyadénylation (Ryan and Bauer, 2008).

La méthylation (ajout d'un groupement méthyle $-CH_3$) peut aussi bien cibler l'ADN, l'ARN que les protéines, au niveau des résidus Arginines. La méthylation de l'ADN est un processus épigénétique permettant notamment la régulation de l'expression des gènes. La méthylation d'une région promotrice conduit à l'inhibition de l'expression du gène en aval. Dans le système murin, la méthylation de l'ADN permet l'expression sélective des allèles parentaux, et la modification du choix du site de polyadénylation (Wood et al., 2008).

Le méthylation de l'ARN au niveau des résidus adénosines – N^6-méthyladensine ou m^6A – est la modification de l'ARN la plus fréquente, retrouvée dans les régions 3'-UTR. Depuis quelques années, cette modification de l'ARN est proposée comme permettant la régulation post-transcriptionnelle des ARNs, par recrutement de protéines adaptatrices et à terme la modulation de l'expression des protéines (Fu et al., 2014; Yue et al., 2015). Chez *S. cerevisiae*, la m^6A n'est retrouvée que lors de la phase S et de la phase G2 du cycle cellulaire.

Lee et Bedford, en 2002 (Lee and Bedford, 2002) ont montré que PABP1, homologue de Pab1, était le substrat de l'arginine méthyltransférase CARM1. L'impact de cette modification sur l'interaction avec le queue poly(A) reste inconnue.

MATERIELS & METHODES

MATERIELS ET METHODES

A. LEVURES : SOUCHES UTILISEES, VECTEURS ET MILIEUX DE CULTURE

A.1. Souches utilisées

Le nom, le génotype et l'origine des souches de *Saccharomyces cerevisiae* utilisées dans cette étude sont répertoriés dans le *TABLEAU 4.1*.

A.2. Vecteurs

Les caractéristiques et l'origine des vecteurs utilisés sont répertoriées dans le *TABLEAU 4.2*.

A.3. Milieux de cultures

Les milieux sous forme solide contiennent 2% d'agar.

- YPD(A) : 1% (p/v) Extrait de levures, 2% bactopeptone, 2% glucose 2%, 20mg/L adénine.
- SC (Milieu synthétique) : 0,67% Yeast Nitrogen Base w/o a.a., 2% glucose, 10% Drop-out 10X. Le milieu peut être supplémenté par +/- 20mg/L d'adénine, +/- 20mg/L de L-histidine, +/- 20mg/L de L-tryptophane, +/- 20mg/L d'uracile ou +/- 30mg/L de L-leucine. Un ou plusieurs de ces cinq composants sont omis lors des tests de sélection basés sur un ou plusieurs marqueurs d'auxotrophie. Un milieu SC sans leucine sera noté SC-L et un milieu sans ces cinq composants sera noté SC –ALWUH.
- SC + GAL: 0,67% Yeast Nitrogen Base w/o a.a., 2% galactose, 10% Drop-out 10X. De la même manière, le milieu peut être supplémenté par +/- 20mg/L d'adénine, +/- 20mg/L de L-histidine, +/- 20mg/L de L-tryptophane, +/- 20mg/L d'uracile ou +/- 30mg/L de L-leucine.
- Drop-out 10X : 20 mg/L L-arginine, 20 mg/L L-méthionine, 30 mg/L L-tyrosine, 30 mg/L L-isoleucine, 30 mg/L L-lysine, 50 mg/L L-phénylalanine, 150 mg/L L-valine, 200 mg/L L-thréonine, 400 mg/L L-sérine, 100 mg/L acide L-glutamique, et 100 mg/L acide L-aspartique.
- SC - 5-FOA : 0,67% Yeast Nitrogen Base w/o a.a., 2% glucose, 0,1% 5-FOA (acide 5-Fluoro-orotique), 20mg/mL uracile, 0.025M KH_2PO_4.
- Milieu de sporulation : 1% acétate de potassium, 0,1% extrait de levures, 0,05% glucose, 2,5% Drop-out 10X.
- Milieu de conservation à -80% : YPD(A) + 15% glycérol.

TABLEAU 4.1 : Souches de *S. cerevisiae* utilisées dans cette étude.

CONSTRUCTION	SOUCHES	GENOTYPES	ORIGINE
	BMA64-1A	*MATa, ade2-1 leu2-3,112 ura3-1 trp1Δ his3-11,15 can1-100 GAL+*	Baudin-Baillieu (1997)
	BMA64-1B	*MATα, ade2-1 leu2-3,112 ura3-1 trp1Δ his3-11,15 can1-100 GAL+*	Baudin-Baillieu (1997)
	YSD10	*MATa, ade2 leu2 ura3 trp1Δ his3, [pNOP1]-TAP::FIP1-TRP1-Kl*	S. Dheur
	YSD12	*MATa, ade2 leu2 ura3 trp1Δ his3, [pNOP1]-TAP::RNA15-TRP1-Kl*	S. Dheur
	YSD39	*MATα, [pNOP1]-TAP::RNA15-URA3, [pNOP1]-TAP::FIP1-TRP1 dans BMA64*	S. Dheur
	YLM206	*ade2-1 leu2-3,112 ura3-1 trp1Δ his3-11,15 can1-100 GAL+, clp1::LEU2, pASZ11-clp1-15, [pNOP1]-TAP::RNA15-URA3*	L. Minvielle-Sébastia
	YLM39-3	*MATa, [pNOP1]-TAP::RNA15-URA3, [pNOP1]-TAP::FIP1-TRP1 dans BMA64*	L. Minvielle-Sébastia
	YLM219-12	*MATa, [pNOP1]-TAP::FIP1-TRP1-Kl, [pNOP1]-ProtA::RNA15-URA3-Kl dans BMA64*	L. Minvielle-Sébastia
	YLM219-21	*MATα, [pNOP1]-TAP::FIP1-TRP1-Kl, [pNOP1]-ProtA::RNA15-URA3-Kl dans BMA64*	L. Minvielle-Sébastia
	YLM221	*MATa, [pNOP1]-TAP::RNA15-TRP1-Kl dans BMA64*	L. Minvielle-Sébastia
	YLM232	*MATa/α, ade2-1/ade2-1 leu2-3,112/leu2-3,112 ura3-1/ura3-1 trp1-1/trp1Δ his3-11,15/his3-11,15 can1-100 GAL+, pcf11::HIS3/pFL38-PCF11*	L. Minvielle-Sébastia
	YLM233	*MATα, pcf11::HIS3, pFL38-PCF11 dans BMA64*	L. Minvielle-Sébastia
	YJG35	*MATa, ade2 leu2 ura3 trp1Δ his3, rna15-1, [pNOP1]-TAP::FIP1-TRP1-Kl*	Ce travail
	YLM225	*MATa, ade2 leu2 ura3 trp1Δ his3, [pNOP1]-TAP::FIP1-URA3-Kl*	L. Minvielle-Sébastia
	YLM126	*MATa , clp1::TRP1 ade2 leu2 ura3 trp1Δ his3 [pHH2-URA3-CLP1]*	L. Minvielle-Sébastia
	YRH37	*clp1::LEU2, ade2-1 leu2-3,112 ura3-1 trp1Δ his3-11,15 can1-100, pASZ11-clp1-15*	R. Haddad
	YLM262	*MATa, [pNOP1]-ProtA-TEV::RNA15-URA3-Kl dans BMA64*	L. Minvielle-Sébastia
WT	YLM235	*pcf11::HIS3, pFL38-PCF11, pFL36ΔSX-PCF11*	Ce travail
	YLM237	*pcf11::HIS3, pFL36ΔSX-PCF11*	Ce travail
	YLM244	*pcf11::HIS3, pFL36ΔSX-PCF11, [pNOP1]-TAP::RNA15-TRP1-Kl,*	Ce travail
	YLM245	*pcf11::HIS3, pFL36ΔSX-PCF11, [pNOP1]-TAP::RNA15-TRP1-Kl,*	Ce travail
	YJG18	*pcf11::HIS3, pFL36ΔSX-PCF11, [pNOP1]-TAP::RNA15-TRP1-Kl, [pNOP1]-TAP::FIP1-URA3-Kl*	Ce travail
	YJG36	*pcf11::HIS3, pFL36ΔSX-PCF11, pGCYC1*	Ce travail
	YJG50	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-PCF11, pHH2-URA3-CLP1*	Ce travail
	YJG51	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-PCF11, pHH2-URA3-CLP1, pASZ11-clp1-1*	Ce travail
	YJG52	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-PCF11, pHH2-URA3-CLP1, pASZ11-clp1-5*	Ce travail
	YJG53	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-PCF11, pHH2-URA3-CLP1, pASZ11-clp1-6*	Ce travail
	YJG57	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-PCF11, pASZ11-clp1-1*	Ce travail
	YJG58	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-PCF11, pASZ11-clp1-5*	Ce travail
	YJG59	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-PCF11, pASZ11-clp1-6*	Ce travail
	YJG66	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-PCF11, pHH2-URA3-CLP1, pASZ11-clp1-17*	Ce travail
	YJG67	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-PCF11, pHH2-URA3-CLP1, pASZ11-clp1-16*	Ce travail
	YJG68	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-PCF11, pHH2-URA3-CLP1, pASZ11-clp1-19*	Ce travail
ΔQ_{20}	YLM239	*pcf11::HIS3, pFL36ΔSX-pcf11ΔQ20*	Ce travail
	YLM246	*pcf11::HIS3, pFL36ΔSX-pcf11ΔQ20, [pNOP1]-TAP::RNA15-TRP1-Kl*	Ce travail
	YLM247	*pcf11::HIS3, pFL36ΔSX-pcf11ΔQ20, [pNOP1]-TAP::RNA15-TRP1-Kl*	Ce travail
	YJG20	*pcf11::HIS3, pFL36ΔSX-pcf11ΔQ20, [pNOP1]-TAP::RNA15-TRP1-Kl, [pNOP1]-TAP::FIP1-URA3-Kl*	Ce travail
	YJG47	*pcf11::HIS3, pFL36ΔSX-pcf11ΔQ20, pGCYC1*	Ce travail

TABLEAU 4.1 : Souches de *S. cerevisiae* utilisées dans cette étude (*Suite*).

Δ(141-233)	YLM236	*pcf11::HIS3, pFL38-PCF11, pFL36ΔSX-pcf11Δ(141-233)*	Ce travail
	YLM238-2	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(141-233)*	Ce travail
	YLM238-3	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(141-233)*	Ce travail
	YLM248	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(141-233), [pNOP1]-TAP::RNA15-TRP1-Kl*	Ce travail
	YLM249	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(141-233), [pNOP1]-TAP::RNA15-TRP1-Kl*	Ce travail
	YLM256	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(141-233), [pNOP1]-TAP::RNA15-URA3-Kl, [pNOP1]-TAP::FIP1-TRP1-Kl*	Ce travail
	YJG17	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(141-233), [pNOP1]-TAP::FIP1-URA3-Kl)*	Ce travail
	YJG39	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(141-233), pGCYC1*	Ce travail
	YLM242	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(141-233), pHH2-ADE2-PCF11*	Ce travail
	YLM243	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(141-233), pHH2-ADE2-pcf11Δ(287-626)*	Ce travail
Δ(155-174)	YJG4	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(155-174)*	Ce travail
	YJG13	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(155-174), [pNOP1]-TAP::RNA15-URA3-Kl*	Ce travail
	YLM257	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(155-174), [pNOP1]-TAP::RNA15-URA3-Kl, [pNOP1]-TAP::FIP1-TRP1-Kl*	Ce travail
	YJG40	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(155-174), pGCYC1*	Ce travail
Δ(180-196)	YJG6	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(180-196)*	Ce travail
	YJG14	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(180-196), [pNOP1]-TAP::RNA15-URA3-KI*	Ce travail
	YLM260	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(180-196), [pNOP1]-TAP::FIP1-TRP1-Kl*	Ce travail
	YJG19	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(180-196), [pNOP1]-TAP::RNA15-URA3-Kl, [pNOP1]-TAP::FIP1-TRP1-Kl*	Ce travail
	YJG42	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(180-196), pGCYC1*	Ce travail
Δ(203-227)	YJG7	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(203-227)*	Ce travail
	YJG15	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(203-227), [pNOP1]-TAP::RNA15-URA3-Kl*	Ce travail
	YLM261	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(203-227), [pNOP1]-TAP::FIP1-TRP1-Kl*	Ce travail
	YLM258	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(203-227), [pNOP1]-TAP::RNA15-URA3-Kl, [pNOP1]-TAP::FIP1-TRP1-Kl*	Ce travail
	YJG43	*pcf11::HIS3, pFL36ΔSX-pcf11Δ(203-227), pGCYC1*	Ce travail
(L189D)	YLM250	*pcf11::HIS3, pFL36ΔSX-pcf11-14 (L189D)*	Ce travail
	YLM253	*pcf11::HIS3, pFL36ΔSX-pcf11-14 (L189D), [pNOP1]-TAP::RNA15-TRP1-Kl*	Ce travail
	YJG44	*pcf11::HIS3, pFL36ΔSX-pcf11-14 (L189D), pGCYC1*	Ce travail
(L213D)	YLM251	*pcf11::HIS3, pFL36ΔSX-pcf11-15 (L213D)*	Ce travail
	YLM254	*pcf11::HIS3, pFL36ΔSX-pcf11-15 (L213D), [pNOP1]-TAP::RNA15-TRP1-Kl*	Ce travail
	YJG45	*pcf11::HIS3, pFL36ΔSX-pcf11-15 (L213D), pGCYC1*	Ce travail
(D220A)	YLM252	*pcf11::HIS3, pFL36ΔSX-pcf11-16 (D220A)*	Ce travail
	YLM255	*pcf11::HIS3, pFL36ΔSX-pcf11-16 (D220A), [pNOP1]-TAP::RNA15-TRP1-Kl*	Ce travail
	YJG46	*pcf11::HIS3, pFL36ΔSX-pcf11-16 (D220A), pGCYC1*	Ce travail
(L189D, L213D)	YJG5	*pcf11::HIS3, pFL36ΔSX-pcf11-17 (L189D, L213D)*	Ce travail
	YJG22	*pcf11::HIS3, pFL36ΔSX-pcf11-17(L189D, L213D), [pNOP1]-TAP::RNA15-TRP1-KI*	Ce travail
	YJG31	*pcf11::HIS3, pFL36ΔSX-pcf11-17(L189D, L213D), [pNOP1]-TAP::RNA15-URA3-KI*	Ce travail
	YJG32	*pcf11::HIS3, pFL36ΔSX-pcf11-17(L189D, L213D), [pNOP1]-TAP::FIP1-TRP1-Kl*	Ce travail
	YJG25	*pcf11::HIS3, pFL36ΔSX-pcf11-17(L189D, L213D), [pNOP1]-TAP::RNA15-URA3-KI, [pNOP1]-TAP::FIP1-TRP1-Kl*	Ce travail

TABLEAU 4.1 : Souches de *S. cerevisiae* utilisées dans cette étude (*Suite*).

(C421S; C424S)	YJG10	*pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S)*	Ce travail
	YJG23	*pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), [pNOP1]-TAP::RNA15-TRP1-KI*	Ce travail
	YJG26	*pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), [pNOP1]-TAP::RNA15-URA3-KI, [pNOP1]-TAP::FIP1-TRP1-Kl*	Ce travail
	YJG37	*pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), pGCYC1*	Ce travail
	YJG48	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), pHH2-URA3-CLP1*	Ce travail
	YJG54	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), pHH2-URA3-CLP1, pASZ11-clp1-1*	Ce travail
	YJG55	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), pHH2-URA3-CLP1, pASZ11-clp1-5*	Ce travail
	YJG56	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), pHH2-URA3-CLP1, pASZ11-clp1-6*	Ce travail
	YJG60	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), pASZ11-clp1-1*	Ce travail
	YJG61	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), pASZ11-clp1-6*	Ce travail
	YJG71	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), pHH2-URA3-CLP1, pASZ11-clp1-16*	Ce travail
	YJG72	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), pHH2-URA3-CLP1, pASZ11-clp1-17*	Ce travail
	YJG73	*clp1:: TRP1, pcf11::HIS3, pFL36ΔSX-pcf11-18(C421S; C424S), pHH2-URA3-CLP1, pASZ11-clp1-19*	Ce travail
(C564S, C567S)	YJG12	*pcf11::HIS3, pFL36-pcf11-19(C564S, C567S)*	Ce travail
	YJG24	*pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), [pNOP1]-TAP::RNA15-TRP1-KI*	Ce travail
	YJG34	*pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), [pNOP1]-TAP::RNA15-URA3-KI*	Ce travail
	YJG33	*pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), [pNOP1]-TAP::FIP1-TRP1-KI*	Ce travail
	YJG38	*pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), pGCYC1*	Ce travail
	YJG49	*clp1:: TRP1, pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), pHH2-URA3-CLP1*	Ce travail
	YJG62	*clp1:: TRP1, pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), pHH2-URA3-CLP1, pASZ11-clp1-1*	Ce travail
	YJG63	*clp1:: TRP1, pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), pHH2-URA3-CLP1, pASZ11-clp1-5*	Ce travail
	YJG64	*clp1:: TRP1, pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), pHH2-URA3-CLP1, pASZ11-clp1-6*	Ce travail
	YJG65	*clp1:: TRP1, pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), pHH2-URA3-CLP1, pASZ11-clp1-17*	Ce travail
	YJG69	*clp1:: TRP1, pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), pHH2-URA3-CLP1, pASZ11-clp1-16*	Ce travail
	YJG70	*clp1:: TRP1, pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), pHH2-URA3-CLP1, pASZ11-clp1-19*	Ce travail
pcf11-13 (D68A, S69A, I70A)	YLM267-1B	*pcf11::HIS3, pFL36ΔSX-pcf11-13*	Sadowski et al, 2003
	YLM267-3D	*pcf11::HIS3, pFL36ΔSX-pcf11-13*	Sadowski et al, 2003
(E199A)	YLM268-1C	*pcf11::HIS3, pFL36ΔSX-pcf11-21(E199A)*	Ce travail
	YLM268-5D	*pcf11::HIS3, pFL36ΔSX-pcf11-21(E199A)*	Ce travail
(R102A)	YLM269-1A	*pcf11::HIS3, pFL36ΔSX-pcf11-22(R102A)*	Ce travail
	YLM269-3C	*pcf11::HIS3, pFL36ΔSX-pcf11-22(R102A)*	Ce travail
	YLM272	*pcf11::HIS3, pFL36ΔSX-pcf11-22(R102A), pGCYC1*	Ce travail
(E199A ; R102A)	YLM273-3D	*pcf11::HIS3, pFL36ΔSX-pcf11-23(E199A ; R102A)*	Ce travail
	YLM274-4A	*pcf11::HIS3, pFL36ΔSX-pcf11-23(E199A ; R102A)*	Ce travail

B. BACTERIES : SOUCHES UTILISEES, VECTEURS ET MILIEUX DE CULTURE

B.1. Souches

Les souches d'*E. coli* utilisées pour l'amplification des vecteurs ainsi que pour la production de protéines recombinantes sont les suivantes :

DH5α (Biolabs – génotype : *fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*).

BL21 (DE3) (Biolabs – génotype : *E. coli B F– ompT hsdSB (rB–, mB–) gal dcm*).

BL21 pLysS (Promega – génotype : *E.coli B F–, ompT, hsdSB (rB–, mB–), gal dcm, λ(DE3), pLysS, Cmr*).

BL21 CodonPlus (DE3)-RIL (Stratagene – génotype : *E.coli B F–, ompT, hsdS(rA–, mA–), gal dcm, argU (AGA, AGG), ileY (AUA), leuW (CUA), Cmr*).

B.2. Vecteurs

Les caractéristiques et l'origine des vecteurs utilisés sont répertoriées dans le *TABLEAU 4.3*.

B.3. Milieux de cultures

Les milieux sous forme solide contiennent 2% d'agar.

Les antibiotiques sont utilisés aux concentrations de : 100µg/mL Ampicilline, 38µg/mL Chloramphénicol, 40µg/mL Kanamycine.

- LB : 0,5% Extrait de levures, 1% bactotryptone et 1% NaCl.
- TB : 1,2% bactotryptone, 2,4% extrait de levures, 0,5% glycérol, 89mM tampon phosphate.
- Milieu de conservation à 80°C : LB + 15% glycérol.

C. TECHNIQUES DE BIOLOGIE MOLECULAIRE

C.1. Transformation des bactéries

C.1.1. Préparation des bactéries électro-compétentes

Les cellules sont cultivées dans 1L de milieu LB, à 37°C, 200rpm, jusqu'à une DO_{600nm} de 0,8. Les cellules sont lavées plusieurs fois avec de grands volumes décroissants d'H_2O stérile à 4°C, puis concentrés dans 2mL d'une solution de glycérol 10% (p/v). Les cellules obtenues sont aliquotées et stockées à -80°C.

TABLEAU 4.2 : Vecteurs levures utilisés dans cette étude.

VECTEURS	CARACTERISTIQUES	ORIGINE
pFL36	From pUC19 - EcoRI/MCS/HindIII - CC LEU2 - ARS-CEN6 - pMB1 ori - ampR	Bonneaud et al, 1991
pFL36ΔSX	Deletion site XhoI (+893 – LEU2 UTR – par filling-in) et SphI (+441 polylinker – end blunting T4 DNA polymerase)	L. Minvielle-Sébastia
pGCYC1	*CYC1* gene under GAL1 promoter - R : Ampicilline /Marqueur: TRP1	Birse et al, 1998
pASZ11-clp1-1	ADE2-ARS/CEN AmpR, clp1-1	R. Haddad
pASZ11-clp1-5	ADE2-ARS/CEN AmpR, clp1-5	R. Haddad
pASZ11-clp1-6	ADE2-ARS/CEN AmpR, clp1-6	R. Haddad
pASZ11-clp1-16	ADE2-ARS/CEN AmpR, clp1-16	R. Haddad
pASZ11-clp1-17	ADE2-ARS/CEN AmpR, clp1-17	R. Haddad
pASZ11-clp1-19	ADE2-ARS/CEN AmpR, clp1-19	R. Haddad
pHH2-ADE2-PCF11	[pCYC1] - HA/6xHis Not1-Term(PGK1)-ADE2-ARS/CEN - AmpR	L. Minvielle-Sébastia
pHH2-ADE2-pcf11Δ(287-626)	Deletion of C-terminal part of PCF11 by EcoRI	L. Minvielle-Sébastia
pFL38-PCF11 (420)	PCF11 gene cloned into the BamHI and XbaI cloning sites of pFL38 - URA3- ARS/CEN-AmpR	Amrani et al, 1997
pFL36ΔSX-PCF11	Genomic PCR fragment of PCF11(-458/+2566) cloned in pFL36ΔSX	L. Minvielle-Sébastia

TABLEAU 4.3 : Vecteurs bactéries utilisés pour l'expression des protéines recombinantes.

VECTEURS	CARACTERISTIQUES	ORIGINES
pGDV1	Produit : GST-CTD R : Ampicilline	Dichtl et al (2001)
pETM30-Hrp1-6His	Produit : 6His-GST-siteTev-Nab4p-6His (NcoI/XhoI) R : Kanamycine / Low copy plasmid	M. Mourat
pETM30-Hrp1	Produit : 6His-GST-siteTev-Nab4p (cloné NcoI/XhoI) R : Kanamycine / Low copy plasmid	F. Hakil
6His-GST		
pETM30-NAB2	Produit : GST-Tev-Nab2p-6His R : Kanamycine / Low copy plasmid	N. Viphakone
pET22-Pab1	Produit : Cter-6His Pab1-6His - Promoteur T7 R : Ampicilline Clonage BamHI/NotI	Amrani et al (1997)
pFS1853	Produit : GST-Yra1 (sans intron) R : Ampicilline	Stutz et al (2000)

C.1.2. Transformation des bactéries électro-compétentes

Les cellules compétentes (25µL par transformation) sont mises en présence de l'ADN transformant pendant 1 minute sur la glace. Après la décharge électrique (2500V - Electroporator 2510 - Eppendorf), 1mL de LB est ajouté et la suspension de cellules est incubée 1 heure, 200rpm, à 37°C avant d'être étalée sur milieu LB contenant l'antibiotique adéquat pour la sélection.

C.2. Transformation des levures

Le protocole de transformation de levures utilisé est adapté du protocole de Gietz et Wood (Gietz and Woods, 2002) « *Quick and Easy Trafo protocol* ». À partir d'une pré-culture liquide ou de cellules fraiches étalées sur boite, 50mL de YPD(A) sont inoculés - DO_{600nm} 0,3. Après 5 à 6 heures d'incubation (DO_{600nm} de 1), les cellules sont lavées 2 fois par 1mL d'H_2O qualité biomol stérile. Ces dernières sont ensuite reprises en présence de 33% PEG 3500, 100mM LiAc, 100µg d'ADN de sperme de saumon préalablement dénaturé (100°C pendant 5 minutes) et 0,1 à 1µg d'ADN transformant. Un choc thermique est alors réalisé : 42°C, 40 minutes. Les cellules sont récoltées, reprises dans 100µL final, puis étalées sur milieu sélectif. Dans le cas des souches portant un allèle ts, le choc thermique est réalisé à 30°C pendant 90 minutes.

C.3. Isolement des acides nucléiques

C.3.1. Préparation d'ADN plasmidique à partir des bactéries

L'extraction d'ADN plasmidique - à partir de bactéries cultivées en milieu liquide - peut être réalisée à l'aide d'un kit Qiagen (« *Qiagen plasmid mini kit* ») ou par la méthode de lyse alcaline (Birnboim and Doly, 1979). Brièvement, les cellules sont culotées et reprises dans 200µL d'H_2O qualité biomol, auxquels sont ajoutés 200µL d'un tampon 0,2N NaOH, 1% SDS. La lyse se poursuit 5 minutes à température ambiante, avant d'être arrêtée par l'ajout d'une solution contenant 5M d'acétate de potassium, 1,5% acide acétique. Le mélange est alors incubé 10 minutes dans la glace puis centrifugé 15 minutes à 14000 rpm, à 4°C. L'ADN, contenu dans le surnageant, est précipité par 400µL d'isopropanol et récupéré par centrifugation, 20 minutes, 14000rpm, à 4°C. Le culot d'ADN est lavé par 1mL d'éthanol, séché et repris dans 30 à 50µL d'H_2O qualité biomol ou tampon AE (10mM TrisHCl pH 9, 0,5mM EDTA).

Pour le stockage des différents plasmides construits, l'ADN plasmidique est extrait à l'aide d'un kit Qiagen – « *Qiagen plasmid midi kit* » – à partir de 50mL de culture en milieu LB liquide.

C.3.2. Préparation d'ADN plasmidique à partir de levures

L'ADN plasmidique de cellules de levure a été extrait selon le protocole décrit par Ward (Ward, 1990). Les cellules sont cultivées dans 3mL de milieu YPD(A), à 24°C, sur la nuit. Les cellules sont récoltées à partir de 1,5mL de culture, et suspendues de nouveau dans 200µL de tampon de lyse (2% Triton X-100, 1% SDS, 100mM NaCl, 10mM TrisHCl pH 8 et 1mM EDTA). 200µL de billes de verres (diamètre 0,45mm) et 200µL d'un mélange phénol/ chloroforme/alcool isoamylique (25/24/1) sont ajoutés. Le mélange est ensuite vortexé pendant 2 minutes. La phase aqueuse est séparée par centrifugation 5 minutes à 13000rpm. Une nouvelle extraction phénol/chloroforme est réalisée et les acides nucléiques sont précipités par l'ajout de 500µL d'éthanol 100%. Après lavage à l'éthanol 70%, le culot est séché et repris dans 50µL d'H_2O qualité biomol ou de tampon AE.

C.3.3. Préparation d'ARNs totaux de levures

Les ARNs totaux de levure ont été extraits selon le protocole d'extraction rapide décrit par Schmitt *et al.,* (Schmitt et al., 1990). Les cellules sont cultivées dans 50mL de milieu YPD(A) à 24°C jusqu'à une DO_{600nm} de 3. Les cellules sont ensuite incubées à 37°C pendant 0, 30, 60 et 120 minutes et l'extraction est réalisée sur 10mL de culture. Les cellules sont récupérées par centrifugation, 4000rpm pendant 5 minutes, puis lavées par 1mL d'H_2O qualité biomol froide, conservée dans la glace.

Les cellules sont reprises dans 500µL de tampon AE froid (50mM acétate de sodium pH 5,3, 10mM EDTA) auxquels sont ajoutés 50µL de SDS 10%. Le mélange est ensuite transféré dans 550µL de phénol acide (phénol/chloroforme/isoamylalcool 25 :24 :1), préalablement équilibré avec du tampon AE et incubé à 65°C. Le mélange est vortexé pendant 30 secondes, puis incubé 30 minutes à 65°C avant d'être refroidi dans un bain de carboglace mélangé à de l'éthanol, jusqu'à apparition de cristaux. La phase aqueuse est séparée par centrifugation et traitée par 550µL phénol acide à température ambiante, puis par 550µL de chloroforme/isoamylalcool (24 :1). Les ARNs contenus dans la phase aqueuse sont précipités en ajoutant 1/10 de volume de 3M d'acétate de sodium pH 5,3 et 3 volumes d'éthanol 100% puis incubés sur la nuit à -80°C. Les échantillons sont centrifugés à 13000rpm, pendant 30

minutes à 4°C. Les culots obtenus sont lavés par 1mL d'éthanol 70% froid et séchés à température ambiante. Les ARNs totaux sont repris dans 200µL d'H_2O qualité biomol, incubés 10 minutes sur la glace, 5 minutes à 65°C puis conservés à -80°C.

C.4. Dosage des acides nucléiques

La quantité d'acides nucléiques est dosée au NanoDrop (ThermoScientific ND-1000) et leur qualité est vérifiée par électrophorèse sur gel d'agarose. La lecture de l'absorbance à 260nm permet de déterminer la concentration en acides nucléiques. Une unité de densité optique correspond à 50µg/mL d'ADN double brin et à 40µg/mL d'ARNs. Le rapport entre les mesures d'absorbance à 260 et 280nm permet d'avoir une estimation du degré de pureté des échantillons. Cette valeur A260nm/A280nm doit être proche de 1,8 pour les préparations d'ADN et proche de 2 pour les ARNs.

C.5. Digestion des plasmides

Pour les étapes de clonage, de vérification ou pour générer un « gap », la digestion d'ADN est réalisée en incubant 1µg du plasmide d'intérêt avec 5 à 10 Unités d'enzyme de restriction correspondante (Biolabs ou Fermentas), pendant 1 à 16 heures, à la température optimale, généralement 37°C. Les doubles digestions sont réalisées dans des tampons compatibles pour les deux enzymes.

C.6. Amplification par PCR

Pour un volume final de 50µL, 1 Unité de la polymérase thermostable Q5 (Biolabs) dans son tampon est ajouté à 1µL de dNTP (10mM), 20ng de matrice, et 3µL de chaque amorce (10µM). Une phase initiale de dénaturation (95°C), hybridation (55°C) et élongation (72°C) est réalisée pour 35 cycles de PCR. Les produits de PCR sont analysés sur gel agarose 1% / TBE 0,5% puis directement purifiés du mélange réactionnel par le kit « *Gene Jet PCR Purification* » (Thermo Scientific).

C.7. Caractérisation du phénotype

C.7.1. Test en goutte / thermosensibilité

La croissance et la thermosensibilité des mutants obtenus sont étudiées sur milieu YPD(A). Une anse de cellules est prélevée à partir de cultures sur YPD(A) et est reprise dans 200µL d'H_2O qualité biomol stérile. Après dilutions sérielles dans 50µL d'H_2O qualité biomol stérile, 5µL de chaque dilution sont déposés sur milieu YPD(A). La thermosensibilité est étudiée en incubant les différentes souches à 24°C, 30°C et 37°C pendant 24 à 48 heures.

La sensibilité au formamide à 30°C est étudiée en incubant les différentes souches sur un milieu YPD(A) contenant 1,5% ou 2% de formamide, pendant 24 à 48 heures.

C.7.2. Test de croissance

Les phénotypes de croissance lente sont également étudiés en milieu liquide, à 30°C ; afin de pouvoir calculer le temps de génération. 200mL de milieu YPD(A) sont inoculés à partir d'une pré-culture sur la nuit pour obtenir une DO_{600nm} de 0,05. Les cultures sont incubées à 30°C, 160rpm pendant 24 à 48 heures et la DO_{600nm} est évaluée toutes les heures à partir de 1mL de culture. Après 24 à 48 heures, l'équivalent de 500µL de chaque culture est lavée dans 1mL d'H_2O qualité biomol stérile, puis étalés sur milieu YPD(A) afin de s'assurer de la non présence de révertants.

D. TECHNIQUES DE GENETIQUES

D.1. Production de diploïdes

La méthode utilisée est adaptée du « *Matchmaker Gold Yeast two-hybrid system user manual* ». Une anse de chaque haploïde MATa et MATα est incubée dans 500µL de YPD(A), sur la nuit à 30°C, 250rpm. Les cellules sont récupérées par centrifugation (3 minutes, 3000rpm) puis lavées 2 fois par 1mL d'H_2O qualité biomol stérile et reprises dans 100µL final. Elles sont ensuite étalées sur milieu sélectif permettant la multiplication des diploïdes uniquement.

Les diploïdes peuvent également être obtenus par croisement en « patch ». Brièvement, une anse de chaque haploïde est étalée sur milieu YPD(A) puis elles sont mélangées. Après incubation 6 à 7 heures à 30°C (24°C pendant 16 heures dans le cas de

TABLEAU 4.4 : Oligonucléotides utilisés lors des clonages et séquençages

OLIGO	SENS	SEQUENCES	UTILISATION
olm444	+	GTTTTTAATAAAGGCCAGTGCTcagcagcaacagcaacaacagcag	pcf11Δ(141-233)
olm445	-	ctgctgttgttgctgttgctgctgAGCACTGGCCTTTATTAAAAAC	pcf11Δ(141-233)
olm449	+	ataagaatgcggccgccATGGATCACGACACAGAAGTTATAG	PCF11 dans pHH2
olm450	-	ataagaatgcggccgcTTATTTTGTGACCAATTTCTTTAAGTCATC	PCF11 dans pHH2
olm451	+	GAAAGAATGAGGTATCACGAACTTtatcacgaaactaaggatatggt	pcf11Δ(Q20)
olm452	-	accatatccttagtttcgtgataAAGTTCGTGATACCTCATTCTTTC	pcf11Δ(Q20)
olm455	+	GAAAAATTAAAAATGAAACTGTTGGTTgacTCTCAGTTGAA ACAAGAACTGAAAAG	pcf11(L189D)
olm456	-	CTTTTCAGTTCTTGTTTCAACTGAGAgtcAACCAACAGTTTCA TTTTTAATTTTTC	pcf11(L189D)
olm457	+	GCTCTGAAACAAGTTCAAATGCAAgacAGGCAGGTCTTCTC ACAGGACCAAC	pcf11(L213D)
olm458	-	GTTGGTCCTGTGAGAAGACCTGCCTgtcTTGCATTTGAACTT GTTTCAGAGC	pcf11(L213D)
olm459	+	CAACTAAGGCAGGTCTTCTCACAGgccCAACAAGTTTTGCA GGAAAGAATG	pcf11(D220A)
olm460	-	CATTCTTTCCTGCAAAACTTGTTGggcCTGTGAGAAGACCTG CCTTAGTTG	pcf11(D220A)
olm461	+	CAGTTTTTAATAAAGGCCAGTGCTtatcacgaaactaaggatatggtcg	pcf11Δ(141-253)
olm462	-	cgaccatatccttagtttcgtgataAGCACTGGCCTTTATTAAAAACTG	pcf11Δ(141-253)
olm463	+	GCAAGCAATGTTACCAACTCCGACTaatcaaccaaatgatgaaaaattaaa aatg	pcf11Δ(155-174)
olm464	-	catttttaatttttcatcatttggttgattAGTCGGAGTTGGTAACATTGCTTGC	pcf11Δ(155-174)
olm465	+	GGCTAAAGAATCAACCAAATGATaaaagagaaaaactaactctcaatgc	pcf11Δ(180-196)
olm466	-	gcattgagagttagtttttctctttttATCATTTGGTTGATTCTTTAGCC	pcf11Δ(180-196)
olm467	+	GAACTGAAAAGAGAAAAACTAACTatgaggtatcacgaacttcagca	pcf11Δ(203-227)
olm468	-	tgctgaagttcgtgatacctcatAGTTAGTTTTTCTCTTTTCAGTTC	pcf11Δ(203-227)
olm529	+	GCTATACGCTTTTTATGCCTTGgcagccgctTGTAAGAATGTTG GCAGC	*pcf11-13*
olm530	-	GCTGCCAACATTCTTACAagcggctgcCAAGGCATAAAAAGCG TATAGC	*pcf11-13*
PCF11_R102	+	GGACAACACCACAgcaACAAAACTCATCAATATG	pcf11(R102A)
PCF11_R102	-	GAGTTTTGTtgcTGTGGTGTTGTCCACCAATAAG	pcf11(R102A)
PCF11_E199	+	CTGAAAAGAgcaAAACTAACTCTCAATGCTCTG	pcf11(E199A)
PCF11_E199	-	GTTAGTTTtgcTCTTTTCAGTTCTTGTTTC	pcf11(E199A)
olm535	+	cggaattctcaGAGCACTGGCCTTTATTAAAAACTGTTC	PCF11-CID (1-142) dans pHH2
olm536	+	cggaattctcaACCGACCATATCCTTAGTTTCGTGATA	PCF11-CID (1-262) dans pHH2
olm518	+	ATGCCTTCCAGATTCACTAAG	Northern blot CYH2
olm519	-	TTAAGCGATCAATTCAACAACAC	Northern blot CYH2
olm520	+	ATGTCTCAAAGAATAATTCAAC	Northern blot TRS31
olm521	-	TCAGCCAAATCTCAAACC	Northern blot TRS31
olm522	+	CAGACAAATCACTCCA	Northern blot 18S
olm484	+	ataagaatgcggccgcCAGATCCGCCAGGCGT	S1 Nucléase CYC1
olm531	-	cgggatccGTTTCAGACCCATCATCTTAC	S1 Nucléase CYC1
olm532	+	cggaattcGGATTTTTTAAAGGGTCAGGAAG	S1 Nucléase snR13
olm533	-	cgggatccGCTTGCTTAGGCCCAACAG	S1 Nucléase snR13
olm538	+	cggaattcAGGATAAGAAAACCATCATTCA	S1 Nucléase snR5
olm539	-	cgggatccGGATGGTATTCAAAACATTTC	S1 Nucléase snR5

mutants portant un allèle thermosensible), une anse de cellules est étalée sur YPD(A) et les diploïdes sont isolés à l'aide d'un micromanipulateur Singer.

D.2. Dissection d'asques

Les souches diploïdes sont incubées sur milieu de sporulation à 24°C. Après formation des asques (3-4 jours), une anse de cellules est prélevée et incubée en présence de 1000 unités de β-Glucoronidase (Sigma) pendant 40 minutes à 30°C. Les asques ainsi traités sont ensuite disséqués sur YPD(A) à l'aide d'un micromanipulateur Singer.

Après croissance, le contenu génétique de chaque spore est analysé sur milieu sélectif après empreinte sur velours.

D.3. Stratégie de « gap repaired »

La construction des différents allèles mutés dans le vecteur navette pFL36ΔSX a lieu directement dans la levure, grâce à sa haute capacité de recombinaison homologue. La souche YLM221 est transformée par 0,1 à 1µg de plasmide pFL36ΔSX préalablement ouvert par la double digestion *Xho*I/*Sph*I, et en présence des couples de fragments obtenus par PCR (*TABLEAU 4.4*). L'ADN plasmidique total des transformants obtenus est purifié selon le protocole décrit par Ward (Ward, 1990) (paragraphe C.3.2.). Le plasmide pFL36ΔSX est isolé du reste du contenu plasmidique en transformant 25µL de bactéries électro-compétentes (DH5α) par l'ADN plasmidique total. Le plasmide pFL36ΔSX portant un gène de résistance à l'ampicilline, seul les transformants possédant ce dernier peuvent se développer sur milieu de sélection LB + Amp. Les plasmides sont extraits par lyse alcaline ; la présence de l'insert est vérifiée par la double digestion *Xho*I/*Sph*I et leur séquence est vérifiée par séquençage.

D.4. Plasmide shuffling

L'impact des différents allèles mutés de *PCF11* est étudié par « plasmide shuffling ». Dans la souche YLM233, la version chromosomique de *PCF11* est remplacé par le marqueur *HIS3* ; cette délétion est secourue par la présence du plasmide pFL38-*PCF11* portant le marqueur *URA3*. Cette souche est transformée par les vecteurs pFL36ΔSX obtenus par « *gap repair* ». Après sélection, les transformants obtenus sont étalés sur milieu SC contenant 0,1% de 5-FOA. La présence du gène *URA3*, codant pour l'oritidine 5'-phosphate décarboxylase, contenu sur la plasmide pFL38-*PCF11* conduit à la conversion du 5-FOA en 5-fluoro-uracile, composé toxique pour la cellules qui bloque la synthèse des acides nucléiques. Cela permet la

chasse du plasmide pFL38-*PCF11* ; les souches obtenues seront ura- et ne porteront qu'une seule copie de *PCF11*, présente sur le plasmide pFL36ΔSX. Le phénotype induit par des différents allèles est ensuite caractérisé.

E. TECHNIQUES DE BIOCHIMIE DES PROTEINES

E.1. Production de protéines recombinantes

E.1.1. Préparation d'extraits protéiques

Des bactéries électro-compétentes BL21 DE3 ou BL21 pLysS sont transformées avec les plasmides codant pour les protéines d'intérêts présentées dans le *TABLEAU 4.3* puis étalées sur milieu sélectif.

500mL de milieu LB sont ensemencés avec une pré-culture obtenue sur la nuit, puis incubés à 37°C, 180rpm jusqu'à obtenir une DO_{600nm} de 1 – 1,2. Pour une production à plus grande échelle, les bactéries sont cultivées dans du milieu TB. Les cultures sont alors incubées à 25°C, 180rpm et une fois la température redescendue, l'expression de la protéine d'intérêt est induite sur la nuit par 0.25mM d'IPTG. Le lendemain, les cellules sont récupérées par centrifugation et reprises dans 40mL de tampon de lyse (PBS 1X, 5mM DTT, 5mg/mL DNAse I, 1mM PMSF, 2µg/mL Leupeptine, 2µg/mL Pepstatine, 5mM Benzamidine). La lyse est effectuée par sonication (3 cycles de 1min puissance 3 continue séparés par des cycles de 1min de refroidissement dans la glace). Le lysat obtenu est incubé 30 minutes, sous agitation à 4°C en présence de 1% de TX-100. Les débris cellulaires sont éliminés par centrifugation différentielle (10500rpm, 4°C, 15 minutes). Le surnageant ainsi obtenu est stocké à -80°C en présence de 10% de glycérol jusqu'à utilisation.

E.1.2. Purification GSH-Sépharose

L'extrait protéique (volume final de 40mL) est incubé en présence de 1,5mL de billes GSH-sépharose (GE Healthcare) - lavées par 2 fois 40mL PBS 1X, 5mM DTT - pendant 3 heures, sous agitation à 4°C. Les billes sont récupérées sur colonne Biorad et lavées par 2 fois 10mL de 1M NaCl puis par 10mL de 150mM NaCl. L'élution est effectuée par incubation des billes en présence de 6 fois 1,5mL de tampon d'élution (125mM TrisHCl pH7,9, 150mM NaCl, 10mM Glutathion réduit ou GSH, 10% glycérol).

Après purification sur GSH-Sépharose, la fraction d'élution contenant **GST-Hrp1** est dialysée contre 2 x 2L de tampon E_{20}^{50} (20mM TrisHCl pH 7,9, 50mM KCl, 0,2% NP-40, 0,5mM DTT, 1mM EDTA, 20% glycérol) puis purifiée par chromatographie à échange d'ions (UnoQ – BioRad) avec un gradient 50-500mM KCl.

Les autres fractions d'élutions obtenues sont dialysées contre 2 fois 2 litres de tampon d'interaction (10mM TrisHCl pH 7,9, 150mM NaCl, 0,1% NP-40, 0,5mM DTT, 0,1mM PMSF, 0,5mM EDTA, 10% glycérol).

E.1.3. Purification Ni-NTA Agarose

L'extrait protéique (volume final de 40mL) est incubé en présence de 1mL de billes NiNTA-agarose (Qiagen) lavées par 2 fois 40mL PBS 1X pendant 2 heures, sous agitation à 4°C, en présence de 10mM d'imidazole. Les billes sont récupérées sur colonne Biorad et lavées par 3 fois 10mL de tampon de lavage (25mM TrisHCl pH 7,9, 1M KCl, 0,01% NP-40, 15mM imidazole, 10% glycérol). L'élution est effectuée par incubation des billes en présence de 6 fois 1mL de tampon d'élution (25mM TrisHCl pH 7,9, 100mM KCl, 0,01% NP-40, 125mM imidazole, 10% glycérol).

Dans le cas de la construction **Pab1-6His**, une seconde purification sur NiNTA-agarose est effectuée après dialyse contre 2 fois 2 litres de tampon d'interaction.

Dans le cas de la construction **GST-TEV-Pab1-6His**, une première purification sur NiNTA-agarose est effectuée. Après vérification de cette première étape, les fractions d'élutions sont incubées 3 heures, sous agitation, à 4°C en présence de 1,5mL de billes GSH-sépharose - lavées par 2 fois 40mL PBS 1X, 5mM DTT. Après fixation, les billes sont lavées par 2 fois 10mL de NaCl 1M puis par 10mL du tampon de clivage (10mM TrisHCl pH 7,9, 150mM NaCl, 0,1% NP-40, 0,5mM EDTA, 1mM DTT). La construction Pab1-6His est récupérée par action de 15 unités de TEV homemade, 16 heures, sous agitation, à 4°C dans 3mL de tampon de clivage. La protéine d'intérêt se trouve alors dans la fraction non retenue.

Les fractions d'élutions obtenues sont ensuite dialysées contre 2 fois 2 litres de tampon d'interaction (10mM TrisHCl pH 7,9, 150mM NaCl, 0,1% NP-40, 0,5mM DTT, 0,1mM PMSF, 0,5mM EDTA, 10% glycérol).

E.2. Préparation d'extraits protéiques de levures

E.2.1. Extraction rapide

Cette méthode est adaptée du protocole de Riezman *et al.,* (Riezman et al., 1983). L'extraction est réalisée sur 2mL de cellules à une DO_{600nm} de 2. Les cellules sont récupérées par centrifugation et reprise dans 250µL de 50mM TrisHCl pH7,9. La suspension cellulaire est ensuite incubée 10 minutes à 4°C en présence de 50µL de mix (2N NaOH, 13% β-MercaptoEthanol, 20mM PMSF). Les protéines sont précipitées, 20 minutes à 4°C par l'ajout de 220µL de TCA 72%, puis récupérés par centrifugation (13000rpm, 30 minutes, 4°C). Le culot protéique est lavé par 1mL d'acétone et séché, avant d'être repris dans 100µL de Laemmli 1X (10µL correspondant à une DO_{600nm} de 0,15).

E.2.2. Préparation d'extraits protéiques dits « 0-40 »

La méthode utilisée est celle décrite par Minvielle-Sebastia *et al.,* (Minvielle-Sebastia et al., 1994) et permet d'obtenir des extraits protéiques enrichis, utilisables lors des tests d'activité. Les cellules sont incubées sur la nuit dans 2 litres de YPD(A) jusqu'à une DO_{600nm} de 3-4. Elles sont ensuite récupérées par centrifugation et le culot est repris dans 80mL de tampon S (50mM Tris-HCl, pH 7,9, 10mM $MgCl_2$ et 1M Sorbitol) et incubé 20 minutes à température ambiante en présence de 30mM de DTT. Les cellules sont récupérées par centrifugation et reprises dans 40mL de tampon S contenant 3mM de DTT. La paroi des cellules est ensuite fragilisée en incubant la suspension cellulaire en présence de 10mg de Zymolyase-20T (Euromedex) pendant 30 à 45 minutes à 30°C. La formation des sphéroplastes est suivie sous microscope ; l'incubation est arrêtée lorsque plus de 90% des cellules sont fragilisées.

Les étapes suivantes sont alors réalisées à 4°C ou sur la glace avec des tampons préalablement refroidis. Les sphéroplastes sont récupérés par centrifugation et lavés par 40mL de tampon S, puis repris dans 40mL de tampon lyse (10mM TrisHCl, pH 7,9, 10mM KAc, 0,5mM DTT) contenant des inhibiteurs de protéases – 0,5mM Phenylmethylsulfonyl fluoride (PMSF), 0,4µg/µL Leupeptine, 0,7µg/µL Pepstatine et 2mM Benzamidine. La lyse totale des cellules est obtenue après une vingtaine de chocs portés à la suspension cellulaire par un système de broyeur Dounce. La concentration en KAc du lysat cellulaire est ensuite ajustée à 200mM et équilibrée 30 minutes sur la glace. L'extrait est soumis à une ultracentrifugation à 20500rpm, pendant 30 minutes, à 4°C (ultracentrifugeuse Beckman Coulter, rotor 50,2 Ti, soit 37000g). Le surnageant est soumis à une seconde ultracentrifugation de 41000rpm,

1 heure, 4°C (soit 150000g). Il est ensuite incubé, sur la nuit, en présence de sulfate d'ammonium (0,226g/mL d'extrait obtenu) afin d'obtenir une saturation 0-40% à 0°C.

Le culot obtenu après centrifugation (15000rpm, 30 minutes, 4°C), est repris dans 1 à 2mL de tampon de dialyse (20mM TrisHCl, pH 7,9, 1mM EDTA, 50mM KAc, 0,5mM DTT, 20% glycérol) contenant des inhibiteurs de protéases (de concentrations finales identiques à celles utilisées pour l'étape de lyse des sphéroplastes). L'extrait obtenu est ensuite dialysé contre 2 litres de tampon de dialyse en présence d'inhibiteurs de protéases (2 fois 2 heures). Le surnageant obtenu après centrifugation est congelé, sous forme d'aliquotes de 50µL, dans l'azote liquide et conservés à -80°C.

E.3. Purification des facteurs protéiques par *TAP-Tagging*

Cette méthode, développée par Puig *et al.,* (Puig et al., 2001) a été utilisée pour purifier les facteurs protéiques associés aux protéines étiquetées en N-terminale - Rna15 et Fip1 en fonction des constructions. Le protocole de préparation des extraits protéiques est identique au protocole de préparation des extraits dits « 0-40 », jusqu'à l'étape de précipitation à l'ammonium sulfate.

La concentration de l'extrait protéique, après centrifugation à haute vitesse, est équilibrée à une concentration finale de 0,1% NP-40 et à 150mM NaCl. L'extrait protéique est ensuite incubé 2 heures, à 4°C sous agitation, en présence de 400µL de suspension de billes IgG-Sépharose (GE Healthcare). Le mélange est passé sur colonne Biorad ; ce qui n'est pas retenu sur la colonne n'est pas conservé. La colonne est ensuite lavée par 3 x 10mL de tampon IPP150 (10mM TrisHCl pH 7,9, 150mM NaCl, 0,1% NP-40) puis par 10mL d'IPP150 contenant 0,5mM EDTA et 1mM DTT. Les billes sont incubées 16 heures, à 4°C, sous agitation, en présence de 2mL de ce tampon et 10µg de TEV protéase homemade (100 unités TEV commercial). L'éluât est recueilli ainsi que le volume mort, après passage de 500µL de tampon sur la colonne ; cette fraction correspond au « ***TEV Eluate*** ». À cet éluât sont ajoutés 6,6µL de $CaCl_2$ 1M et trois volumes (6,6mL environ) de tampon de fixation sur la calmoduline (IPP150 + 1 mM MgAc, 1 mM Imidazole, 2 mM $CaCl_2$, 10 mM β-MercaptoEthanol). Le mélange est incubé en présence de 400µL de billes de Calmoduline-sépharose (GE Healthcare) pendant 1 heure, à 4°C, sous agitation. Le volume non retenu, le « *Calmodulin Flow-Through* » est stocké à -80°C en présence de 10% de glycérol. Les billes sont lavées par 3 x 10mL de tampon de fixation sur la calmoduline. Les facteurs sont élués, sous forme de 5 fractions par 400µL de tampon d'élution (IPP150 + 1 mM MgAc, 1 mM

TABLEAU 4.5 : Anticorps utilisés dans cette étude.

ANTICORPS		DILUTIONS	ORIGINE
PRIMAIRE	anti-Rna15	1/10000^{e}	polyclonal
	anti-Rna14	1/3000^{e}	polyclonal
	anti-Clp1	1/1000^{e}	polyclonal
	anti-Pcf11	1/2000^{e}	polyclonal
	anti-GST	1/5000^{e}	monoclonal
	anti-Fip1	1/5000^{e}	polyclonal
	anti-Pab1	1/10000^{e}	polyclonal
SECONDAIRE	anti-souris	1/6000^{e}	monoclonal
	anti-lapin	1/10000^{e}	polyclonal

Imidazole, 2 mM EGTA, 10 mM β-MercaptoEthanol, 10% glycérol). Les fractions sont congelées dans l'azote liquide et conservées à -80°C. 10µL de chaque fraction sont déposés sur gel dénaturant SDS-PAGE 12% puis colorés au nitrate d'argent afin de vérifier la qualité de la purification et d'estimer la quantité de complexe purifié.

Pour la purification des facteurs CPF – l'étiquette est située sur la sous unité Fip1p – trois lavages de 10mL avec une salinité de 300mM NaCl puis 1 lavage de 10mL à 150mM NaCl sont effectués après l'étape de fixation sur les billes de calmoduline. Cela permet d'éliminer les facteurs CF IA et/ou CF IB pouvant co-purifier lors de préparation (Kessler et al., 1996).

E.4. Western Blot

Les échantillons sont déposés sur un gel d'acrylamide dénaturant SDS-PAGE puis transférés sur membrane de nitrocellulose 0,45µM grâce à un système semi-sec (Poly Labo - 3mA/cm^2, 45 minutes) ou un système Trans Blot Turbo (Biorad - 1,3A, 25V, 7 minutes) dans du tampon de transfert (25 mM Tris-HCl, 192 mM glycine, 0,1% SDS, 20% éthanol). Après coloration au Rouge Ponceau S, la membrane est saturée en présence de PBS 1X, 0,1% Tween20 et 5% lait (w/v) pendant une heure puis incubée en présence de l'anticorps primaire selon la dilution indiquée (dans du PBS 1X, 0,1% Tween20 et 5% lait w/v) pendant une heure à température ambiante. Après trois lavages PBS 1X, 0,1% Tween20, la membrane est incubée avec l'anticorps secondaire dilué dans le même tampon que l'anticorps primaire, pendant 45 minutes. La membrane est à nouveau lavée 3 fois puis incubée en présence du substrat de la peroxydase (« *SuperSignal West Pico Chemiluminescent Substrate* » - Thermo Scientific). Le signal est détecté par autoradiographie.

Les anticorps et les dilutions utilisées sont résumés dans le *TABLEAU 4.5*.

E.5. Test d'activité kinase

E.5.1. Activité associée aux facteurs

L'activité kinase retrouvée associée aux facteurs CF IA et CPF est testée à partir de 4,4µL de facteurs purifiés. Les facteurs sont incubés 30 minutes à 30°C en présence de 10mM HEPES pH8, 2mM MgAc, 1mM DTT, 0,05% NP-40 et 0,2µL de [γ^{32}P]-ATP (25µCi/6000Ci/mmole) (*TABLEAU 4.6*). La réaction est arrêtée par ajout de 2µL de tampon Laemmli 5X. Les échantillons sont incubés à 100°C pendant 2 minutes puis les protéines sont

TABLEAU 4.6 : Tests d'activité kinase associée aux facteurs.

MIX 10X	**VOLUME FINAL (100 µL)**	**CONCENTRATION FINALE**
HEPES pH8 (1M)	10 µL	100mM
MgAc (1M)	2 µL	20mM
DTT (1M)	1 µL	10mM
NP-40 (10%)	5 µL	0,5%
H_2O	82 µL	-

MIX ESSAI KINASE	**VOLUME FINAL**	
Facteurs	4,4 µL	
Mix10X	1 µL	5,6 µL
[γ^{32}P]-ATP	0,2 µL	
H_2O	4,4 µL	

séparées par SDS-PAGE en conditions dénaturantes. Après fixation pendant 45 minutes (40% Éthanol, 10% Acide acétique) et séchage du gel, le transfert d'un phosphate radioactif est observé par autoradiographie (Molecular Imager Pharos FX - Biorad).

E.5.2. Phosphorylation CTD par la MAP Kinase

La phosphorylation du CTD par la MAP Kinase est réalisée selon le protocole décrit par Schreieck *et al.,* en 2014 (Schreieck et al., 2014). Pour un volume final de 25µL, 400ng de protéines GST-CTD sont incubées pendant 1 heure à 30°C, en présence de 80 Unités d'enzyme MAP Kinase (New England BioLabs) et 1X NEB Buffer pour protéines kinases.

Dans le cas d'un marquage radioactif, 200µM d'ATP et 500µCi/µmol de [γ^{32}P]-ATP (25µCi/6000Ci/mmole) sont ajoutés au mix réactionnel. La réaction est stoppée par l'ajout de 2µL de Laemmli 5X, et les protéines sont séparées par SDS-PAGE en conditions dénaturantes. Après fixation (40% Éthanol, 10% Acide acétique) et séchage du gel, le transfert d'un phosphate radioactif est observé par autoradiographie.

Dans le cas d'une phosphorylation « à froid », 400µM d'ATP sont ajoutés au mix réactionnel. La réaction est stoppée par l'ajout de 10mM d'EDTA.

E.6. Test d'activité d'hydrolyse de l'ATP

Le protocole utilisé est adapté du protocole décrit pas Hughes *et al.,* (Hughes and Romanos, 1993). La présence d'une activité ATPasique associée aux facteurs CF IA et CPF est analysée de la même manière que l'activité kinase, à partir de 4,4µL de facteurs purifiés. Les facteurs sont incubés 30 minutes à 30°C en présence de 10mM HEPES pH8, 2mM MgAc, 1mM DTT, 0,05% NP-40, 50µM ATP non radioactif et 0,25µL de [γ^{32}P]-ATP (25µCi/6000Ci/mmole). La réaction a lieu 30 minutes à 37°C dans le cas de la phosphatase alcaline, utilisée comme témoin. La réaction est arrêtée par l'ajout de 0,2M EDTA. 1µL de chaque est réaction est déposée sur une plaque de CCM et la migration s'effectue pendant 45 à 60 minutes en présence de 1M d'acide acétique et 0,5M acétate de lithium. Les résultats sont ensuite observés par autoradiographie (Molecular Imager Pharos FX - Biorad).

E.7. Test d'interaction CF IA

E.7.1. GST-CTD

Après phosphorylation par la MAP Kinase de 400ng de GST-CTD, 30µL de billes GST-Sépharose sont ajoutés au mélange réactionnel. La fixation a lieu pendant 1 heure à 4°C sous faible agitation. Les billes sont récupérées par centrifugation et lavées par 3 x 300µL de tampon d'interaction (paragraphe E.1.3.) contenant 500mM NaCl. L'interaction avec le complexe CF IA est analysée en ajoutant 150µL de *TEV Eluate*, sur la nuit à 4°C sous faible agitation. Les billes sont lavées par 2 x 300µL de tampon d'interaction contenant 500mM NaCl, 2 x 300µL 1M NaCl puis 1 x 300µL 200mM NaCl. L'analyse se fait directement par Western Blot en chargeant 1/3 des billes reprises dans 10µL de tampon de charge 5X et incubées à 90°C pendant 5 minutes. Les anticorps utilisés sont dirigés contre le tag GST, Pcf11 et Rna15.

E.7.2. GST-Yra1

Lors des expérience préliminaires, 400ng de GST-Yra1 sont incubée en présence de 30µL de billes GST-Sépharose. La fixation a lieu pendant 1 heure à 4°C sous faible agitation. Les billes sont récupérées par centrifugation et lavées par 3 x 300µL de tampon d'interaction (paragraphe E.1.3.) contenant 500mM NaCl. L'interaction avec le complexe CF IA est analysée en ajoutant 150µL de *TEV Eluate*, sur la nuit à 4°C sous faible agitation. Les billes sont lavées par 2 x 300µL de tampon d'interaction contenant 500mM NaCl, 2 x 300µL 1M NaCl puis 1 x 300µL 200mM NaCl. L'analyse se fait directement par Western Blot en chargeant 1/3 des billes reprises dans 10µL de tampon de charge 5X et incubées à 90°C pendant 5 minutes. Les anticorps utilisés sont dirigés contre le tag GST, Pcf11 et Rna15.

F. TECHNIQUES DE BIOCHIMIE DES ARNS

F.1. Plasmides

Les caractéristiques des vecteurs utilisés sont répertoriées dans le *TABLEAU 4.7*

TABLEAU 4.7 : Plasmides et séquences des ARN synthétiques CYC1 et CYC1-pré. Les éléments cis impliqués dans la maturation en 3' sont représentés (« **l'*efficiency element*** » riche en AU ; « **le *positioning element*** » riche en A ; le **site de clivage** ; les éléments riche en U **UUE** – *Upstream U-rich Element* - et **DUE** – *Downstream U-rich Element*).

VECTEURS	TRANSCRIT		ORIGINES
pG4-CYC1	CYC1	gggagaccggaagcuugcaugccugcagguCGAUAUCAUGUAAUUAGUU AUGUCACGCUUACAUUCACGCCCUCCCCCCACAUCCGCUCUAAC CGAAAAGGAAGGAGUUAGACAACCUGAAGUCUAGGUCCCUAUU UAUUUUUUUAUAGUUAUGUUAGUAUUAAGAACGUUAUUUA UAUUUCAAAUUUUUUCUUUUUUUUCUGUACAGACGCGUGUACG CAUGUAACAUUAUACUGAAAACCUUGCUUGAGAAGGUUUUGGG ACGCUCGAcucuagaggaucccccgggauccgagcucgaauu	Minvielle-Sébastia *et al.*, 1994
pG4-CYC1pré	CYC1-pré	gggagaccggaagcuugcaugccugcaggucgacucuagaCGAUAUCAUGU AAUUAGUUAUGUCACGCUUACAUUCACGCCCUCCCCCCACAUCC GCUCUAACCGAAAAGGAAGGAGUUAGACAACCUGAAGUCUAGG UCCCUAUUUUAUUUUUUUAUAGUUAUGUUAGUAUUAAGAACG UUAUUUAUAUUUCAua	Preker *et al.*, 1995

TABLEAU 4.8 : Composition des mix réactionnel utilisés lors de la reconstitution des activités de maturation de l'extrémité 3'

COMPOSITION MILIEU REACTIONNEL	VOLUME FINAL = 20µL
Polyéthylène glycol PEG 8000 (8 %)	5 µL
EDTA ou MgAc (100 mM)	0,4 µL
KAc (1 M)	1,5 µL
ATP (100 mM)	0,4 µL
Créatine phosphate (1 M)	0,4 µL
Créatine kinase (10 mg/mL)	0,2 µL
Nondiet P-40 0,5%	0,4 µL
Dithiotréitol 0,1M	0,2 µL
BSA (10 mg/mL)	0,2 µL
RNAguard (34 U/mL)	0,1 µL
Extrait protéique	2-4 µL
Substrat ARN marqué (20-40 fmoles de CYC1 ou CYC1-pré) repris dans l'eau	Qsp 20 µL

COMPOSITION TAMPON PROTEINASE K	VOLUMES
Tampon 2X (200mM Tris-HCl pH 7.9, 300mM NaCl, 25mM EDTA, 2% SDS)	50 µL
Protéinase K (20mg/mL)	1 µL
Glycogène (20mg/mL)	0,25 µL
Eau	23,75 µL

F.2. Production d'ARNs *in vitro*

F.2.1. Transcription d'ARNs *in vitro*

Les ARNs non radio-marqués utilisés comme contrôles sont produits par trasncription T7 grâce au kit MEGAscript (Ambion). Brièvement, 1µg de matrice ADN (plasmide préalablement linéarisé en aval de l'insert) est incubée - 4 heures pour des transcrits supérieur à 500pb ou sur la nuit pour des transcrits inférieur à 500pb - en présence des différents composants nécessaire à la réaction. L'excédent de matrice est éliminé par incubation 15 minutes à 37°C en présence de 2 unités de TURBO DNAse. La réaction est ensuite stoppée par l'ajout de 115µL d'H_2O et 15µL d'ammonium acétate. Les transcrits sont purifiés par une extraction phénol/chloroforme et précipités 3 heures en présence de 150µL d'isopropanol. Les ARNs synthétiques sont ensuite séparés sur gel 6% polyacrylamide (acrylamide/bisacrylamide 19 :1)/TBE 1X/8,3M urée et repérés sur le gel par UV Shadow. La bande correspondante est découpée et l'ARN est élué comme décrit dans le paragraphe F.2.2.1. Après centrifugation 30 minutes, 13000rpm, 4°C, le culot d'ARNs est lavé par 1mL d'éthanol 70%, séché puis repris dans 50µL d'H_2O.

Dans le cas de l'ARN correspondant à snR5 – utilisé comme témoin dans les essais de protection à la digestion par la S1 nucléase – la transcription est réalisée par transcription SP6. Le protocole est identique à celui utilisé lors de la transcription d'ARN radiomarqué. Les NTPs sont ajoutés au mix réactionnel pour une concentration finale de 0,5mM. La réaction est effectuée pendant 2 heures à 40°C, en présence de 25 unités d'ARN polymérase SP6. Les ARN sont ensuite précipités et purifiés comme décrit dans le paragraphe F.2.2.1. L'ARN est repéré sur le gel par UV Shadow. La bande correspondante est découpée et l'ARN est élué comme décrit dans le paragraphe F.2.2.1.

F.2.2. Transcription *in vitro* d'ARNs radiomarqués

F.2.2.1. Transcription in vitro d'ARNs radiomarqués

La maturation en 3' des ARNs messagers est étudiée par reconstitution de cette activité *in vitro*. Un ARN synthétique est fourni comme substrat lors de la réaction. Deux types d'ARNs peuvent être utilisés : *CYC1* ou *CYC1*pre ; correspondant à la région 3' non traduite de l'ARN du gène *CYC1*, codant pour l'iso-1-cytochrome C. Les matrices d'ADN employées sont les plasmides pG4-*CYC1* (promoteur T7 – Fragment +351/+588 – (Minvielle-Sebastia et al., 1994)) et pG4-*CYC1pre* (promoteur T7 – Fragment +351/+505 – (Preker et al.,

1995)) linéarisés respectivement par *EcoR*I et *Nde*I. La synthèse est catalysée par l'ARN polymérase du bactériophage T7 (25 unités ; Promega) dans son tampon ; auquel est ajouté : 50ng/µL ADN matrice, 0,01% TX-100, 0,5mM ATP, 0,5mM CTP, 0,1mM UTP, 0,1mM GTP, 1mM m^7GpppG, 2 unités/µL d'inhibiteur de RNAse. Le mélange réactionnel contient 1µL de [α^{32}P]-UTP (25µCi/800Ci/mmole). Après incubation 2 heures à 37°C, l'ARN est précipité par l'ajout de 500µL d'éthanol 100%, 130µL d'H_2O qualité biomol, 50µL d'acétate d'ammonium et 0,5µL de glycogène (Euromedex, 20mg/mL). Les ARNs sont récupérés par centrifugation, puis lavés par 1mL d'éthanol 70% et séchés pour être repris dans 20µL de tampon de charge (98% formamide, 10mM EDTA, 0,125% de Xylène cyanole FF et 0,125% de bleu de bromophénol). Après dénaturation 5 minutes à 90°C, les ARNs synthétiques sont séparés par électrophorèse sur gel 6% polyacrylamide (acrylamide/bisacrylamide 19 :1) / TBE 1X / 8,3M urée. La migration est effectuée à 900V pendant 45 minutes environ.

L'ARN est repéré sur le gel par autoradiographie. La bande est découpée et l'ARN est élué dans 400µL de tampon d'élution (0,75M acétate ammonium, 10mM MgAc, 0,1mM EDTA et 0,1% SDS) contenant 4µL d'un mélange phénol/chloroforme/alcool isoamylique (25/24/1) et 0,5µL de glycogène (Boehringer, 20mg/mL) ou 1µL de glycoblue (ThermoFisher Scientific, 15mg/mL). Après incubation sur la nuit à 37°C, le surnageant est mélangé à 400µL d'un mélange phénol/chloroforme/alcool isoamylique (25/24/1). La phase aqueuse est récupérée par centrifugation et précipitée par 1mL d'éthanol 100% froid et stockée à -20°C.

F.2.2.2. Calculs

Ces mesures sont réalisées deux fois.

Activité totale (A) :

1µL de mélange réactionnel initial est prélevé et dilué au centième. On mesure l'activité X (en coups par minutes, cpm), présent dans 5µL de cette dilution.

$A = (X/5) . 100 . 20 = X . 400$

Activité spécifique (B) :

Le mélange réactionnel contient 2.10^{-9} mol d'UTP. L'activité spécifique B du mélange réactionnel est exprimée en cpm/pmol UTP

$B = (A/2.10^{-9}) . 10\text{-}12 = A . 5\ 10^{-4}$

Activité incorporée (C) :

L'activité de 10µL (des 1400µL final) de suspension d'ARN, et noté Y, est mesurée

C = (1400/10) . Y = 140 . Y

Quantité d'UTP incorporé (D) :

Elle est exprimée en pmol d'UTP.

D = C / B

Concentration de la suspension d'ARN (F) :

L'ARN *CYC1* contient 97 résidus Uracile et *CYC1*-pré en contient 61.

La quantité (notée E) d'ARN dans la suspension finale, exprimée en pmol d'ARN, est alors égale à :

- pour *CYC1* : $E = (1/97) . D$
- pour *CYC1-pré* $E = (1/61) . D$

La concentration (F) de la suspension d'ARN est exprimée en fmol/µL

F = E / 1,4

F.3. Réaction de reconstitution des activités de maturation en 3' des ARN messagers

Après purification des facteurs multi protéiques, à partir des souches portant des protéines sauvages ou mutantes, leurs activités de maturation en 3' des ARN messagers peuvent être étudiées par reconstitution *in vitro*. Le mélange réactionnel (*TABLEAU 4.8*) utilisé et l'ARN substrat choisi permettent d'étudier les activités de clivage seul, de polyadénylation seule ou de clivage et de polyadénylation couplées. La présence de 2mM d'EDTA permet d'analyser le clivage seul d'un ARN substrat *CYC1*. La présence de 2mM de MgAc permet d'analyser le clivage et la polyadénylation d'un ARN substrat *CYC1*, ou la polyadénylation de l'ARN substrat pré-clivé - *CYC1*-pré. Le mélange réactionnel est incubé 1 heure à 30°C. 75µL d'un tampon contenant 1µL de protéinase K (20mg/mL), 0,25µL de glycogène (Boehringer, 20mg/mL) ou 0,5µL de glycoblue (ThermoFisher Scientific, 15mg/mL), et 50µL de tampon 2X (200mM TrisHCl pH 7,9, 300mM NaCl, 25mM EDTA, 2% SDS) sont ajoutés et le mélange est incubé 30 minutes à 42°C. Cette étape permet la dégradation des protéines et la libération des ARNs, qui peuvent être récupérés par précipitation par ajout de 2,5 volumes d'éthanol 100% froid. Après centrifugation, les culots d'ARNs sont lavés par 1mL d'éthanol 70%, séchés puis repris dans 6µL de tampon de charge et dénaturés 5 minutes à 90°C.

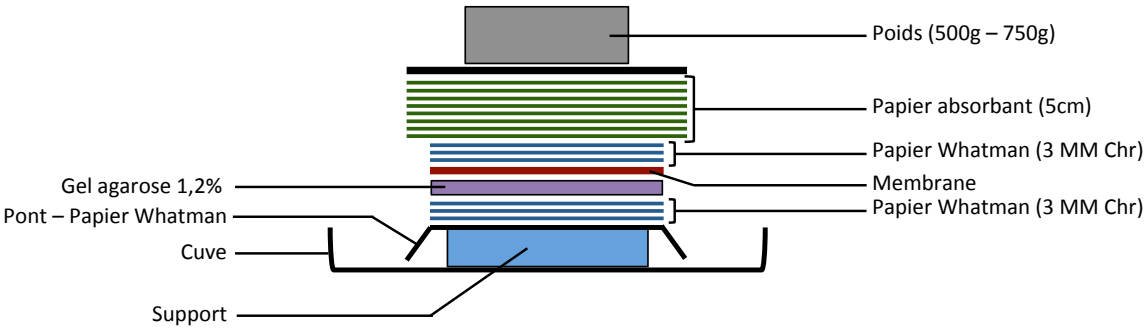


Figure 4.01 : Transfert sur membrane de Nylon.

Les ARNs sont chargés sur gel 6% polyacrylamide (acrylamide/bisacrylamide 19 :1)/ TBE 1X / 8,3M urée. La séparation des ARNs est effectuée à 1000V constant, pendant 45 minutes. La position des ARNs sur le gel est analysée par autoradiographie (Molecular Imager Pharos FX - BioRad).

F.4. Northern Blot

F.4.1. Préparation et migration des échantillons

Les ARNs sont séparés en fonction de leur taille par électrophorèse, sur gel agarose 1,2% dans du tampon 1X MOPS, en conditions dénaturantes (1,8% formaldéhyde). 5µg d'ARNs sont déposés pour chaque condition, en présence de tampon de charge (50% formamide, 6,2% formaldéhyde, 1X MOPS, 0,125% de Xylène cyanole FF, 0,125% de bleu de bromophénol, 1ng/mL Bromure d'éthidium, 10% glycérol), pour un volume final de 20µL. La migration est réalisée dans un tampon MOPS 1X, et est suivie sous UV (254nm).

F.4.2. Transfert sur membrane de Nylon

Les ARNs sont transférés sur membrane de Nylon N+ (Amersham Hbond N+, GE Healthcare) par capillarité dans du 20X SSC (**Figure 4.1**). Le transfert est effectué à température ambiante, sur la nuit. L'efficacité du transfert est vérifiée par coloration du gel au Bromure d'éthidium (EtBr ou BET). La membrane est exposée au rayonnement ultra-violet, afin de fixer de manière permanente l'ARN sur la membrane. Elle est ensuite rincée deux fois par du SSC 2X puis colorée au bleu de méthylène (0,04% de bleu de méthylène, 0,3M sodium acétate, pH5.3).

F.4.3. Préparation des sondes

La détection des ARNs *CYH2* et *TRS31* est effectuée par amorçage aléatoire ou « random priming ». La séquence codante est amplifiée par PCR à partir de 100ng d'ADN génomique et purifiée grâce au kit « GeneJet PCR purification kit » (ThermoScientific). Les produits de PCR obtenus servent alors de matrice pour la synthèse de sondes grâce au kit « *Prime-a-gene Labelling System* » (Promega) en présence de [α^{32}P]-CTP (25µCi/3000Ci/mmole). 50µL d'un mélange de sondes aléatoires sont obtenus, purifiés par chromatographie d'exclusion (Colonnes MicroSpin Sephadex G-25 – GE Healthcare) et stockés à -20°C.

L'ARN 18S est détecté un oligonucléotide (5' – CAGACAAATCACTCCA 3') marqué à son extrémité 5'. L'oligonucléotide (dilué à 100fmole/µL) est incubé 1h à 37°C en présence de 5µL de [γ^{32}P]-ATP (25µCi/6000Ci/mmole) et de 20 unités de Polynucléotide kinase T4 (T4 PNK, ThermoScientific). Les sondes sont purifiées par chromatographie d'exclusion G25 et stockées à -20°C.

F.4.4. Hybridation et révélation

La membrane est pré-hybridée à 42°C, 10 à 12 heures présence de 15mL tampon d'hybridation (SSC 5X, Denhardt's 5X - Eurobio, SDS 0,5%) auxquels sont ajoutés 3mg d'ARNt et 100µg d'ADN de sperme de saumon dénaturés et soniqués. L'hybridation est effectuée sur la nuit, à 42°C en présence de 50µL de sonde. La membrane est ensuite lavée 3 fois, à 60°C, en présence de 15mL de SSC 2X + 0,1% SDS et 1 fois en présence de 15mL de SSC 0,1X + 0,5% SDS. La détection des ARNs marqués est effectuée par autoradiographie (Molecular Imager Pharos FX - BioRad).

F.5. Essai de protection contre la digestion par la S1 Nucléase

F.5.1. Préparation des marqueurs de tailles

Deux marqueurs de tailles sont déposés sur chaque gel afin d'identifier de manière précise les produits obtenus par digestion à la S1 Nucléase. 1µg de produits de clivage du vecteur pBR322 (digéré par l'enzyme *Msp*I) ou 1µg du marqueur de taille 100pB (Invitrogen) est incubé 1 heure en présence de 20 unités de T4 PNK, et 3µL de [γ^{32}P]-ATP (25µCi / 6000Ci/mmole). Après dilution à 1fmol/µL, les marqueurs sont purifiés par chromatographie d'exclusion G25 et stockés à -20°C. L'équivalent de 400cps est déposé sur gel.

F.5.2. Préparation des sondes

Les sondes sont obtenues par digestion des plasmides pGEM4-*CYC1*, pGEM4-snR13-*TRS31*, pGEM4-snR5-*HEM4* par *Eco*RI/*Pvu*II, *Xho*I/*Pvu*II et *Nde*I/*Pvu*II respectivement.

Le marquage des sondes est réalisé extemporanément, par 3 à 9 Unités d'enzyme de Klenow. 30ng de sondes (*CYC1* et snR13*/TRS31*) sont incubés en présence de l'enzyme ainsi que de 500Ci/mmole de [α^{32}P]-TTP (25µCi / 3000Ci/mmole) et 0,25mM dATP. Pour snR5*/HEM4*, 20ng de sondes sont utilisés. Après 15 minutes de réaction à 25°C, 0,25mM de chaque dNTPs sont ajoutés au mix réactionnel pendant 5 minutes. Le volume est ensuite

ajusté à 50µL, et les sondes sont purifiées par chromatographie d'exclusion G25. Avant d'effectuer la réaction de protection, les sondes sont précipitées par l'ajout de 1/10 de volume de 3M d'acétate de sodium pH 5,3, 2,5 volumes d'éthanol 100% et 0,5µL de Glycoblue (15mg/mL). Le culot, contenant 30ng de sondes, est ensuite lavé par 1mL d'éthanol 70% et repris dans 25µL de tampon d'hybridation (80% formamide déionisé, 400mM NaCl, 10mM PIPES pH 6,4, 0,05% SDS, 1mM EDTA).

F.5.3. Essai de protection contre la digestion par la S1 Nucléase

Cette méthode est adaptée du protocole de Sittman *et al.,* (Sittman et al., 1983) Pour chaque condition, 30µg d'ARNs totaux ou 2µg d'ARNt de levures sont évaporés dans un tube de 1,5mL. 20 à 30ng de sondes reprises dans 25µL de tampon d'hybridation sont ensuite ajoutés. Le mélange est incubé 10 minutes à 90°C, refroidit 30 secondes dans un bain de carboglace mélangée à de l'éthanol puis incubés à 75°C dans un bain à sec progressivement refroidit à 58°C. Cette étape d'hybridation à lieu sur la nuit. La réaction est refroidie à 25°C (environ 2 heure) et 300µL de tampon de digestions S1 (50mM acétate de sodium pH 4,5, 280mM NaCl, 4,5mM $ZnSO_4$) contenant 150 unités de S1 nucléase, sont ajoutés (200 unités pour l'étude de snR5/*HEM4*). La réaction de digestion a lieu pendant 75 minutes à 25°C. Les hybrides ARN/ADN sont précipités par l'ajout de 2,5 volumes d'éthanol 100%, et incubés 1 heure à -20°C. Après centrifugation, les culots sont lavés par 1mL d'éthanol 70%, et repris dans 6µL de tampon de charge FA. Les hybrides ARN/ADN sont dénaturés 5 minutes à 90°C et chargés sur gel polyacrylamide (acrylamide/bisacrylamide 19 :1) / TBE 1X / 8,3M urée - 6% pour *CYC1* et snR5*/HEM4* ou 8% pour snR13*/TRS31*. La séparation est effectuée à 600V, pendant 2 heures. La position des ADN radiomarqués est analysée par autoradiographie (Molecular Imager Pharos FX – Biorad).

CHAPITRE 1 : ÉTUDE FONCTIONNELLE DES SOUS-DOMAINE DE PCF11

CHAPITRE 1 : ETUDE FONCTIONNELLE DES SOUS-DOMAINES DE PCF11

A. RAPPEL SUR PCF11

Pcf11 fait partie du sous-complexe de polyadénylation CF IA. Cette protéine de 72kDa est nécessaire aux activités de clivage et de polyadénylation. Pcf11 est formée de sept régions distinctes dont l'organisation est présentée dans la **Figure 2.01A** (Amrani et al., 1997b). La présence d'une répétition de vingt glutamines dans la région centrale de la protéine laisse suggérer que Pcf11 est une protéine multifonctionnelle, impliquée dans plusieurs fonctions en rapport avec la transcription par l'ARN polymérase II. De nombreuses données fonctionnelles sont disponibles, mais la structure de cette protéine n'est pas connue dans son entier. Seule la structure de la région N-terminale de la protéine (CID pour « *CTD interaction domain* » – 1-141 et 2nd NTD pour « *2nd N-terminal domain* » – 141-233) est disponible ; le CID a été défini comme nécessaire et suffisant pour la terminaison de transcription, via l'interaction avec le CTD de la Pol II et le pré-ARNm. La région C-terminale (288 à 626) est indispensable pour les activités de clivage et de polyadénylation. La structure du domaine d'interaction avec Clp1 est également disponible (Dupin and Fribourg, 2014; Noble et al., 2007).

A.1. Région N-terminale

La région N-terminale de la protéine, des acides aminés 1 à 253 est formée de deux domaines : le domaine d'interaction avec le CTD de l'ARN polymérase II et le 2nd NTD, récemment identifié (Xu et al., 2015).

Le domaine d'interaction avec le CTD de l'ARN polymérase II (CID) comprend les 141 premiers acides aminés. Le CID est nécessaire pour la viabilité chez *S. cerevisiae* (Sadowski et al., 2003) Ce domaine a été identifié par homologie de séquence avec d'autres protéines interagissant avec le CTD, comme Nrd1 ou Rtt103, puis confirmé par des analyses fonctionnelles. Le CID de Pcf11 présente une structure proche de celles des CID identifiés (Kubicek et al., 2012). La structure du CID des acides aminés 1 à 138 a été décrite par

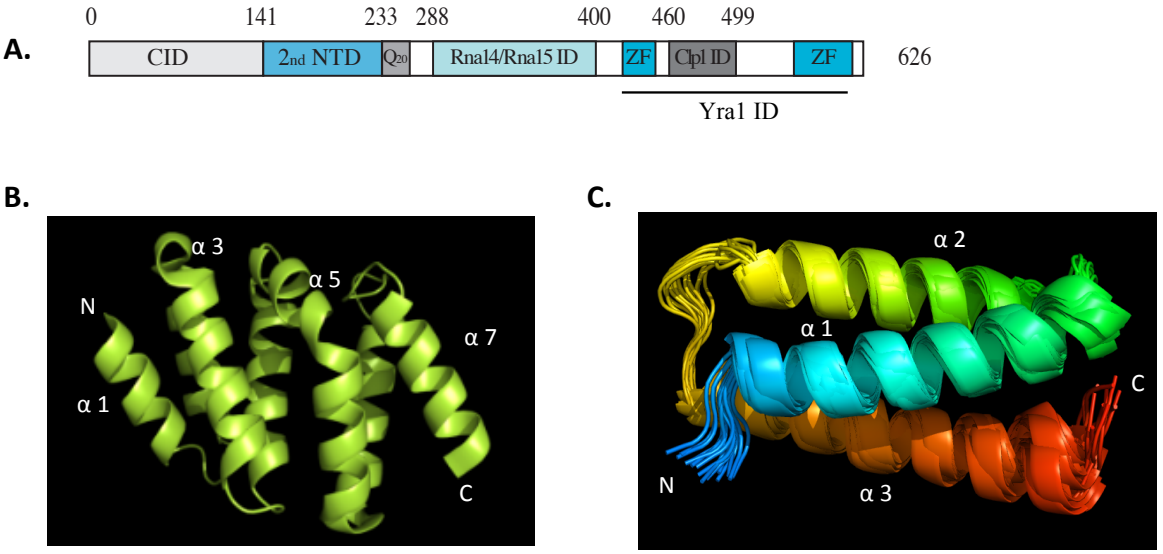


Figure 2.01 : Organisation et structure de Pcf11. **(A.)** Représentation schématique de l'organisation des différents domaines de Pcf11. Les différents domaines sont à l'échelle et leur taille est donnée en acides aminés. CID (1-141) : « *CTD Interaction Domain* ». 2nd NTD (141-233) : « *2nd N-terminal domain* ». Q_{20} (233-253) : motif Q_{20} formé d'une répétition de 20 résidus glutamines. Rna14/RNa15 ID (288-400) : région contenant les domaines d'interaction avec Rna14 et Rna15. Clp1 ID (460-499) : région contenant le domaine d'interaction avec Clp1. ZF (420-450 et 560-610) : région comprenant deux motifs en doigt de zinc. Yra1 ID (420-608) : région d'interaction minimale avec Yra1.
(B.) Structure du CID « *CTD Interaction Domain* » de Pcf11 (1-141). Le CID est formé de sept hélices α numérotées de 1 à 7. Le domaine est orienté de N-terminal (N) vers C-terminal (C). D'après Noble et *al.*, en 2005.
(C.) Structure du 2nd NTD de Pcf11 (141-233). Le domaine est orienté de N-terminal (N) vers C-terminal (C). Le 2nd NTD est formé de trois hélices α numérotées de 1 à 3 : hélice α1 en **bleu**, hélice α2 en **vert** et hélices α3 en **rouge**. D'après Xu *et al.,* 2015.

Meinhart et Cramer (Meinhart and Cramer, 2004), puis par Noble *et al.,* (Noble et al., 2005). Il est formé de huit hélices α liées par une séquence de 2 à 5 résidus (**Figure 2.01B**). Les hélices forment un domaine compact et stable qui permet la formation d'un domaine hydrophobe au centre de la molécule. Les hélices α7 et α8 sont les plus flexibles de la structure. L'interaction avec la forme P-SER^{2} du CTD se fait via les hélices α2, α4 et α7 et requiert la présence d'au moins deux répétitions de l'heptapetide. *In vivo*, cela permettrait l'augmentation de l'affinité. Le CID permet l'interaction avec les résidus TYR^{1} et PRO^{3} du CTD mais pas avec les résidus SER^{2} (Noble et al., 2005). *In vivo*, le CID peut interagir avec le CTD hypo-phosphorylé (Licatalosi et al., 2002) et avec sa forme en fin d'élongation – P-SER^{2} et PRO^{3}/ PRO^{6} *trans* (Barilla et al., 2001). L'interaction avec l'ARN, préférentiellement GA, se fait via des résidus situés entre les hélices α1/α2 et α3/α4, ainsi que des résidus situés sur l'hélice α7 (Hollingworth et al., 2006). Les interactions entre le CTD et l'ARN requièrent certains résidus en commun avec ses deux fonctions ; les expériences par SPR et RMN ont permis de mettre en évidence un mécanisme de compétition (Hollingworth et al., 2006).

D'un point de vue fonctionnel, des études (Barilla et al., 2001; Birse et al., 1998) ont mis en évidence le lien entre l'interaction avec le CTD et la terminaison de transcription. Sadowski *et al.,* (Sadowski et al., 2003) ont montré que la mutation des résidus conservés du CID, D68A, S69A, I70A (correspondant à l'allèle *pcf11-13*) inhibe l'interaction avec le CTD et la terminaison de transcription. La région comprenant le CID (1-180) est définie comme nécessaire et suffisante pour la terminaison de transcription. Un défaut de terminaison de la transcription n'est pas corrélé avec un défaut de l'activité de clivage. Pcf11 est impliquée dans la terminaison de transcription des ARNs codants et des ARNs non codants (terminaison NRD-dépendante) (Grzechnik et al., 2015), de manière indépendante de sa fonction dans la maturation en 3'. Dans le cas de la terminaison des ARNs codants, Pcf11 via son CID, est recrutée par interaction avec la forme P-SER^{2} du CTD et par interaction avec le facteur de terminaison Rtt103 (Lunde et al., 2010). Dans le cas de la terminaison des ARNs non codants (snoRNAs, CUTs), le recrutement de Pcf11 par interaction avec la forme P-SER^{2} du CTD permettrait de déstabiliser l'interaction entre Nrd1 et le CTD. Nrd1, via son CID, peut recruter le complexe TRAMP. Pcf11 permettrait le chargement de Sen1, une hélicase ARN/ADN, sur l'ADN ce qui déstabiliserait l'ARN polymérase II, permettant son décrochage (Grzechnik et al., 2015). Il a été montré que le CID de Pcf11 permet, par interaction avec l'ARN, le désassemblage du complexe d'élongation (Zhang et al., 2005).

Le 2nd NTD est formé des acides aminés 141 à 233. Récemment décrit (Xu et al., 2015), l'analyse par RMN a permis de déterminer que ce domaine est monomérique et d'une structure comparable à celle du CID. Le 2nd NTD est formé de trois hélices α liées par des séquences non structurées de 6 et 7 résidus, formant un domaine hydrophobe au centre de la molécule (**Figure 2.01C**). La structuration de ce domaine est indépendante de la présence du CID. L'hélice α3 comporte huit résidus glutamines exposés en surface vers la région C-terminale de la protéine. Les études préliminaires ont montré que le 2nd NTD n'interagit pas avec la forme P-SER^2 du CTD et que sa présence ne modifie pas la fixation du CID sur l'heptapetide. Le rôle de l'isomérisation des résidus prolines (PRO^3/ PRO^6) reste à déterminer. Le 2nd NTD interagit avec l'ARN, préférentiellement de séquence GA (C. Mackereth, *données non publiées*). Cette interaction fait intervenir des résidus situés en amont de l'hélice α1 et dans la séquence de liaison entre les hélices α2 et α3. Le CID et le 2nd NTD interagissent de manière coopérative pour l'interaction avec l'ARN (C. Mackereth, *données non publiées*). Aucune fonction n'est attribuée au 2nd NTD. Cependant, la structuration spécifique du 2nd NTD et son interaction avec l'ARN laisse supposer une fonction particulière.

Les régions N-terminale et C-terminale de Pcf11 sont séparées par une répétition de vingt résidus glutamines – motif Q_{20} (Amrani et al., 1997b). La délétion de ce motif ou d'une partie de ce motif n'induit pas de modification de croissance chez *S. cerevisiae*. Cependant, la délétion de cette répétition, en association avec la délétion du 2nd NTD, aggrave le phénotype observé (cette étude).

A.2. Région C-terminale

Sadowski *et al.,* (Sadowski et al., 2003) ont montré que les mutations du CID situées dans la partie N-terminale de la protéine ne modifiaient pas les activités de clivage et de polyadénylation. La région C-terminale, comprenant les acides aminés 266 à 626, est définie comme nécessaire et suffisante aux activités de clivage et de polyadénylation *in vitro.* En accord avec cette fonction, la région C-terminale comprend les domaines d'interaction avec les autres partenaires du complexe CF IA : Rna14/Rna15 et Clp1 (**Figure 2.01A**). Deux régions contenant des motifs en doigt de zinc sont également retrouvées encadrant le domaine d'interaction avec Clp1.

Le domaine d'interaction avec Rna14 et Rna15 est défini par les acides aminés 288 à 400 – **Rna14/Rna15 ID**. La structure de ce domaine de Pcf11 n'est pour l'instant pas résolue. La région d'interaction au niveau de Rna14 n'est également pas définie, mais l'organisation de la protéine suggère que ce domaine pourrait être situé dans la région C-terminale de la protéine (« *monkey tail domain* »). Pcf11 interagit avec les 50 acides aminés en C-terminal de Rna15, donc la structure a été décrite par Qu *et al.,* en 2007, par homologie avec CstF-64 (Qu et al., 2007). Le domaine de Rna15 interagissant avec Pcf11 est formé de deux hélices α (quatre tours) connectées par une hélice α plus petite (deux tours).

Pcf11 interagit avec la quatrième sous-unité du complexe CF IA, Clp1, via les résidus 460 à 499 (**Clp1 ID**) (Sadowski et al., 2003) (Noble et al., 2007). La structure de ce domaine a été décrite par Noble *et al.,* en 2007 (Noble et al., 2007) et affinée en 2014 par Dupin *et al.,* (Dupin and Fribourg, 2014). Le **Clp1 ID** de Pcf11 contient trois résidus (R480, W482 et W489) fortement conservés , impliqués dans des liaisons hydrogènes et des interactions hydrophobes avec Clp1. La région d'interaction de Clp1 comprend quatre résidus fortement conservés Y332A, F333A, Y334A et G335A ; la mutation de ces quatre résidus n'est pas viable chez *S. cerevisiae* (L. Minvielle-Sébastia, *données non publiées*). Des mutations de la région d'interaction entre Clp1-Pcf11 au niveau de Clp1, modifient l'assemblage du complexe CF IA *in vivo* et lors de la co-purification par *TAP-tagging* (Haddad et al., 2012).

Enfin, Sadowski *et al.,* (Sadowski et al., 2003) ont décrit la présence de deux régions, encadrant le domaine d'interaction avec Clp1, contenant des motifs supposés lier un ou plusieurs atomes de zinc. Cependant ces domaines restent à ce jour peu étudiés. D'un point de vue structural, le premier motif de type **Cys_2His_2**, situé entre les résidus 420 à 450 est replié sous la forme ββα et décrit par la séquence X_2-**Cys**-$X_{2,4}$-**Cys**-X_{12}-**His**-$X_{3,4,5}$-**His.** Le second motif (situé entre les résidus 560 à 610) est caractérisé par les résidus **Cys_2HisCys**, caractéristique des motifs de type « *zinc knuckle* », formé de deux feuillets β liés par une boucle suivie d'une hélice α (Krishna et al., 2003). Chaque motif est décrit comme pouvant potentiellement lier un atome de zinc. Aucune fonction n'est pour l'instant attribuée à ces deux motifs. Les motifs liant le zinc sont impliqués dans des interactions protéine-protéine et/ou protéine-ARN (Matthews and Sunde, 2002).

Plus récemment, la région 420 à 608 de Pcf11 a été définie comme nécessaire pour l'interaction avec Yra1, protéine adaptatrice impliquée dans l'export (Johnson et al., 2011) (**Figure 2.01A**). Cette région comprend les deux motifs supposés lier le zinc et le domaine d'interaction avec Clp1 ; cependant, les résidus impliqués dans l'interaction ne sont pas

connus. Les motifs liant le zinc sont des candidats potentiels de cette interaction. Yra1 est, comme Pcf11, recrutée par interaction avec le CTD de l'ARN polymérase II (MacKellar and Greenleaf, 2011).

B. OBJECTIFS DE L'ETUDE

L'objectif de ce travail est d'étudier le rôle du 2nd NTD (141-233) et des motifs liant le zinc (420 à 450 et 560 à 610), dans (i) la viabilité cellulaire, (ii) les activités de maturation de l'extrémité 3' et (iii) la terminaison de transcription des ARNs codants et non codants. Nous avons choisi de mettre en place une stratégie de génétique inverse. Ce travail a été réalisé en collaboration avec deux équipes de biologie structurale à l'IECB : l'équipe de C. Mackereth (RMN) et l'équipe de S. Fribourg (cristallographie).

Dans la première partie de ce travail, sont décrits les résultats obtenus pour l'étude du 2nd NTD. Nous avons, tout d'abord, construits différents mutants de délétions et ponctuels à partir des données obtenues par RMN pour étudier l'impact du 2nd NTD sur la viabilité cellulaire. Nous avons ensuite caractérisé le phénotype des mutants obtenus, en particulier lors de la délétion du domaine entier, puis l'assemblage du complexe CF IA. Nous avons ensuite étudié l'impact des mutations sur les activités de maturation de l'extrémité 3' (clivage et clivage/polyadénylation) et sur la terminaison de transcription des ARNm et des snoRNAs. Les différents mutants utilisés dans cette étude ainsi que les résultats obtenus pour chaque mutant sont récapitulés dans le *TABLEAU 2.1*.

Dans la seconde partie de ce travail, nous avons essayé de comprendre le rôle des motifs liant le zinc, dans les fonctions de Pcf11. Les travaux ont été menés en parallèle, avec la même logique de travail : obtention des mutants, caractérisation du phénotype, impact sur l'assemblage du complexe et étude fonctionnelle à proprement dite - rôle dans la maturation de l'extrémité 3' et la terminaison de transcription. Les domaines liant le zinc encadrant la région d'interaction avec Clp1, nous avons enfin recherché une létalité synthétique entre différents allèles ou une aggravation du phénotype de croissance de chacun des mutants pris individuellement. Les différents mutants obtenus dans cette étude ainsi que les résultats obtenus pour chaque mutant sont récapitulés dans le *TABLEAU 2.2*.

C. ÉTUDE DU 2ND NTD DE PCF11

C.1. Résultats

C.1.1. Obtention et caractérisation des mutants

C.1.1.1. Obtentions des mutants

La première étape dans l'étude du 2nd NTD de Pcf11 a été de créer différents mutants, soit par délétion du domaine entier ou une des trois hélices formant ce domaine, soit par des mutations ponctuelles. En parallèle, nous avons souhaité étudier le rôle du motif **Q_{20}**, supposé jouer le rôle de *linker* entre les parties N-terminale et C-terminale de la protéine (Amrani et al., 1997b) (Sadowski et al., 2003) ainsi que l'interaction du 2nd NTD avec le CID présent en amont de la protéine. Les différents mutants obtenus sont présentés dans la **Figure 2.02B**.

En délétant la totalité du domaine (**Δ(141-233)** c.), nous souhaitions savoir si ce domaine était essentiel pour la fonction de la protéine. Nous avons également obtenu un mutant délété pour le 2nd NTD et une partie du motif Q_{20} - **Δ(141-241)**. Le phénotype associé à cette mutation est plus marqué que celui associé au mutant délété pour le 2nd NTD seul ; il ne sera pas traité dans le reste de l'étude du 2nd NTD.

Les acides aminés mutés ont été choisis sur la base de la structure obtenue par RMN. L'objectif était de modifier les interactions entre les hélices deux à deux, afin de déstructurer le domaine ; la double mutation L189D, L213D (***pcf11-17*** – i.) avait pour objectif de modifier les interaction entre les trois hélices du domaine et de déstructurer le 2nd NTD dans son ensemble. Les acides aminés mutés sont situés dans la région hydrophobe formée par les trois hélices (**Figure 2.02C**).

- La Lysine en position 189 est située dans la zone d'interaction entre les hélices α1 et α3 et est remplacée par un Acide aspartique (***pcf11-14***) g.
- La Lysine en position 213 est située dans la zone d'interaction entre les hélices α2 et α3 et est remplacée par un Acide aspartique (***pcf11-15***) h.
- L'Acide aspartique en position 220 est situé dans l'hélice α3 et est remplacé par une Alanine (***pcf11-16***) j.

Avec la délétion de l'une des trois hélices, nous avions pour objectif d'étudier l'importance de chacune des hélices dans la structure du 2nd NTD et dans sa fonction.

- Délétion de l'hélice α1 - **Δ(155-174)** d.
- Délétion de l'hélice α2 - **Δ(180-196)** e.
- Délétion de l'hélice α3 - **Δ(203-227)** f.

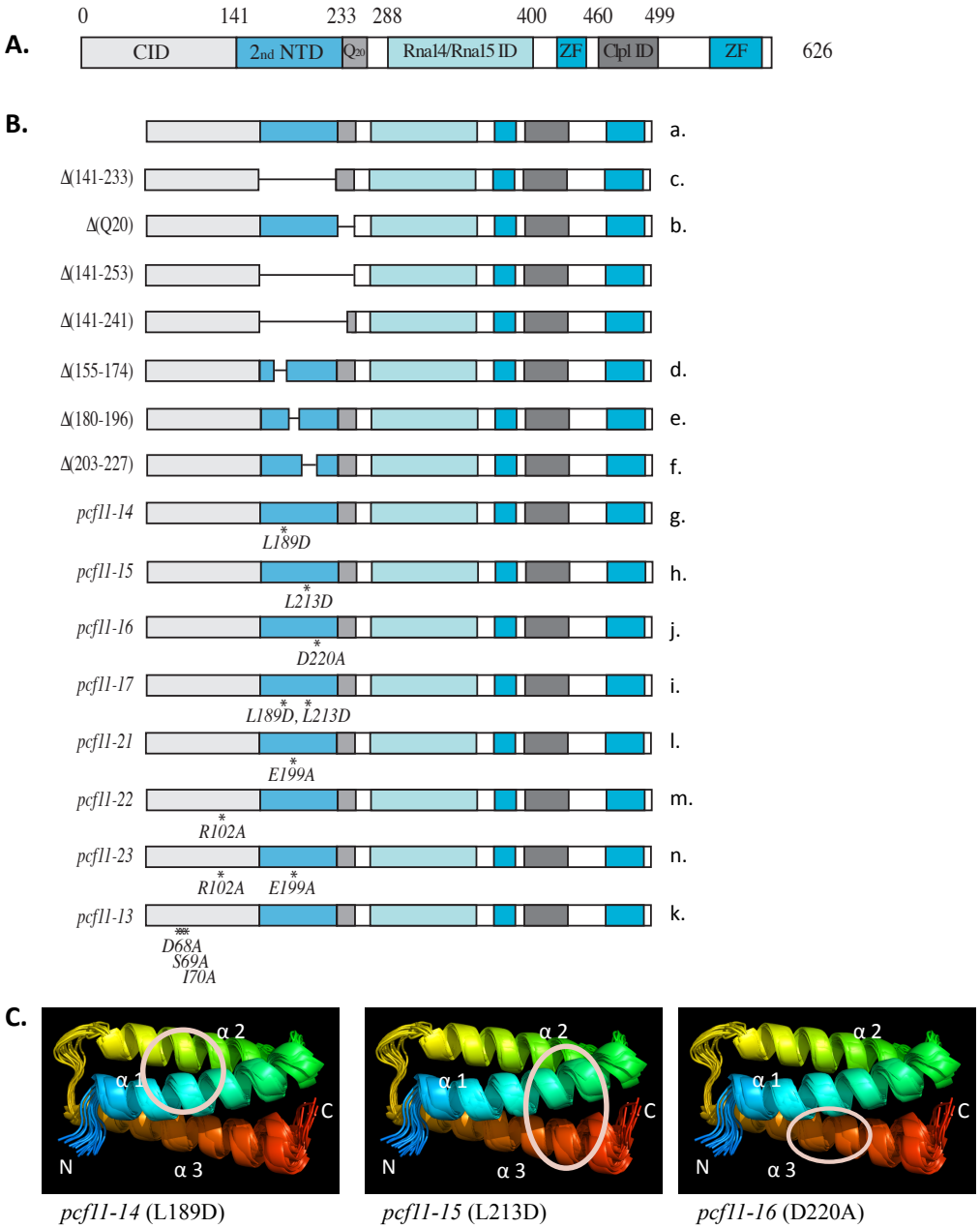


Figure 2.02 : Organisation des différents mutants construits pour l'étude du 2nd NTD.

(A.) Représentation schématique de l'organisation des différents domaines de Pcf11. Les différents domaines sont à l'échelle et leur taille est donnée en acides aminés. CID (1-141) : « *CTD Interaction Domain* ». 2nd NTD (141-233) : « *2nd N-terminal domain* ». Q_{20} (233-253) : motif Q_{20} formé d'une répétition de 20 résidus glutamines. Rna14/RNa15 ID (288-400) : région contenant les domaines d'interaction avec Rna14 et Rna15. Clp1 ID (460-499) : région contenant le domaine d'interaction avec Clp1. ZF (420-450 et 560-610) : région comprenant deux motifs en doigt de zinc.

(B.) Représentation des différents mutants utilisés dans l'étude du 2nd NTD, annotés de a. à n.. Les délétions sont représentées par des bandes horizontales. Les * représentent les mutations ponctuelles ; les résidus mutés sont indiqués en italique.

(C.) Représentation des acides aminés mutés sur la structure du 2nd NTD. Leur localisation est représentée par un cercle blanc. La mutation L189D est située dans la zone d'interaction des hélices α1 et α3, la mutation L213D dans la zone d'interaction entre les hélices α2 et α3 et la mutation D220A au niveau de l'hélice α3.

Le mutant ***pcf11-13*** (D68A, S69A, I70A – k.) est utilisé dans cette étude comme contrôle ; les mutations introduites au niveau du CID ne modifient pas les activités de maturation en 3' des ARN messagers mais altèrent l'interaction avec le CTD de l'ARN polymérase II et la terminaison de transcription (Sadowski et al., 2003).

Parallèlement pour étudier l'interaction avec l'ARN, nous avons étudié le phénotype de croissance de deux mutants, obtenus par nos collaborateurs à l'IECB, où l'interaction avec le pré-ARNm est diminuée. La double mutation R102A ; E199A (***pcf11-23*** - n) avait pour objectif de voir s'il y a une synergie des effets.

- L'Arginine en position 102 en Alanine (***pcf11-22*** – m.). L'interaction avec l'ARN est réduite, mais la structure du 2nd NTD est conservée ainsi que l'interaction avec le CTD de l'ARN polymérase II.
- L'Acide glutamique en position 199 en Alanine (***pcf11-21*** - l.). Ce résidu est situé dans la région *linker* entre les hélices α2 et α3. La structure du 2nd NTD n'est pas modifiée mais l'interaction avec l'ARN est réduite.

Comme tous les gènes codant pour les composants de la machinerie de polyadénylation, *PCF11* est essentiel chez *Saccharomyces cerevisiae* (Amrani et al., 1997b). L'effet des différentes mutations est étudié indirectement. Au laboratoire, nous disposons d'une souche, la souche YLM233, délétée du gène *PCF11* au niveau du chromosome ; cette délétion est complémentée par le vecteur d'expression pFL38-*PCF11* (pNA39) (Amrani et al., 1997b) portant le marqueur *URA3*. Les différentes mutations étudiées sont introduites dans le plasmide pFL36ΔSX-*PCF11* (pFL36 dans lequel les sites *Sph*I et *Xho*I ont été supprimés et portant le marqueur *LEU2*) par recombinaison homologue chez la levure. Après transformation de la souche YLM233 par le plasmide pFL36ΔSX portant un allèle sauvage ou mutant du gène *PCF11*, le plasmide pFL38-*PCF11* est chassé sur milieu contenant 0,1% d'acide 5-Fluoro-orotique (ou 5-FOA). Les souches obtenues ne portent qu'une copie de *PCF11* portée par le plasmide pFL36ΔSX. Cette approche a pu être utilisée pour l'obtention des mutants ponctuels et également des mutants de délétion du 2nd NTD, des trois hélices du 2nd NTD ainsi que la délétion du motif Q_{20}. La délétion du 2nd NTD induit un phénotype de croissance lente, décrit plus loin dans ce manuscrit. La double délétion [2ndNTD+Q_{20}] correspondant au mutant **Δ(141-253)** n'a pas pu être étudiée par cette approche. En effet, les différentes souches obtenues après la chasse du plasmide sur 5-FOA ne présentaient pas de phénotype de croissance lente comme observé lors de la délétion du 2nd NTD. De plus, la vérification sur milieu sélectif (milieu SC-L-H en absence et en présence d'uracile) et la taille

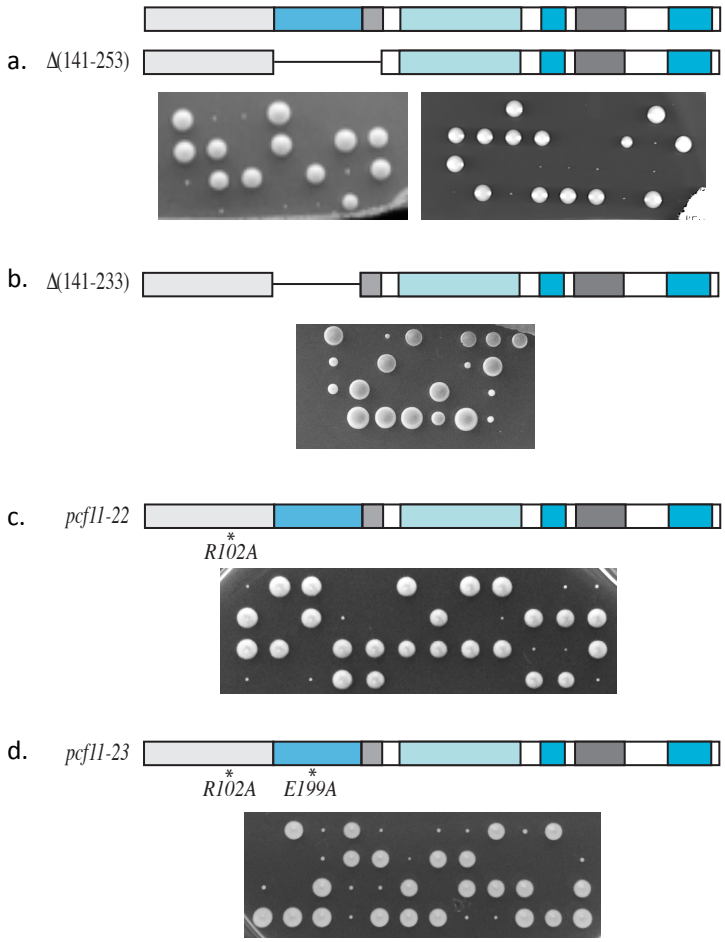


Figure 2.03 : La délétion du 2nd NTD induit un phénotype de croissance lente chez *S. cerevisiae* à 24°C. Étude de la ségrégation des spores pour les différents mutants du CID et du 2nd NTD. Les spores sont séparés à l'aide d'un micromanipulateur sur milieu YPD(A) puis incubés à 24°C. ***a.*** Ségrégation des spores pour le mutant (141-253) correspondant à la délétion [2nd NTD+Q_{20}] après 72 heures d'incubation. ***b.*** Ségrégation des spores pour le mutant (141-223) délété pour le 2nd NTD après 72 heures d'incubation. ***c.*** Ségrégation des spores pour le mutant *pcf11-22* (R102A) après 6 jours d'incubation. ***d.*** Ségrégation des spores pour le mutant *pcf11-23* (R102A, E199A) après 6 jours d'incubation.

de la copie de *PCF11* exprimée dans cette souche a permis de montrer que ces souches portaient et exprimaient une version sauvage du gène *PCF11* (*données non montrées*). Cela peut être dû à la présence du plasmide pFL38-*PCF11* qui n'est pas correctement éliminé lors de la chasse sur 5-FOA ou à la létalité de la double délétion, entraînant la production de révertants pour pallier cette dernière.

Afin d'étudier l'impact réel de cette mutation, nous avons transformé la souche diploïde YLM232 (MAT a/α) par le plasmide portant la délétion **Δ(141-253)**. La souche YLM232 est une souche diploïde hétérozygote, portant une délétion du gène *PCF11* au niveau chromosomique – *pcf11*::*TRP1*/*PCF11*. Après sporulation, l'étude de la ségrégation des spores a permis de montrer que cette double délétion n'induit pas de létalité chez *S. cerevisiae*, mais entraîne un phénotype de croissance lente important (**Figure 2.03** a.) en comparaison des spores portant une copie sauvage du gène et à la délétion du 2nd NTD uniquement (**Figure 2.03** b.). De par le phénotype de croissance lente important et la capacité à former rapidement des révertants, nous avons décidé de ne pas inclure ce mutant dans le reste de cette étude. Nous avons également utilisé cette approche pour étudier la mutation R102A (***pcf11-22***) et la double mutation R102A, E199A (***pcf11-23***). L'étude de la ségrégation des spores (**Figure 2.03** c. et d.) permet de mettre en évidence un phénotype de croissance lente important induit par la mutation R102A. La double mutation n'est pas létale mais induit un phénotype de croissance lente important, ce mutant ne sera pas utilisé dans le reste de l'étude.

C.1.1.2. Caractérisation du phénotype associé

Après obtention des mutants, nous avons étudié le phénotype de chaque délétion et mutation ponctuelle, sur milieu complet YPD(A) à différentes températures. Les phénotypes obtenus sont présentés dans la **Figure 2.04A** et regroupés dans le *TABLEAU 2.1*. La délétion du 2nd NTD (**Δ(141-233))** induit un phénotype de croissance lente aux températures physiologiques de croissance et une létalité à 37°C (ligne c en comparaison à la ligne a), ce qui démontre un rôle important de ce domaine pour la viabilité cellulaire. Pour obtenir une estimation du phénotype de croissance lente causée par la délétion du 2nd NTD, nous avons effectué des courbes de croissance à partir de culture en milieu liquide à 30°C durant 24 et 48 heures en comparaison avec la souche YLM237, portant le plasmide pFL36ΔSX-*PCF11* et la souche délétée pour le motif $\mathbf{Q_{20}}$ (**Figure 2.05A**). Comme observé dans les tests en goutte (ligne b en comparaison à la ligne a), la délétion du motif $\mathbf{Q_{20}}$ ne modifie pas la croissance des

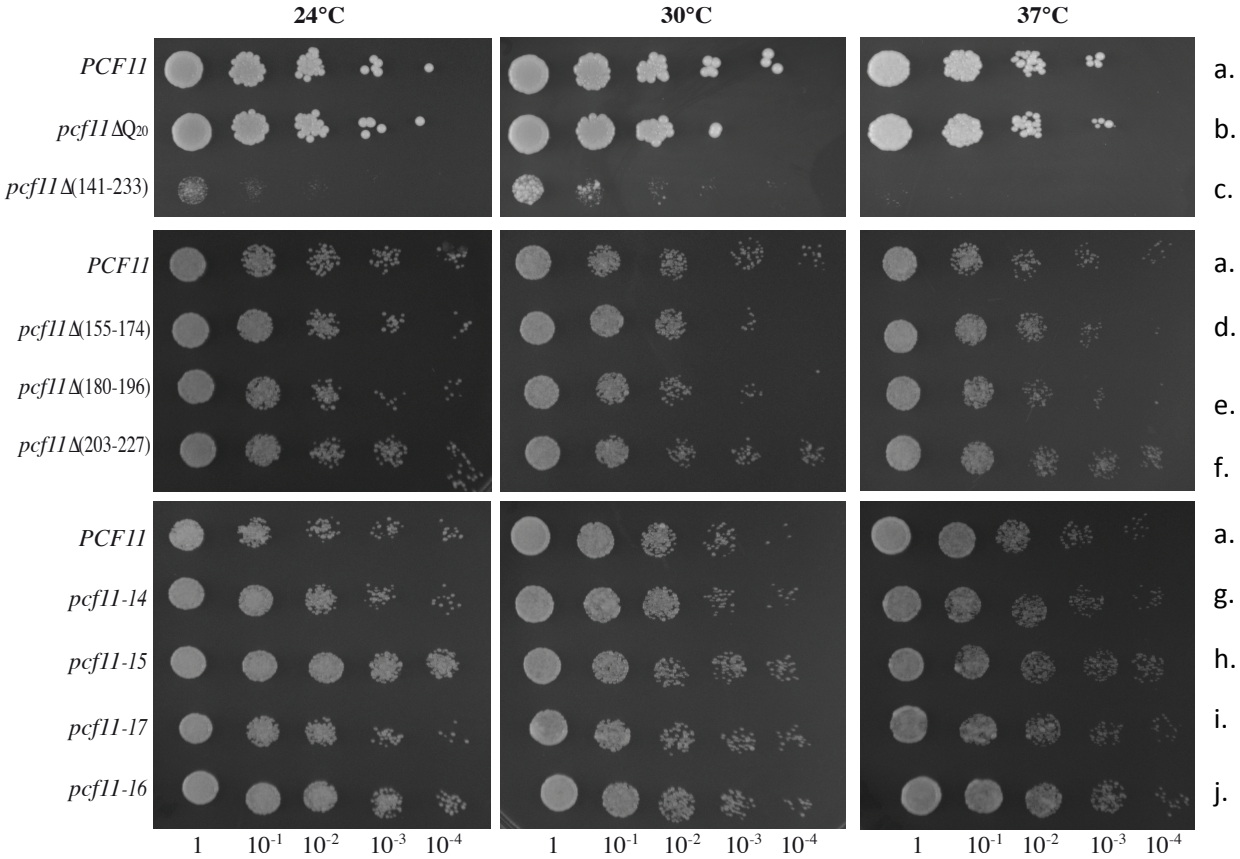


Figure 2.04 : Phénotypes associés aux différentes mutations du 2nd NTD. Analyse des phénotypes associés aux mutants du 2nd NTD de délétions et aux mutants ponctuels à 24°C, 30°C et 37°C sur milieu YPD(A). ***a.*** témoin positif ; souche exprimant une version sauvage de *PCF11*. ***b.*** pcf11ΔQ$_{20}$: souche délétée pour le motif Q$_{20}$. ***c.*** pcf11Δ(141-233) : souche délétée pour le 2nd NTD. ***d.*** pcf11Δ(155-174) : souche délétée pour l'hélice α1. ***e.*** pcf11Δ(180-196) : souche délétée pour l'hélice α2. ***f.*** pcf11Δ(203-227) : souche délétée pour l'hélice α3. ***g.*** *pcf11-14* : mutant ponctuel (L189D). ***h.*** *pcf11-15 :* mutant ponctuel (L213D). ***i.*** *pcf11-17 :* mutant ponctuel (L189D, L213D). ***j.*** *pcf11-16 :* mutant ponctuel (D220A).

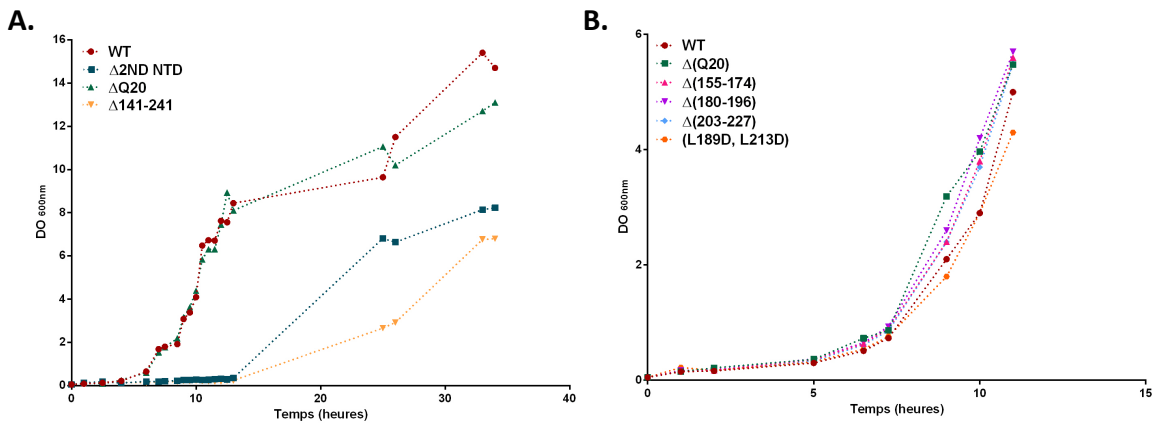


Figure 2.05 : Caractérisation du phénotype de croissance lente du mutant délété pour le 2nd NTD.
(A.) Courbe de croissance en milieu liquide à 30°C pendant 48 heures. La DO à 600nm est donnée en fonction du temps d'incubation en heures. (● - WT) souche exprimant une version sauvage de *PCF11*. (■) pcf11Δ(141-233) : souche délétée pour le 2nd NTD. (**Δ**) pcf11ΔQ$_{20}$: souche délétée pour le motif Q$_{20}$. (**Δ**) pcf11Δ(141-241) : souche délétée pour le 2nd NTD et une partie du motif Q$_{20}$.
(B.) Courbes de croissance en milieu liquide à 30°C pendant 12 heures. La DO à 600nm est donnée en fonction du temps d'incubation en heures. (● - WT) souche exprimant une version sauvage de *PCF11*. (■) pcf11ΔQ$_{20}$: souche délétée pour le motif Q$_{20}$. (**Δ**) pcf11Δ(155-174) : souche délétée pour l'hélice α1. (**Δ**) pcf11Δ(180-196) : souche délétée pour l'hélice α2. (◆) pcf11Δ(203-227) : souche délétée pour l'hélice α3. (◆) *pcf11-17 :* mutant ponctuel (L189D, L213D).

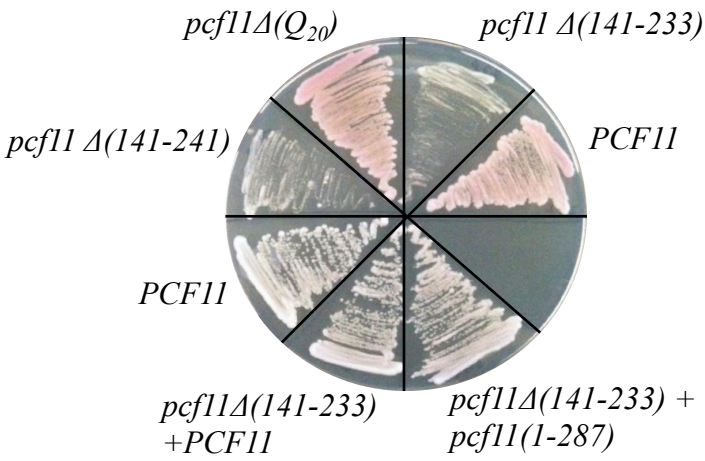


Figure 2.06 : La *trans*-complémentation de la délétion du 2nd NTD par la protéine entière ou la région N-terminale permet de restaurer le phénotype sauvage. *PCF11* : souche portant une copie sauvage de *PCF11*. pcf11ΔQ$_{20}$: souche délétée pour le motif Q$_{20}$. pcf11Δ(141-233) : souche délétée pour le 2nd NTD. pcf11Δ(141-233)+*PCF11* : souche délétée pour le 2nd NTD et complémenté par la protéine entière. pcf11Δ(141-233)+*pcf11(1-287)* : souche délétée pour le 2nd NTD et complémentée par la partie N-terminale de la protéine. Les phénotypes sont analysés sur milieu sélectif.

souches (temps de génération d'environ 1h52 contre 1h39 pour la souche sauvage – courbe **Δ**). Le temps de génération estimé est plus important pour la souche mutante **Δ(141-233)** que pour la souche sauvage (temps de génération estimé de 5h – courbe ■). Cela confirme l'importance du 2nd NTD pour la croissance. On observe également que le mutant **Δ(141-241)** présente un phénotype de croissance lente plus prononcé que le mutant **Δ(141-233)** (DO_{600nm} *PCF11* :0,96 ; *pcf11Δ(141-233)* :0,216 et *pcf11Δ(141-241)* :0,162 courbe **Δ**).

Dans le cas de Pcf11, la complémentation en *trans* par la protéine entière ou une partie de la protéine – le fragment N-terminal – permet de déterminer si deux mutations sont situées sur une même unité fonctionnelle. Cette approche a été utilisée par Sadowski *et al.* (Sadowski et al., 2003) ; le phénotype de croissance de la souche portant l'allèle *pcf11-13* peut être restauré par complémentation du fragment N-terminal 1-260 de Pcf11. De la même manière, la *trans*-complémentation de la souche **Δ(141-233)** par le plasmide pHH2 portant une version sauvage de *PCF11* (YLM242) ou la partie N-terminale de la protéine Δ(287-626) (YLM243) permet de restaurer un phénotype sauvage (**Figure 2.06**). Cela suggère que le phénotype de croissance lente observé peut être directement attribué à l'absence du 2nd NTD et que la fonction du 2nd NTD est partiellement découplée de la fonction de la région C-terminale de la protéine.

La délétion du motif $\mathbf{Q_{20}}$ ne modifie pas la croissance des souches aux températures étudiées (24°C, 30°C ou 37°C). De manière surprenante, les délétions de chacune des trois hélices du 2nd NTD n'induit pas de modification du phénotype à 24°C, 30°C et 37°C par rapport à la souche portant une version sauvage de *PCF11*, de même que les mutations ponctuelles du 2nd NTD étudiées (**Figure 2.04A**, ligne d à j). Nous avons également effectué des courbes de croissance en milieu YPD(A) liquide, à 30°C pendant 12 heures (**Figure 2.05B**). Les différents mutants testés ne présentent pas de différence de croissance, en comparaison avec la souche portant une version sauvage de *PCF11*, permettant ainsi de confirmer les résultats obtenus précédemment.

Les mutations ponctuelles introduites dans la région hydrophobe formée par les trois hélices du 2nd NTD n'induisent pas de phénotype associé à la température. Pour aller plus loin dans l'étude de ces mutations et pour tenter d'aggraver leur effet, nous avons étudié le phénotype de ces souches à 30°C sur milieu YPD(A) en présence de 2% de formamide. De la même manière que la sensibilité à l'éthanol, une sensibilité au formamide est décrite comme illustrant une modification de la stabilité des protéines (Hampsey, 1997). Les résultats obtenus

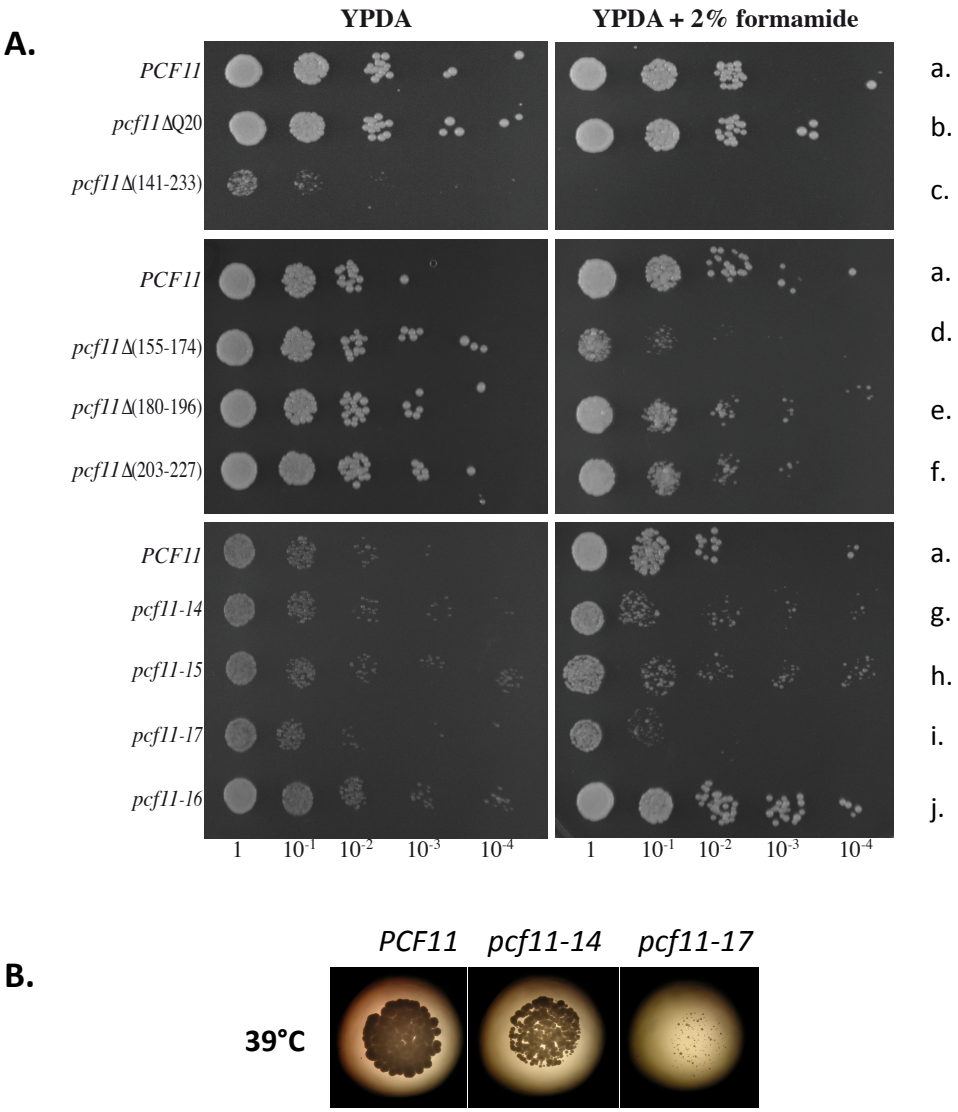


Figure 2.07 : Étude de la sensibilité au formamide des différents mutants sur 2nd NTD.

(A.) Phénotypes associés aux différentes mutations du 2nd NTD à 30°C sur milieu YPD(A) en absence (*à gauche*) et en présence de 2% de formamide (*à droite*). ***a.*** témoin positif ; souche exprimant une version sauvage de *PCF11*. ***b.*** pcf11ΔQ$_{20}$: souche délétée pour le motif Q$_{20}$. ***c.*** pcf11Δ(141-233) : souche délétée pour le 2nd NTD. ***d.*** pcf11Δ(155-174) : souche délétée pour l'hélice α1. ***e.*** pcf11Δ(180-196) : souche délétée pour l'hélice α2. ***f.*** pcf11Δ(203-227) : souche délétée pour l'hélice α3. ***g.*** *pcf11-14* : mutant ponctuel (L189D). ***h.*** *pcf11-15 :* mutant ponctuel (L213D). ***i.*** *pcf11-17 :* mutant ponctuel (L189D ; L213D). ***j.*** *pcf11-16 :* mutant ponctuel (D220A).

(B.) Aspect des colonies pour les mutations des allèles *pcf11-14* et *pcf11-17* à 39°C sur YPD(A) en comparaison à la souche exprimant une version sauvage de *PCF11* (*à gauche*). Les colonies sont observées à la loupe binoculaire.

en présence de 2% de formamide sont présentés **Figure 2.07**. La délétion du 2nd NTD induit un arrêt de croissance en présence de 2% de formamide (ligne c). La délétion de l'hélice **α1** (correspondant à la délétion **Δ(155-174)** ligne d) conduit à une sensibilité au formamide à 30°C, ce qui n'est pas le cas pour les autres mutants des hélices **α2** et **α3**, du motif $\mathbf{Q_{20}}$ ainsi que le mutant ***pcf11-16*** (D220A) (lignes b, e, f, et j). Les mutations ponctuelles des souches ***pcf11-14*** (L189D) et ***pcf11-15*** (L213D) situées dans la poche hydrophobe formée par les hélices α induisent également une sensibilité au formamide (lignes g et h). Le mutant ***pcf11-17*** (ligne i) portant la double mutation (L189D, L213D), présente logiquement un phénotype plus marqué. Cela suggère que la modification des interactions entre les trois hélices du 2nd NTD pourrait conduire à une modification de la stabilité du 2nd NTD. La sensibilité au formamide des mutants ***pcf11-14*** (L189D) et ***pcf11-17*** (L189D, L213D) est corrélée à l'analyse des colonies obtenues pour ces différents mutants à une température de 39°C (**Figure 2.07B**). À cette température élevée, après 72 heures d'incubation, on observe une diminution de la taille des colonies pour les souches portant l'allèle ***pcf11-14*** (L189D) en comparaison à la souche portant une version sauvage de *PCF11*. Pour les souches portant l'allèle ***pcf11-17***, les colonies présentent une taille hétérogène, beaucoup plus petite en comparaison à la souche portant une version sauvage de *PCF11*.

La mutation R102A (***pcf11-22***) conduit à un phénotype de croissance lente important à 24°C et 30°C (**Figure 2.08B**, ligne m) et une thermosensibilité à 37°C. Le phénotype en présence de 2% de formamide n'a pas été étudié pour ce mutant. Le mutant ***pcf11-21*** (E199A) ne présente pas de phénotype, en milieu solide aux températures testées (**Figure 2.08B**, ligne l) ni en présence de 1,5% formamide (**Figure 2.08C**, ligne l.). Les résultats préliminaires obtenus par RMN indiquent que l'interaction du CID et du 2nd NTD avec l'ARN est réduite par ces deux mutations. L'interaction entre le CID et l'ARN serait donc essentielle pour la viabilité et prédominante par rapport à l'interaction entre le 2nd NTD et l'ARN. Cependant, le cumul de ces deux mutations - ***pcf11-23*** (R102A, E199A) – aggrave le phénotype observé à 24°C, 30°C et 37°C. Le 2nd NTD pourrait renforcer l'interaction du CID avec l'ARN. L'interaction de la partie N-terminale de la protéine (CID + 2nd NTD) serait essentielle pour la viabilité. Le phénotype associé à la souche ***pcf11-13*** (D68A, S69A, I70A) est identique à celui décrit par Sadowski *et al.,* 2003 (Sadowski et al., 2003). Les mutations induisent un léger phénotype de croissance lente à 24°C et 30°C, ainsi qu'une sensibilité à la

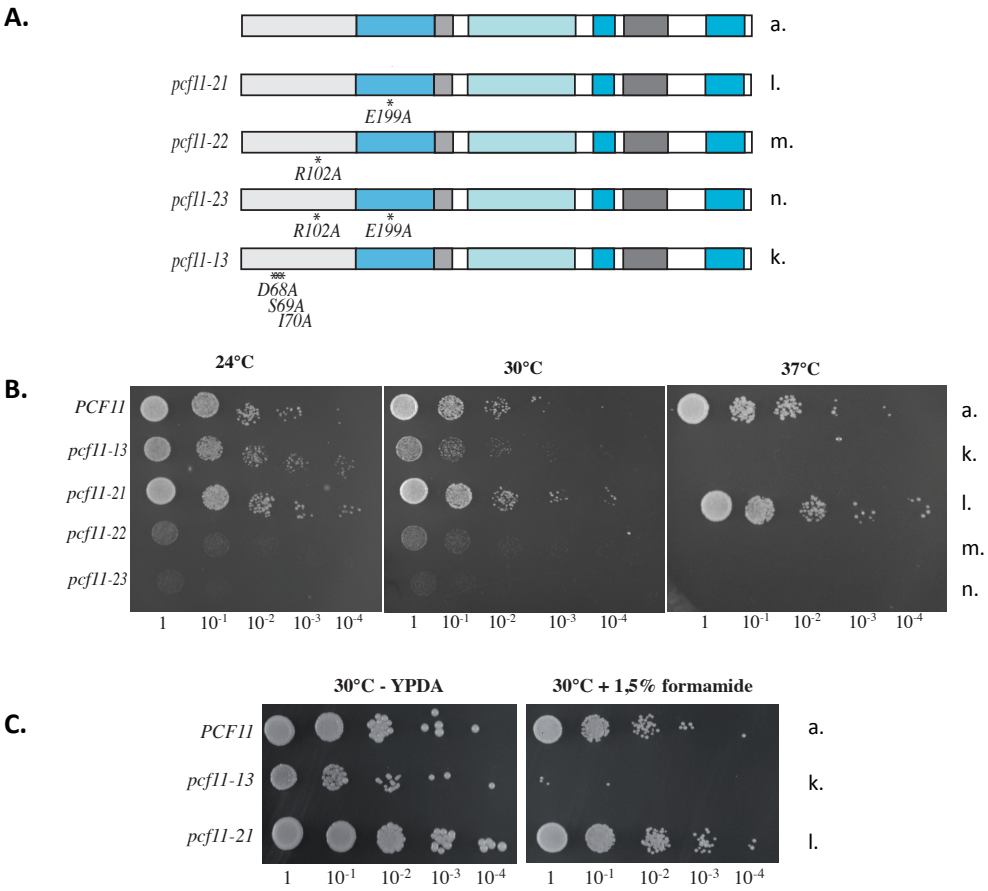


Figure 2.08 : Étude des mutants du 2nd NTD et du CID **(A.)** Rappel de l'organisation des différents mutants. Les différents domaines sont à l'échelle et leur taille est donnée en acides aminés. Les * représentent les mutations ponctuelles ; les résidus mutés sont indiqués en italique. **(B.)** Phénotypes des mutants du CID et du 2nd NTD du milieu YPD(A) à 24°C, 30°C et 37°C. ***a.*** témoin positif ; souche exprimant une version sauvage de Pcf11. ***k.*** *pcf11-13* : contrôle positif ; mutant ponctuel (D68A, S69A, I70A). ***l.*** *pcf11-21* : mutant ponctuel (E199A). ***m.*** *pcf11-22* : mutant ponctuel (R102A). ***n.*** *pcf11-23* : mutant ponctuel (R102A, E199A). **(C.)** Phénotypes des mutants du CID et du 2nd NTD en présence de 1,5% de formamide sur YPD(A) à 30°C. ***a.*** témoin positif ; souche exprimant une version sauvage de *PCF11*. ***k.*** *pcf11-13* : mutant ponctuel (D68A, S69A, I70A). ***l.*** *pcf11-21* : mutant ponctuel (E199A). Les mutants *pcf11-22* et *pcf11-23* présentent un phénotype de croissance lente marqué à 30°C ; le phénotype en présence de formamide n'a pas été étudié.

température à 37°C (**Figure 2.08B**, ligne k) et en présence de 1,5% de formamide dès 30°C (**Figure 2.08C**, ligne k.).

En conclusion, nous disposons de 15 souches, dont trois possédant un phénotype de croissance très affectée - **Δ(141-233)**, **Δ(141-253)** et ***pcf11-22*** (R102A). Les autres souches obtenues ne présentent pas de phénotypes marqués aux températures étudiées. La sensibilité en présence de 2% de formamide suggère cependant que la double mutation ponctuelle (L189D, L213D) du 2nd NTD portée par l'allèle ***pcf11-17*** induit une instabilité du domaine **(141-233)** de la protéine. Les phénotypes de chacune des souches sont récapitulés dans le *TABLEAU 2.1*.

C.1.2. Assemblage du complexe CF IA

Des études précédentes, menées sur Rna14, Rna15 (Moreno-Morcillo et al., 2011) et Clp1 (Haddad et al., 2012; Noble et al., 2007) ont montré que certaines mutations ponctuelles dans l'un des partenaires du complexe CF IA pouvait déstabiliser le complexe dans son ensemble et affecter son assemblage *in vivo,* visible lors de la purification (**Figure 2.09A**). Nous avons cherché à vérifier si les mutations introduites au niveau du 2nd NTD pouvaient avoir ce même effet. Pour cela, nous avons choisi d'utiliser la méthode de purification par *TAP-tagging*, développée par Puig *et al.,* (Puig et al., 2001) et adaptée à notre modèle d'étude au laboratoire. Le complexe CF IA est purifié par deux étapes successives de chromatographie d'affinité, grâce à une étiquette positionnée, dans notre cas, sur Rna15. Si l'intégrité du complexe n'est pas modifiée, Rna15 et ses différents partenaires sont purifiés à l'issue des deux étapes de chromatographie. Cette méthode permet d'obtenir un matériel natif, en quantité suffisante pour le reste de l'étude. L'activité de maturation de l'extrémité 3' des facteurs est testée comme présenté dans le paragraphe C.1.3.

Nous avons construit différentes souches par croisement des souches mutantes décrites précédemment avec la souche YSD12, disponible au laboratoire. Cette souche porte une version étiquetée de Rna15 au niveau du chromosome - [p*NOP1*]-TAP::*RNA15-TRP1*-Kl. Après sporulation, séparation des tétrades et sélection des spores, le phénotype des souches obtenues est vérifié par étalement sur milieu YPD(A). Le complexe CF IA est purifié à partir de 2 litres de culture grâce à un protocole standard développé au laboratoire. Les résultats obtenus sont présentés dans la **Figure 2.09B**. En comparaison des facteurs purifiés à partir

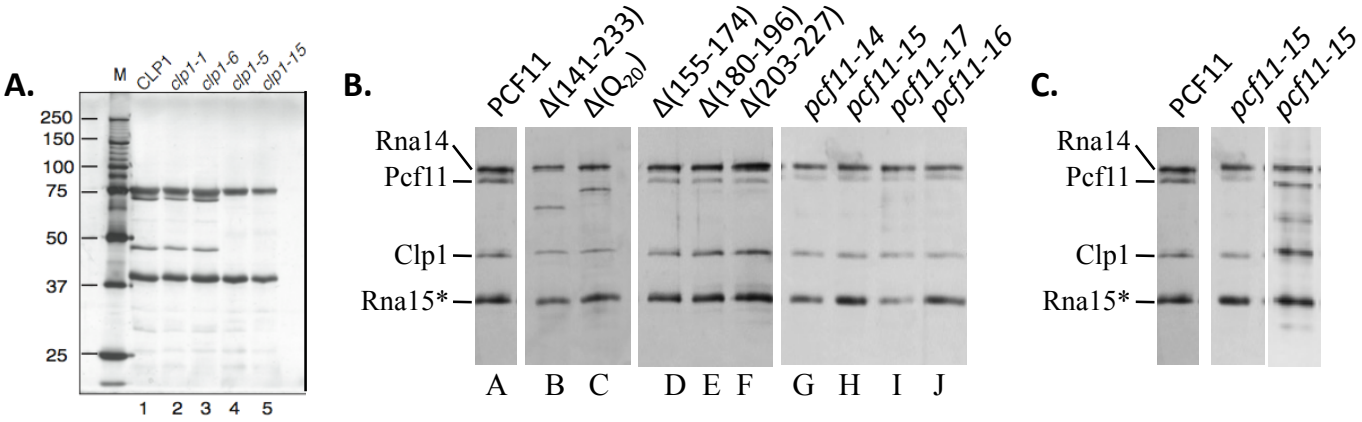


Figure 2.09 : L'assemblage du complexe CF IA n'est pas perturbé par l'insertion des mutations au niveau du 2nd NTD de Pcf11. **(A.)** Purification des facteurs du complexe CF IA par TAP-tagging en présence de différents allèles de *CLP1*. Les facteurs sont séparés par SDS-PAGE et puis colorés au nitrate d'argent. Le poids moléculaire apparent des protéines est estimé par rapport au marqueur de taille (*à gauche*) en kDa. ***1.*** Témoin positif. ***2.*** *clp1-1.* ***3.*** *clp1-6.* ***4.*** *clp1-5.* ***5.*** *clp1-15.* D'après Haddad *et al.,* 2012. **(B.)** Purification des facteurs du complexe CF IA purifiés par TAP-tagging sur la sous-unité Rna15 (*). Les facteurs sont séparés sur gel SDS-PAGE et coloré au nitrate d'argent. ***A.*** témoin positif. ***B.*** pcf11Δ(141-233) : souche délétée pour le 2nd NTD. ***C.*** pcf11ΔQ20 : souche délétée pour le motif Q20. ***D.*** pcf11Δ(155-174) : souche délétée pour l'hélice α1. ***E.*** pcf11Δ(180-196) : souche délétée pour l'hélice α2. ***F.*** pcf11Δ(203-227) : souche délétée pour l'hélice α3. ***G.*** *pcf11-14* : mutant ponctuel (L189D). ***H.*** *pcf11-15 :* mutant ponctuel (L213D). ***I.*** *pcf11-17 :* mutant ponctuel (L189D ; L213D). ***J.*** *pcf11-16 :* mutant ponctuel (D220A). **(C.)** Analyse par SDS-PAGE et coloration au nitrate d'argent des fractions d'élutions obtenues pour deux purifications indépendantes du complexe CF IA en présence de l'allèle *pcf11-15*, en comparaison à un complexe CF IA d'une souche exprimant une version sauvage de *PCF11*.

d'une souche portant une version sauvage de *PCF11*, la délétion du **2nd NTD** (**Δ(141-233)**) ou du domaine **Q_{20}** ne modifie pas l'interaction entre les différents partenaires du complexe (pistes A, B et C) ; Rna15, Rna14 et Clp1 sont visibles, ainsi que Pcf11 aux tailles attendues – 61kDa et 69kDa respectivement. De la même manière, la délétion de l'une ou l'autre des trois hélices formant le 2nd NTD ne modifie pas l'assemblage du complexe (pistes D, E et F). Pour les mutants simples ponctuels, l'interaction entre les différents partenaires n'est pas modifiée (pistes G, H et J). Dans le cas du mutant ***pcf11-15*** (L213D) et en fonction des purifications, nous pouvons observer des bandes intermédiaires correspondant à des produits de dégradation (**Figure 2.09C**) - plus nombreux que pour le complexe CF IA^{PCF11}. Cela pourrait illustrer une instabilité de la protéine, comme proposé par l'étude de la sensibilité au formamide. Cette instabilité est également visible lors de la purification du facteur portant la double mutation (L189D, L213D) - allèle ***pcf11-17*** (**Figure 2.09B**, piste I). L'assemblage du complexe CF IA n'est pas modifié, cependant on observe une coloration de plus faible intensité correspondant à Pcf11 (**Figure 2.10A**). L'analyse par Western blot avec des anticorps dirigés contre les différents partenaires du complexe (**Figure 2.10B**) confirme le résultat obtenu par coloration au nitrate d'argent. Rna14, Rna15 et Clp1 sont détectés en quantité identique à celle obtenue pour le complexe sauvage. On observe une diminution du signal correspondant à Pcf11 dans sa taille entière (72 kDa) et une augmentation des bandes de plus faible poids moléculaire pouvant correspondre à un produit de dégradation - bande correspondant à une taille d'environ 42kDa. Ce fragment majoritaire pourrait correspondre à la partie C-terminale de la protéine (acides aminés 253 à 626) protégée de la dégradation par l'interaction avec Rna14/Rna15 d'une part et de Clp1 d'autre part. L'insertion des deux mutations en position 189 et 213 induit la dégradation ou l'instabilité de la partie N-terminale de la protéine dans les conditions de purifications utilisées mais également *in vivo* comme suggéré par l'étude en présence de 2% de formamide. Cette instabilité peut être causée par la modification des interactions non covalentes qui, à terme, pourraient conduire au dépliement du domaine, favorisant son accessibilité aux protéases. Les résultats obtenus sont en accord avec le choix des mutations ; l'objectif était de modifier l'interaction et d'induire une instabilité au niveau du 2nd NTD.

En conclusion, l'assemblage du complexe n'est pas modifié par la délétion du **2nd NTD** ou d'une partie de ce domaine. Comme suggéré par la sensibilité au formamide, la double mutation dans la souche ***pcf11-17*** (L189D, L213D) induit une instabilité de la partie

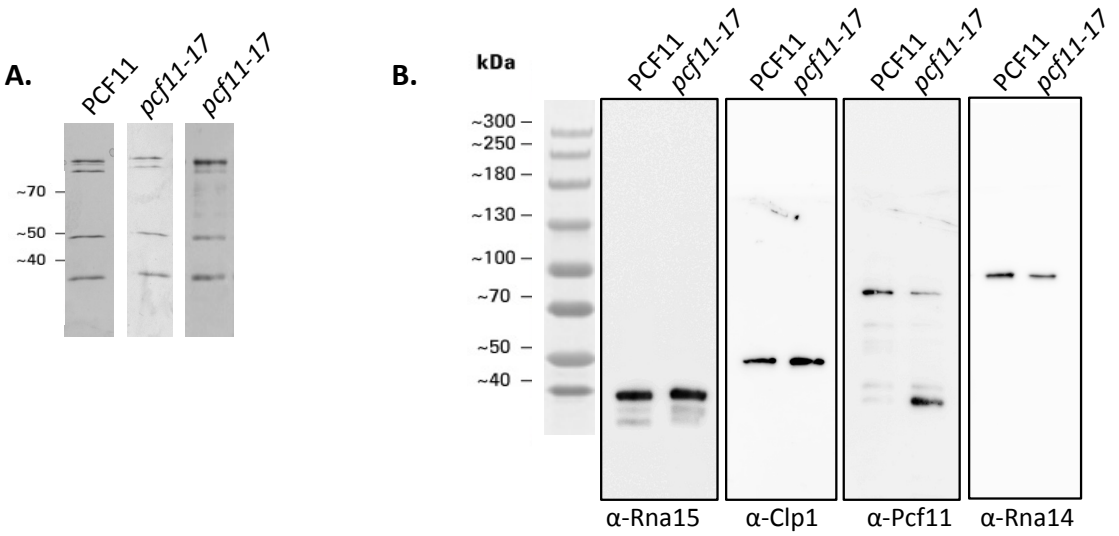


Figure 2.10 : La double mutations *pcf11-17* (L189D, L213D) conduit à une instabilité de la partie N-terminale de Pcf11.

(A.) Analyse par SDS-PAGE et coloration au nitrate d'argent des fractions d'élutions obtenues pour deux purifications indépendantes du complexe CF IA en présence de l'allèle *pcf11-17*, en comparaison à un complexe CF IA d'une souche exprimant une version sauvage de *PCF11*. Le poids moléculaire apparent des protéines est estimé par rapport au marqueur de taille (*à gauche*) en kDa.

(B.) Analyse par Western Blot des différentes sous-unités du complexe CF IA d'une souche exprimant une version sauvage de *PCF11* (*à gauche*) et d'une souche exprimant l'allèle *pcf11-17* (*à droite*). Le poids moléculaire apparent des protéines est estimé par rapport au marqueur de taille (*à gauche*) en kDa. Les anticorps utilisés sont indiqués ; la quantité de protéines analysée est normalisée par rapport à Rna15.

N-terminale de la protéine (*TABLEAU 2.1*). Pour la suite de cette étude, nous avons décidé de ne pas utiliser toutes les souches obtenues précédemment ; les mutants ponctuels ***pcf11-14*** (L189D), ***pcf11-15*** (L213D) et ***pcf11-16*** (D220A) ne présentent pas de phénotype marqué et l'assemblage du complexe est intact. Ils seront donc écartés du reste de l'étude. L'étude fonctionnelle est réalisée sur les mutants suivants :

- Délétion du 2nd NTD - **Δ(141-233)**
- Délétion du motif **Q_{20} - Δ(Q_{20})**
- Délétion de l'hélice α1 - **Δ(155-174)**
- Délétion de l'hélice α2 - **Δ(180-196)**
- Délétion de l'hélice α3 - **Δ(203-227)**
- Double mutation ponctuelle L189D et L213D - ***pcf11-17***

C.1.3. Test d'activité de maturation de l'extrémité 3' des ARN messagers

Certaines études (Ghazy et al., 2012; Grzechnik et al., 2015; Noble et al., 2007; Sadowski et al., 2003) ont montré que Pcf11 est une protéine bifonctionnelle : la partie N-terminale est impliquée dans la terminaison de la transcription par l'ARN polymérase II ; la partie C-terminale est impliquée dans la maturation de l'extrémité 3' des ARN messagers. Selon cette description, il semblerait plus vraisemblable que le 2nd NTD soit impliqué dans la terminaison de transcription. Cependant, nous nous sommes tout de même intéressés au rôle du 2nd NTD dans la maturation de l'extrémité 3' des ARN messagers.

Les activités de clivage et polyadénylation peuvent être étudiées *in vitro* en utilisant des extraits protéiques ou en utilisant des facteurs purifiés par *TAP-tagging* en présence des différentes protéines nécessaires à la réaction – Hrp1 et Pab1 (Amrani et al., 1997a; Kessler et al., 1997). Nous avons décidé d'utiliser cette dernière approche et co-purifier les complexes CF IA et CPF par *TAP-tagging*, étiquetés respectivement au niveau des sous-unités Rna15 et Fip1. La co-purification des deux complexes [CF IA+CPF] à partir d'une même souche permet d'éviter la contamination du complexe CPF par des facteurs du complexe CF IA endogènes. Le complexe CPF est étroitement lié au complexe CF IA, par interaction avec les sous-unités Rna14, Pcf11 et Clp1. Malgré une forte stringence (400mM NaCl) lors de la purification, l'interaction entre les facteurs ne peut être abolie, et une activité de clivage est observée pour le complexe CPF seul, dû à la contamination par des complexes CF IA. La co-purification des complexes [CF IA+CPF] à partir d'une même souche permet de s'assurer que

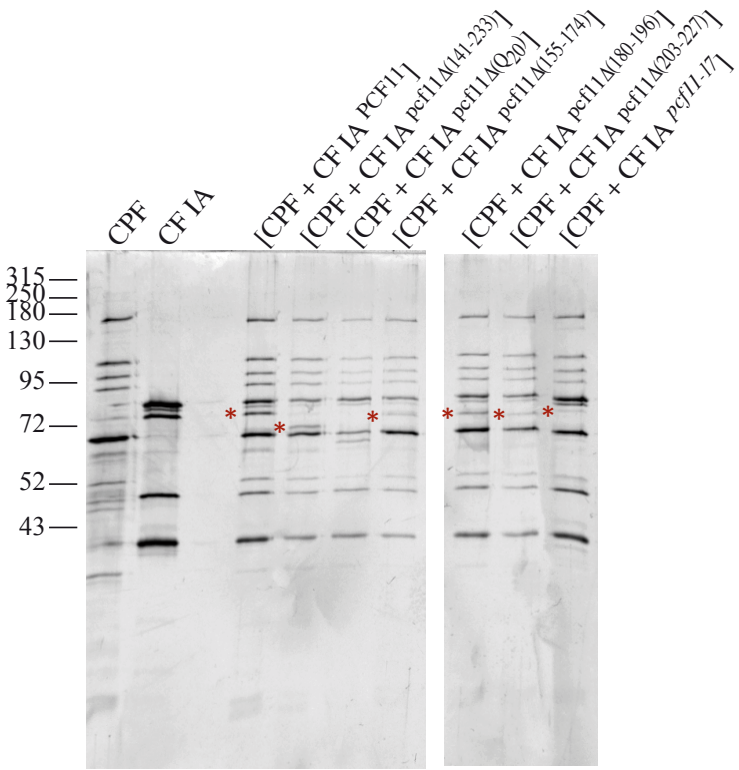


Figure 2.11 : Double purification par tap-tagging des complexes [CF IA+CPF] pour les différents mutants du 2nd NTD *in vitro*. Analyse par SDS-PAGE et coloration au nitrate d'argent des complexes [CF IA+CPF] utilisés dans les expériences de reconstitution des activités de maturation en 3'. Les facteurs sont purifiés par TAP-tagging sur les sous-unités Rna15 et Rna15/Fip1. Le poids moléculaire apparent des protéines est estimé par rapport au marqueur de taille (*à gauche*) en kDa. Les complexes CPF et CF IA sauvages sont utilisés comme contrôles. Les différentes versions de *PCF11* sont indiquées par *.

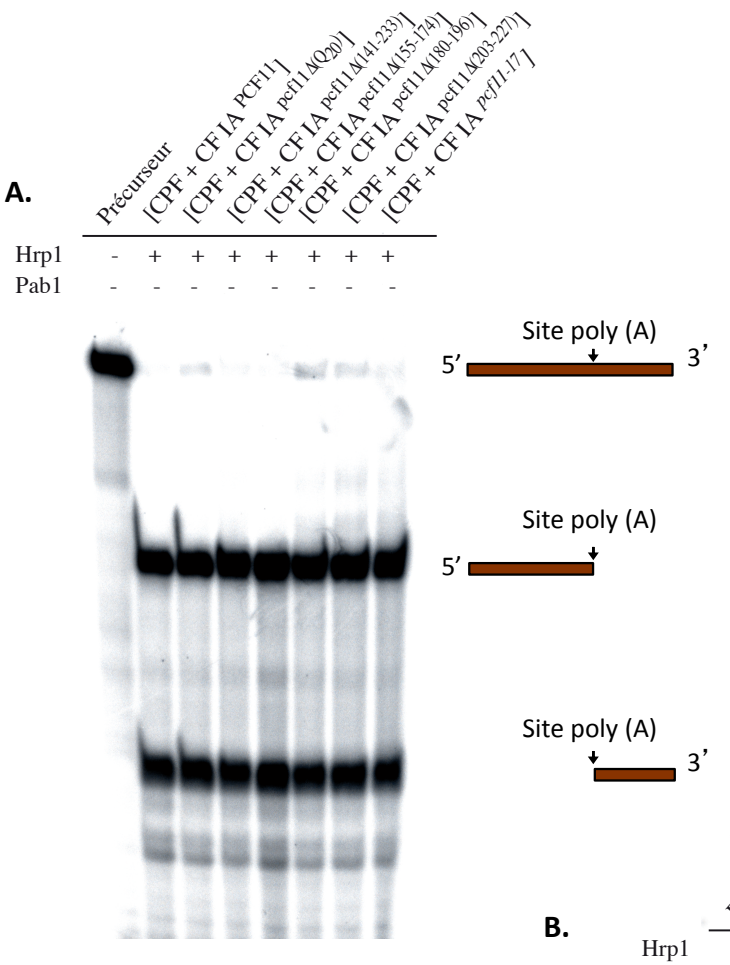
les activités observées sont strictement représentatives des mutations étudiées. Nous avons construit différentes souches par croisement des souches mutantes décrites précédemment avec la souche YLM39-3 ([p*NOP1*]-TAP::*RNA15-URA3*, [p*NOP1*]-TAP::*FIP1-TRP1*) et la souche YLM225 ([p*NOP1*]-TAP::*FIP1-URA*), disponible au laboratoire. Après sporulation, séparation des tétrades et sélection des spores, le phénotype des souches obtenues est vérifié par étalement sur milieu YPD(A). Les facteurs sont purifiés par double affinité et leur pureté est vérifiée par migration sur gel d'électrophorèse en conditions dénaturantes (SDS-PAGE) et coloration au nitrate d'argent (**Figure 2.11**).

Les activités de maturation de l'extrémité 3' sont étudiées en présence d'un ARN substrat *CYC1* radio-marqué (Butler and Platt, 1988). Cet ARN substrat correspond à la région 3' non traduite de l'ARN du gène *CYC1*, codant pour l'iso-1-cytochrome C. L'activité de clivage peut être découplée de celle de la polyadénylation en présence de 2mM d'EDTA ; l'activité de poly(A) polymérase de Pap1 nécessitant la présence d'ion Mg^{2+} (Hirose and Manley, 1997). En contexte sauvage, l'ARN est clivé au site naturel de polyadénylation de l'ARN CYC1. Cela permet la formation d'un fragment 5' de 186 nucléotides et un fragment 3' de 116 nucléotides. Les produits de la réaction sont analysés sur un gel acrylamide en conditions dénaturantes.

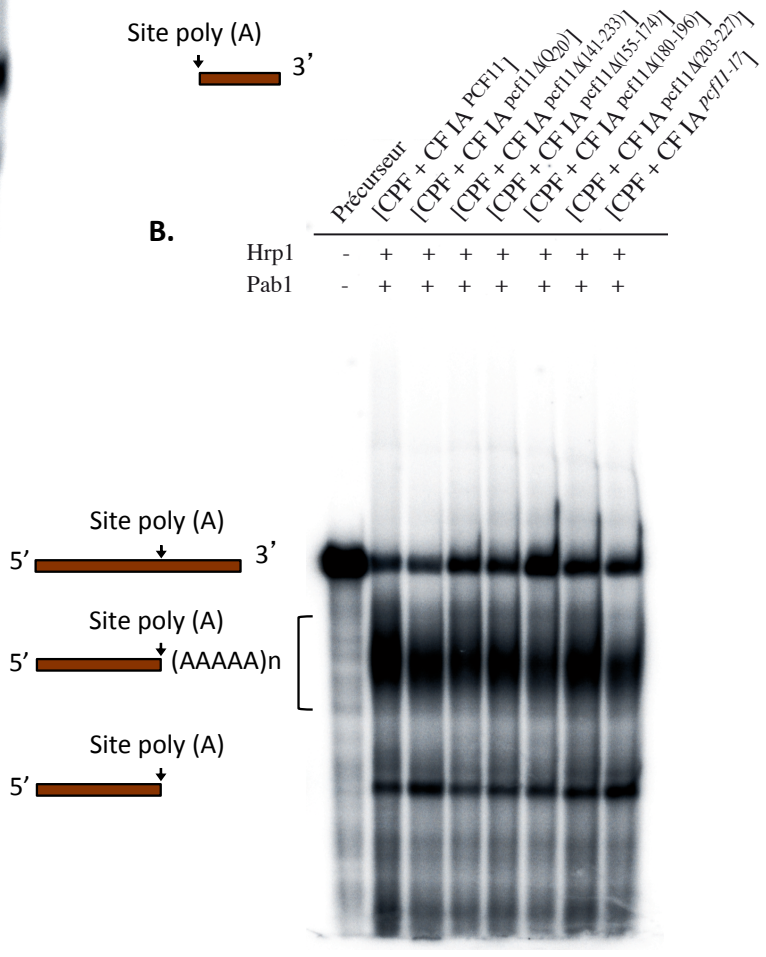
À 30°C, les facteurs des complexes [CF IA+CPF] des souches testées présentant un phénotype de croissance sauvage (motif **Q_{20}**, hélice **α1**, hélice **α2**, hélice **α3** et ***pcf11-17*** (L189D, L213D)) clivent l'ARN précurseur avec la même efficacité en comparaison avec la souche sauvage (**Figure 2.12A**). Les facteurs de la souche délétée pour le **2nd NTD** (**Δ(141-233))** présentent également une activité de clivage similaire à celle des facteurs des complexes [CPF+CF IA^{PCF11}]. Les activités de clivage et de polyadénylation sont ensuite testées sur le même ARN précurseur en absence d'EDTA et en présence de 2mM d'acétate de magnésium (Minvielle-Sebastia et al., 1994). Pour l'ensemble des mutants testés, à 30°C, l'ARN substrat est clivé puis le fragment 5' est polyadénylé – correspondant à l'ajout d'environ 70 résidus adénosines - avec la même efficacité en comparaison des facteurs des complexes [CPF+CF IA^{PCF11}] (**Figure 2.12B**).

En conclusion, *in vitro*, les activités de maturation de l'extrémité 3' des ARNs messagers, aussi bien de clivage et de polyadénylation, ne sont pas modifiées par les délétions ou les mutations du **2nd NTD** (*TABLEAU 2.1*). Ce domaine n'est pas nécessaire pour cette étape ce qui est en accord avec les résultats obtenus précédemment, caractérisant uniquement la

Figure 2.12 : Le 2nd NTD et le motif Q_{20} ne sont pas impliqués dans la maturation de l'extrémité 3' des ARN messagers.



(A.) Essai de reconstitution des activités de clivage en présence de l'ARN CYC1 radiomarqué. Les produits de la réaction sont séparés sur gel de polyacrylamide 6% (acrylamide/bisacrylamide 19 :1) / TBE 1X / 8,3M urée et révélés par autoradiographie. L'organisation des fragments obtenus est présentée *à droite*. L'ajout (+) ou l'absence (-) des protéines nécessaires à la réaction sont indiquées au dessus de chaque ligne. Hrp1 : 50ng.



(B.) Essai de reconstitution des activités de clivage et de polyadénylation en présence de l'ARN CYC1 radiomarqué. Les produits de la réaction sont séparés sur gel de polyacrylamide 6% (acrylamide/bisacrylamide 19 :1) / TBE 1X / 8,3M urée et révélés par autoradiographie. L'organisation des fragments obtenus est présentée *à gauche*. L'ajout (+) ou non (-) des protéines nécessaires à la réaction sont indiquées au dessus de chaque ligne. Hrp1 : 50ng. Pab1 : 140ng.

partie C-terminale de la protéine (des acides aminés 250 à 626) comme nécessaire et suffisante pour la maturation de l'extrémité 3'.

C.1.4. Terminaison de la transcription par l'ARN polymérase II

Nous nous sommes ensuite intéressé à la terminaison de la transcription, étroitement liée à la partie N-terminale de la protéine. Les études récentes ont permis de mettre en évidence le rôle majeur d'un très grand nombre de composants de la machinerie de polyadénylation dans la terminaison de la transcription de certains ARN messagers et des snoRNAs – (Birse et al., 1998), Pta1 (Ghazy et al., 2009), Pti1 et Ref2 (Dheur et al., 2003), Ssu72 (Steinmetz and Brow, 2003), Pcf11 (Grzechnik et al., 2015), Rna14 (Fatica et al., 2000), Rna15 (Morlando et al., 2002). En 2003, Sadowski *et al.,* (Sadowski et al., 2003) ont décrit la partie N-terminale de Pcf11 – des acides aminés 1 à 180 – comme nécessaire et suffisante pour la terminaison de transcription à partir d'analyses run-on sur la région 3'UTR de *CYC1*. Les analyses récentes (Grzechnik et al., 2015; Lunde et al., 2010) confirment l'implication de Pcf11 dans la terminaison de transcription des ARNs codants et non codants. Nous avons cherché à déterminer si le 2nd NTD situé en aval du CID était nécessaire pour cette étape. Deux approches ont été utilisées : le détection par Northern blot et des essais de protection contre la digestion par la S1 nucléase. Pour cela nous avons extrait les ARNs totaux des différents mutants étudiés, après 0, 30, 60 et 120 minutes à 37°C.

C.1.4.1. Northern blot

Nous avons d'abord étudié l'implication de ce domaine par Northern blot. Deux gènes ont été choisis pour cible :

- ***CYH2*** – gène contenant un intron, codant pour la protéine L28 de la sous-unité 60S du ribosome (**Figure 2.13A**).
- ***TRS31*** – codant pour un composant du complexe de transport TRAPP ; situé en aval de **snR13**, snoRNA à boîte C/D (**Figure 2.13D**). Ils sont organisés en tandem, dans le même sens de transcription.

L'épissage au niveau de ***CYH2*** est assez lent comparativement à d'autres gènes, ce qui permet de détecter, par Northern blot, le pré-ARNm non épissé (Swida et al., 1988). Un défaut de terminaison de transcription se traduit, au niveau de ***CYH2***, par l'accumulation de bandes dites extérieures (Holbein et al., 2011) correspondant à la terminaison au niveau des

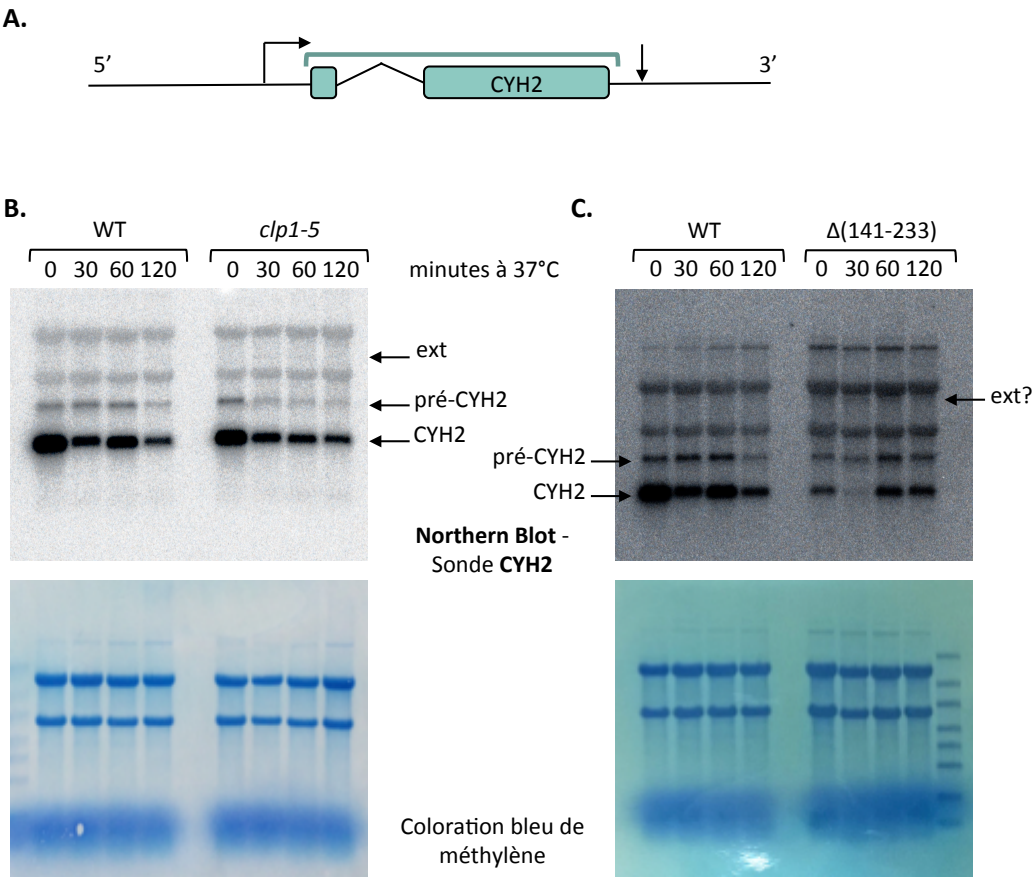


Figure 2.13 : Étude du rôle du 2nd NTD dans la terminaison de transcription par Northern Blot.
(A.) Représentation du gène *CYH2*. La région détectée est encadrée en **bleu**.
(B.) Résultats préliminaires obtenus pour les souches portant une version sauvage de *PCF11* et l'allèle *clp1-5*. 5µg d'ARNs totaux sont analysés après 0, 30, 60 et 120 minutes à 37°C. Northern blot dirigé contre *CYH2* (*haut*). Les fragments correspondant à l'ARNm mature (CYH2 – 450 nucléotides) et au pré-ARNm (pré-CYH2 – 961 nucléotides) sont indiqués. Les bandes extérieures, correspondant à des ARNs de grandes tailles illustrant un défaut de terminaison de transcription sont indiquées par « *ext* ». Révélation après 4 jours d'exposition. La coloration de la membrane au bleu de méthylène est utilisée comme contrôle de charge (*bas*).
(C.) Northern blot dirigé contre *CYH2* (*haut*), après modification du protocole d'hybridation, pour les souches portant une version sauvage de *PCF11* et délétée pour le 2nd NTD Δ(141-233). 5µg d'ARNs totaux sont analysés après 0, 30, 60 et 120 minutes à 37°C. Les fragments correspondant à l'ARNm mature (CYH2 – 450 nucléotides) et au pré-ARNm (pré-CYH2 – 961 nucléotides) sont indiqués. Les bandes extérieures sont indiquées par « ext ». Révélation après 10 jours d'exposition. La coloration de la membrane au bleu de méthylène est utilisée comme contrôle de charge (*bas*). **(D.)** Représentation schématique de snR13 (en **bleu**) et de *TRS31* (en **rose**). La région détectée est encadrée en **rose**.
(E.) Résultats préliminaires obtenus pour les souches portant une version sauvage de *PCF11* et l'allèle *clp1-5*. 5µg d'ARNs totaux sont analysés après 0, 30, 60 et 120 minutes à 37°C et séparés sur un gel agarose 1,2%. Northern blot dirigé contre *TRS31* (*haut*). Le fragment correspondant à TRS31 (852 nucléotides) est indiqué par une flèche. Révélation après 48 heures d'exposition. La coloration de la membrane au bleu de méthylène est utilisée comme contrôle de charge (*bas*).

régions situées en aval du gène. Nous avons tout d'abord réalisé la mise au point sur les ARNs des souches sauvages et portant l'allèle ***clp1-15*** (YRH37) puisque Clp1 a été montrée comme influant sur la terminaison de transcription des ARNs codants et non codants de types snoRNAs (Haddad et al., 2012; Holbein et al., 2011). Les résultats préliminaires sont présentés dans la **Figure 2.13B**. L'ARN messager mature (*CYH2*) ainsi que le pré-ARN messager (pré-*CYH2*) sont détectés aux tailles attendues et en quantité décroissante après incubation à 37°C. Cependant les ARNs ribosomiques sont également fortement détectés au niveau de la zone de migration des bandes extérieures, correspondant à des ARNs de grandes tailles attendus en cas de défaut de terminaison. Nous avons essayé de réduire ce bruit de fond en modifiant différentes étapes du protocole :

- les températures et le temps d'hybridation. Les sondes sont obtenues par « *random priming* » ce qui permet d'obtenir des sondes très concentrées mais qui peut conduire à plus de bruit de fond. L'augmentation de la température permet de favoriser uniquement les hybridations spécifiques.
- augmentation de la stringence des lavages – augmentation de la température et augmentation de la quantité de SDS en fin de lavage de 0,1 à 0,5%.
- modification de la composition du tampon d'hybridation. Les premiers essais ont été réalisés avec du tampon Church. Nous avons ensuite modifié la composition du tampon d'hybridation - SSC 5X, Denhardt's 5X, SDS 0,5% - et ajouté des ARNt et de l'ADN de sperme de saumon. L'objectif étant de mieux saturer la membrane afin de diminuer la fixation des sondes au niveau des ARNs ribosomiques.
- modification de la membrane utilisée (Nylon N+ et XL).

Nous avons réussi à diminuer légèrement le bruit de fond mais les ARNs ribosomiques sont toujours détectés et restent visibles après révélation. Un défaut de terminaison de transcription peut tout de même être observé pour la souche *clp1-5* mais pas pour la souche délétée pour le 2nd NTD (**Figure 2.13B et C**). Comparativement à la quantité d'ARNs ribosomique, l'ARN mature (CYH2) et le pré-ARNm (pré-CYH2) sont détectés en plus faible quantité dans le mutant **Δ(141-233)** par rapport à la souche sauvage. Cela suggère un défaut dans l'une des étapes permettant la formation des ARN messagers matures. La détection de manière stable du pré-CYH2 permet d'éliminer l'hypothèse d'un défaut dans l'épissage du précurseur. De plus, les études précédentes ont montré que le 2nd NTD n'induisait pas de modifications de la maturation de l'extrémité 3'. Nous pouvons ainsi suggérer que le 2nd NTD est impliqué dans la terminaison de transcription de cet ARNm. Cependant, l'analyse par

Northern blot ne permet pas de valider cette hypothèse, la détection des ARNs ribosomiques ne permet pas une résolution optimale de cette zone. À court terme, nous souhaitons éliminer les ARNs ribosomiques, par le biais d'un kit (RiboMinus™ - Fisher Scientific), ce qui nous permettrait d'éliminer le bruit de fond et d'obtenir une meilleure résolution afin de valider le rôle du 2nd NTD.

L'étude de ***TRS31*** permet d'analyser la terminaison de transcription d'un snoRNA situé en amont. De la même manière que pour l'étude de *CYH2*, nous avons rencontré des difficultés liées à la détection importante des ARNs ribosomiques (**Figure 2.13E**). Le bruit de fond n'a pas pu être totalement éliminé (**Figure 2.13F**) mais nous avons pu nous affranchir de cette contrainte en augmentant la résolution du gel (2% **Figure 2.13E** contre 1,2% **Figure 2.13F**) ; les produits migrants à des tailles plus petites que les ARNs ribosomiques. **TRS31** est détecté en très faible quantité dans la souche sauvage, de même que dans la souche témoin portant l'allèle ***clp1-15***. Le défaut de terminaison de transcription de *snR13* peut interférer avec la transcription de *TRS31* qui n'est pas ou peu transcrit (« *transcritpional collision* ») (Prescott and Proudfoot, 2002) et ne permet pas sa détection dans les conditions utilisées. La délétion du **2nd NTD** permet de détecter un transcrit de plus grande taille après 0, 30, 60 ou 120 minutes d'incubation à 37°C - correspondant à un ARN codant pour **snR13** et **TRS31** de 1400 nucléotides. L'intensité de la bande détectée suggère une forte accumulation et une stabilité importante de ce transcrit de plus grande taille. Le 2nd NTD serait donc au moins impliqué dans la terminaison de transcription des snoRNAs, permettant d'expliquer le phénotype marqué observé aux températures étudiées. Le fragment détecté est un fragment de 1400 nucléotides correspondant uniquement à la région snR13/*TRS31*. Un défaut de terminaison de transcription des ARNs codants conduirait à la détection d'un fragment de plus grande taille, voir d'un transcrit instable. Contrairement aux résultats suggérés par l'analyse de *CYH2*, la délétion du 2nd NTD n'induirait pas de défaut de terminaison de transcription de l'ARN codant pour *TRS31*. La détection des ARNs ribosomiques n'exclut cependant pas la présence d'un fragment de plus grande taille, de taille identique à celles des ARNs ribosomiques ou de faible intensité masqué par cette détection aspécifique.

Malgré les problèmes liés à la détection des ARNr, nous avons étudié la terminaison de transcription pour les autres mutants utilisés dans le reste de l'étude. Les résultats sont présentés dans la **Figure 2.13G**. L'ARN correspondant à ***TRS31*** n'est pas détecté dans les

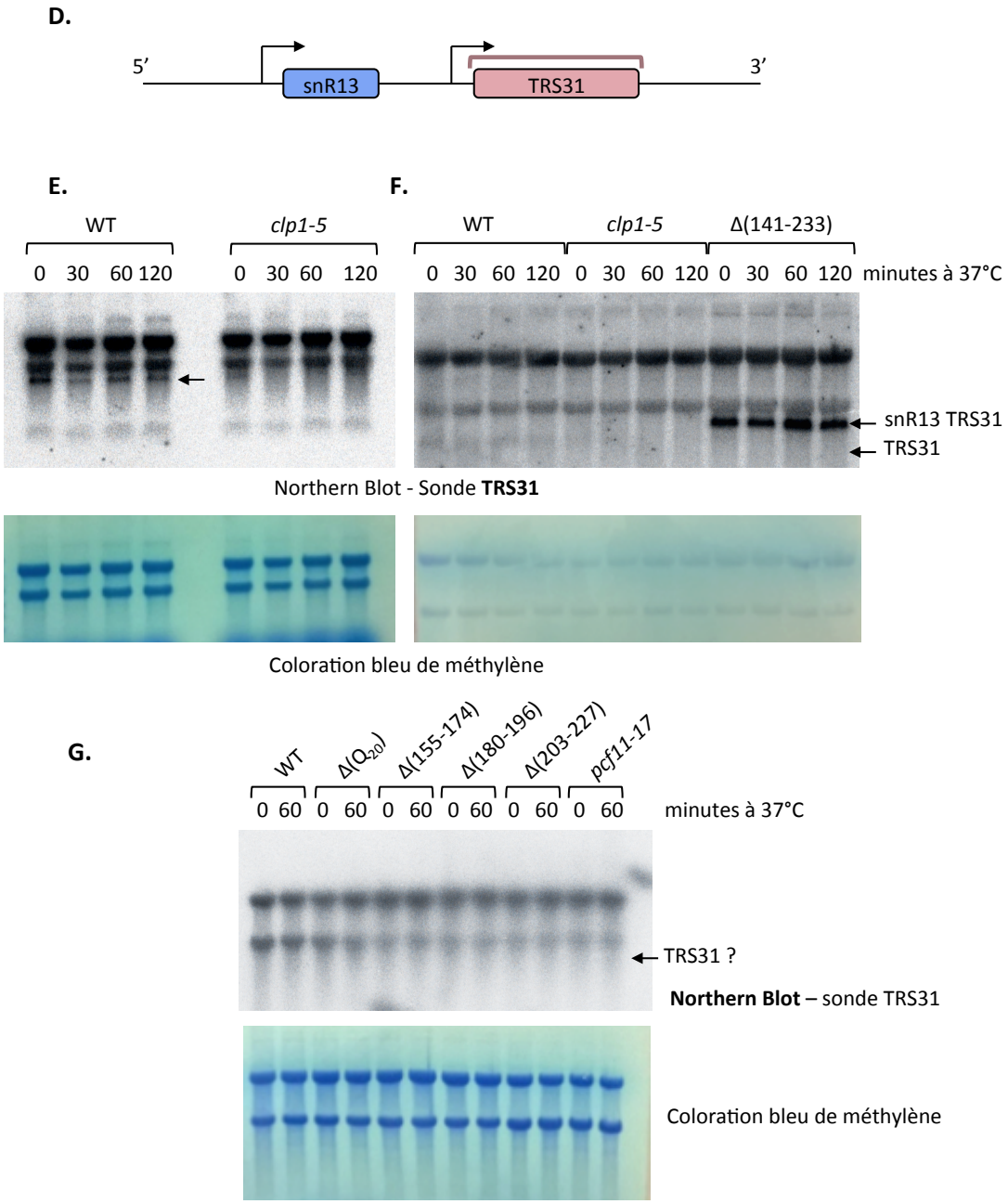


Figure 2.13 : (*Suite*).

(F.) Northern blot dirigé contre *TRS31* (*haut*), après modification du protocole d'hybridation, pour les souches portant une version sauvage de *PCF11*, l'allèle *clp1-5* et délétée pour le 2nd NTD Δ(141-233). 5µg d'ARNs totaux sont analysés après 0, 30, 60 et 120 minutes à 37°C et séparés sur un gel agarose 2%. Le fragment correspondant à TRS31 (852 nucléotides) et celui correspondant à la région snR13/TRS31 (1400 nucléotides) sont indiqués par des flèches. Révélation après 72 heures d'exposition. La coloration de la membrane au bleu de méthylène est utilisée comme contrôle de charge (*bas*)**.**

(G.) Northern blot dirigé contre TRS31 (*haut*), après modification du protocole d'hybridation, pour les autres mutants de Pcf11. ΔQ_{20} : souche délétée pour le motif Q_{20}. Δ(155-174) : souche délétée pour l'hélice α1. Δ(180-196) : souche délétée pour l'hélice α2. Δ(203-227) : souche délétée pour l'hélice α3. *pcf11-17 :* mutant ponctuel (L189D ; L213D). 5µg d'ARNs totaux sont analysés après 0 et 60 minutes à 37°C. Révélation après 48 heures d'exposition. La coloration de la membrane au bleu de méthylène est utilisée comme contrôle de charge (*bas*)**.**

différentes souches. Cependant, l'absence de détection du transcrit correspondant à **snR13-TRS31**, en forte quantité dans la souche **Δ(141-233),** peut suggérer qu'aucun des mutants du domaine $\mathbf{Q_{20}}$, des hélices **α1**, **α2**, **α3** ainsi que ***pcf11-17*** (L189D, L213D) n'induirait de défaut de terminaison au niveau de ce snoRNA. Ces conclusions restent néanmoins hypothétiques et doivent être validées.

En conclusion, malgré une forte détection des ARN ribosomiques, nous avons pu mettre en évidence le rôle majeur du **2nd NTD** dans la terminaison de transcription du snoRNA du type C/D **snR13**. Un défaut de terminaison n'est observé que par la délétion du domaine dans son ensemble, ce qui est en accord avec les phénotypes observés. Seule la délétion du 2nd NTD induit un phénotype de croissance lente marqué. Ces résultats restent à confirmer par Northern blot après élimination des ARNs ribosomiques et par essai de protection contre la digestion par la S1 Nucléase. Les analyses par Northern blot n'ont pas permis de déterminer le rôle du 2nd NTD de Pcf11 dans la terminaison des ARNs codants.

C.1.4.2. Essai de protection contre la digestion par la S1 Nucléase

Afin de valider les résultats préliminaires obtenus par Northern blot, nous avons choisi d'analyser la terminaison de transcription par un essai de protection contre la digestion par la S1 Nucléase, développé notamment pour l'étude des ARNm codants pour les histones chez les métazoaires (Lanzotti et al., 2002). Nous avons tout d'abord choisi de nous intéresser à **snR13**.

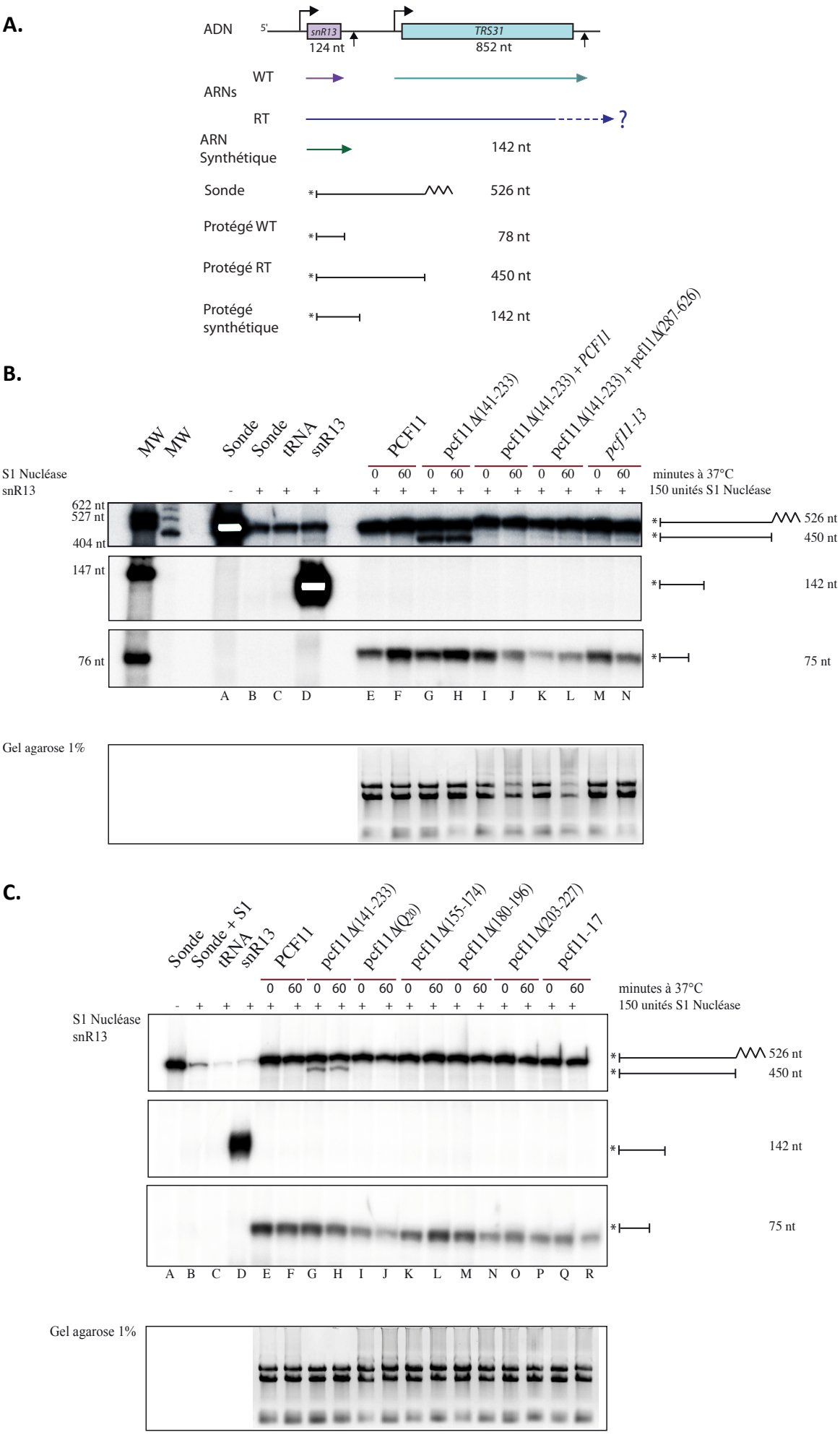
Cet essai est fondé sur la formation d'hybrides ARN-ADN entre les ARNs totaux extraits des différentes souches et une sonde ADN radiomarquée à l'extrémité 5' qui encadre la région à étudier, suivie d'une séquence non spécifique d'environ 70 nucléotides. Dans notre cas, la région à étudier comprend : snR13, la région intergénique et une partie de *TRS31* (**Figure 2.14A**). Une fois ces hybrides formés, la S1 nucléase dégrade spécifiquement les acides nucléiques ADN ou ARN simple brin, permettant de détecter les régions double brin, protégées de la dégradation. Dans le cas d'une souche sauvage, la sonde, marquée à l'extrémité 5' est hybridée uniquement avec l'ARN correspondant au snoRNA mature, et un fragment protégé de petite taille sera visible après migration. Un défaut de terminaison est caractérisé par un défaut de décrochage de l'ARN polymérase II et conduit à la synthèse de transcrits de plus grande taille ; et la formation d'hybride avec la totalité de partie spécifique de la sonde ADN. Le fragment protégé après action de la S1 nucléase aura une taille proche

Figure 2.14 : Étude du rôle du 2nd NTD dans la terminaison de transcription par essai de protection contre la digestion par la S1 Nucléase – snR13 boîte C/D.

(A.) Représentation schématique de la zone étudiée. snR13 (124 nucléotides) est représenté en **violet**, TRS31 est représenté en **bleu** (852 nucléotides). Les ARNs de plus grande taille illustrant un défaut de terminaison de transcription sont indiqués en **bleu**. L'ARN synthétique utilisé comme témoin positif est représenté en **vert**. Le marquage de la sonde à l'extrémité 5' est représenté par un * ; la séquence non spécifique est représenté par des zigzags.

(B.) 30ng de sondes radiomarquées à l'extrémité 5' sont incubés seule (ligne *A*), ou en présence de 150 unités de S1 Nucléase pendant 1h et des ARNs étudiés (ligne *B à R*). *B* : sonde seule. *C* : 2µg de ARNt de levure. *D* : 2µg de transcrit synthétique snR13. L'analyse est effectuée sur 30µg d'ARNs totaux après 0 et 60 minutes d'incubation à 37°C. Les fragments protégés de la digestion par la S1 Nucléase sont séparés sur gel de polyacrylamide 8% et révélés par autoradiographie. Les fragments obtenus sont représentés *à droite*. *E et F* : ARNs totaux extraits d'une souches portant une version sauvage de *PCF11*. *G et H* : ARNs totaux extraits de la souche délétée pour le 2nd NTD Δ(141-233). *I et J* : ARNs totaux extraits de la souche délétée pour le motif Q_{20}. *K et L* : ARNs totaux extraits de la souche délétée pour l'hélice α1 Δ(155-174). *M et N* : ARNs totaux extraits de la souche délétée pour l'hélice α2 Δ(180-196). *O et P* : ARNs totaux extraits de la souche délétée pour l'hélice α3 Δ(203-227). *Q et R* : ARNs totaux extraits portant l'allèle *pcf11-17* (L189D, L213D). *MW* : marqueurs de taille en nucléotides. Les carrés blancs correspondent aux zones saturées.

(C.) *B à N* : ajout de 150 unités de S1 Nucléase pendant 1h. *B* : sonde seule. *C* : 2µg de ARNt de levure. *D* : 2µg de transcrit synthétique snR13. L'analyse est effectuée sur 30µg d'ARNs totaux après 0 et 60 minutes d'incubation à 37°C. Les fragments protégés de la digestion par la S1 Nucléase sont séparés sur gel de polyacrylamide 8% et révélés par autoradiographie. Les fragments obtenus sont représentés *à droite. E et F* : ARNs totaux extraits d'une souches portant une version sauvage de *PCF11*. *G et H* : ARNs totaux extraits de la souche délétée pour le 2nd NTD Δ(141-233). *I et J* : ARNs totaux extraits de la souche délétée pour le 2nd NTD Δ(141-233) complémenté par la protéine entière. *K et L* : ARNs totaux extraits de la souche délétée pour le 2nd NTD Δ(141-233) complémenté par la partie 1-287 de la protéine. *M et N* : ARNs totaux extraits de la souche portant l'allèle *pcf11-13* (D68A, S69A, I70A).

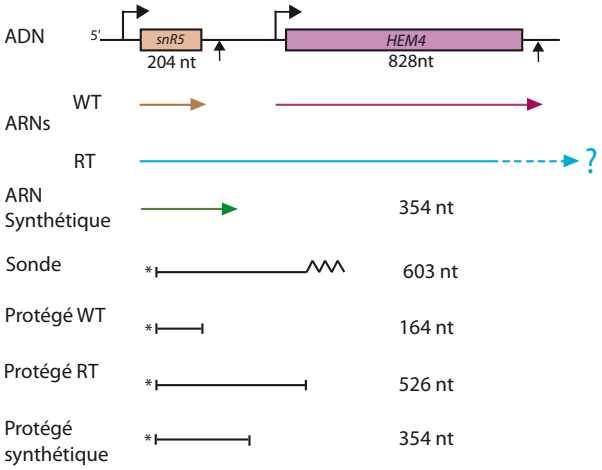

A.
ADN
5'
snR13
TRS31
124 nt
852 nt
WT
ARNs
RT
?
ARN Synthétique
142 nt
Sonde
526 nt
Protégé WT
78 nt
Protégé RT
450 nt
Protégé synthétique
142 nt
B.
MW
MW
Sonde
Sonde
tRNA
snR13
PCF11
pcf11Δ(141-233)
pcf11Δ(141-233) + PCF11
pcf11Δ(141-233) + pcf11Δ(287-626)
pcf11-13
S1 Nucléase
snR13
0 60 0 60 0 60 0 60 0 60
minutes à 37°C
- + + + + + + + + + + + + + + + +
150 unités S1 Nucléase
622 nt
527 nt
404 nt
147 nt
76 nt
526 nt
450 nt
142 nt
75 nt
A B C D E F G H I J K L M N
Gel agarose 1%
C.
Sonde
Sonde + S1
tRNA
snR13
PCF11
pcf11Δ(141-233)
pcf11Δ(Q20)
pcf11Δ(155-174)
pcf11Δ(180-196)
pcf11Δ(203-227)
pcf11-17
0 60 0 60 0 60 0 60 0 60 0 60 0 60 0 60
minutes à 37°C
- + + + + + + + + + + + + + + + +
150 unités S1 Nucléase
S1 Nucléase
snR13
526 nt
450 nt
142 nt
75 nt
A B C D E F G H I J K L M N O P Q R
Gel agarose 1%

de celle de la sonde ; la S1 nucléase permettant la dégradation de la séquence non spécifique de la sonde, d'environ 70 nucléotides.

Les résultats obtenus pour les différents mutants sont présentés dans la **Figure 2.14B**. L'analyse est effectuée sur 30µg d'ARNs totaux ; l'équivalent de 5µg d'ARN totaux est déposé sur gel d'agarose 1%, pour contrôler les quantités utilisées et pour normaliser les résultats obtenus. Les ARNs de la souche portant une version sauvage de *PCF11* (lignes E et F) permettent la détection de deux signaux, après 0 et 60 minutes d'incubation à 37°C : un premier (526 nucléotides) correspond à la sonde, présente en excès dans l'essai ; la seconde à un signal migrant à environ 75 nucléotides, et qui correspond au fragment protégé par l'hybridation avec snR13. Pour les ARNs extraits de la souche délétée pour le 2nd NTD de Pcf11 (**Δ(141-233)**), plusieurs bandes sont visibles après 0 et 60 minutes d'incubation à 37°C (lignes G et H) : une correspondant à la sonde en excès (526 nucléotides), une correspondant au fragment protégé par l'hybridation avec snR13 (75 nucléotides) et une dernière, plus petite que la sonde, de 450 nucléotides. Cette dernière bande correspond à l'hybridation de l'ensemble de la partie spécifique de la sonde avec un transcrit correspondant à la région entière étudiée. Ces résultats illustrent la présence de deux populations d'ARNs dans la souche délétée pour le 2nd NTD : une présentant une terminaison de transcription normale, et une population de taille plus grande illustrant une transcription aberrante, due à un défaut de terminaison. La présence de ces deux populations de snoRNAs sera analysée dans la discussion.

Nous avons testé les différentes souches mutantes obtenues précédemment par complémentation avec le gène codant pour la protéine entière (**YLM242** - lignes I et J) ou par la partie N-terminale de la protéine (1 à 287 - **YLM243** - lignes K et L) et qui permettent de restaurer un phénotype sauvage (**Figure 2.14B**). Deux bandes sont observées : la sonde et un fragment protégé de petite taille ; une bande de très faible intensité correspondant à un fragment protégé de grande taille peut également être visible. La complémentation avec la protéine entière mais aussi avec la partie N-terminale de Pcf11 permet ainsi de restaurer une efficacité de terminaison de transcription proche de celle observée pour la souche sauvage. Nous avons également voulu tester la souche ***pcf11-13*** (lignes M et N) comprenant trois mutations dans le CID (D68A, S69A, I70A), et conduisant à un défaut de terminaison de transcription (Sadowski et al., 2003). De manière surprenante, pour une même quantité d'ARNs totaux, le fragment protégé de grande taille, illustrant un défaut de terminaison de

A.



B.

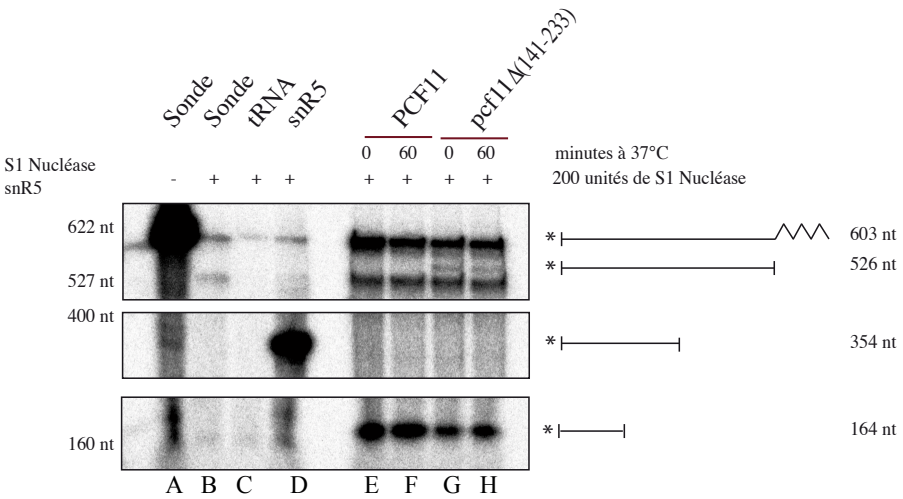


Figure 2.15 : Étude du rôle du 2nd NTD dans la terminaison de transcription par essai de protection contre la digestion par la S1 Nucléase – snR5 boîte H/ACA.

(A.) Représentation schématique de la zone étudiée en essai de protection contre la S1 Nucléase. snR5 (204 nucléotides) est représenté en **marron**, TRS31 est représenté en **rose** (828 nucléotides). Les ARNs de plus grande taille illustrant un défaut de terminaison de transcription sont indiqués en **bleu**. L'ARN synthétique utilisé comme témoin positif est représenté en **vert.** Le marquage de la sonde à l'extrémité 5' est représenté par un * ; la séquence non spécifique est représenté par des zigzags.

(B.) 30ng de sondes radiomarquées à l'extrémité 5' sont incubés seule (ligne *A*), ou en présence de 200 unités de S1 Nucléase pendant 1h30 et des ARNs étudiés (ligne *B à H*). *B* : sonde seule. *C* : 2µg de ARNt de levure. *D* : 1µg de transcrit synthétique snR5. L'analyse est effectuée sur 30µg d'ARNs totaux après 0 et 60 minutes d'incubation à 37°C. Les fragments protégés de la digestion par la S1 Nucléase sont séparés sur gel de polyacrylamide 8% et révélés par autoradiographie. Les fragments obtenus sont représentés *à droite*. *E et F* : ARNs totaux extraits d'une souches portant une version sauvage de *PCF11*. *G et H* : ARNs totaux extraits de la souche délétée pour le 2nd NTD Δ(141-233).

transcription, est détecté en plus faible quantité, en comparaison avec la souche **Δ(141-233)**. Cette différence sera discutée plus loin dans ce chapitre.

Nous avons ensuite testé les autres mutants de *PCF11* obtenus. Les résultats sont présentés dans la **Figure 2.14C**. La délétion du motif $\mathbf{Q_{20}}$, (lignes I et J), de l'hélice **α1** (lignes K et L), **α2** (lignes M et N), **α3** (lignes O et P) ou la double mutation (L189D, L213D) de l'allèle ***pcf11-17*** (lignes Q et R) ne modifient pas le profil obtenu par rapport à une souche sauvage : deux bandes sont observées, la sonde et un fragment protégé de 78 nucléotides. Ainsi nous pouvons conclure que ces mutants ne présentent pas de défaut de terminaison de transcription pour l'unité de transcription testée, **snR13** et dans les conditions testées.

Nous nous sommes ensuite intéressé à **snR5** (204 nucléotides), snoRNA à boîte H/ACA. Il est organisé en tandem avec *HEM4*, situé en aval dans le même sens de transcription (**Figure 2.15A**). De la même manière que pour snR13, nous souhaitions mettre en évidence la formation de transcrits de plus longue taille, illustrant un défaut de terminaison. Les résultats préliminaires confirment un défaut de terminaison pour ce snoRNA induite par la délétion du 2nd NTD. Dans la souche portant une version sauvage de *PCF11*, plusieurs fragments sont détectés correspondant à la sonde et à un fragment protégé de petite taille. Dans la souche portant la délétion du 2nd NTD, en plus de la sonde et d'un fragment de petite taille, un fragment protégé de 526 nucléotides est détecté (**Figure 2.15B**). Cependant, la résolution et les conditions de détection ne sont pas optimales. La méthode devra être affinée ultérieurement.

En conclusion, les essais de protection contre la digestion par la S1 nucléase confirment les résultats obtenus précédemment par Northern blot. La délétion du 2nd NTD induit un défaut de terminaison de transcription par l'ARN polymérase II des snoRNAs à C/D et H/ACA (*TABLEAU 2.1*). Ce défaut de terminaison pourrait expliquer en partie le phénotype observé par la délétion du 2nd NTD. La délétion d'une partie du domaine ainsi que les différentes mutations ponctuelles n'induisent pas de défaut de terminaison de transcription. Les essais de protection contre la digestion par la S1 Nucléase contre des ARNs codants (*CYC1*) n'ont pour l'instant pas permis d'obtenir de résultats exploitables.

TABLEAU 2.1 : Récapitulatif des différents mutants obtenus dans cette étude et des phénotypes associés. Les résultats obtenus lors de l'analyse de l'assemblage, des essais de maturation de l'extrémité 3' et de terminaison de transcription sont indiqués. OK : résultats comparable à ceux obtenus pour la souche portant une version sauvage de *PCF11*.

Constructions	Phénotype	CF IA	3' processing		Terminaison transcription			
			Clivage	Polyadénylation	Northern CYH2	Northern TRS31	S1 Nuclease snR13	S1 Nuclease snR5
PCF11	Sauvage à 24°C, 30°C et 37°C - Temps de génération 1h39 - Pas de sensibilité 2% formamide	OK	OK	OK	OK	OK	OK	OK
pcf11-ΔQ_{20}	Sauvage à 24°C, 30°C et 37°C - Temps de génération 1h52 - Pas de sensibilité 2% formamide	OK	OK	OK		OK	OK	
pcf11-Δ(141-233)	Croissance lente à 24°C et 30°C, thermosensibilité à 37°C - Temps de génération 5h - Sensible 2% formamide	OK	OK	OK	Défaut terminaison transcription ?	Défaut terminaison transcription	Défaut terminaison ttranscription	Défaut terminaison transcription
pcf11-Δ(141-241)	Croissance lente à 24°C et 30°C plus prononcée que pcf11-Δ(141-233), thermosensibilité à 37°C - Sensible 2% formamide							
pcf11-Δ(141-253)	Croissance lente à 24°C et 30°C plus prononcée que pcf11-Δ(141-233), thermosensibilité à 37°C							
pcf11-Δ(155-174) α1	Sauvage à 24°C, 30°C et 37°C - Sensible 2% formamide	OK	OK	OK		OK	OK	
pcf11-Δ(180-196) α2	Sauvage à 24°C, 30°C et 37°C - Faible sensibilité 2% formamide	OK	OK	OK		OK	OK	
pcf11-Δ(203-227) α3	Sauvage à 24°C, 30°C et 37°C - Faible sensibilité 2% formamide	OK	OK	OK		OK	OK	
pcf11-14 (L189D)	Sauvage à 24°C, 30°C et 37°C, thermosensibilité à 39°C - Faible sensibilité 2% formamide	OK						
pcf11-15 (L213D)	Sauvage à 24°C, 30°C et 37°C - Faible sensibilité 2% formamide	OK						
pcf11-16 (D220A)	Sauvage à 24°C, 30°C et 37°C - Faible sensibilité 2% formamide	OK						
pcf11-17 (L189D, L213D)	Sauvage à 24°C, 30°C et 37°C, thermosensibilité à 39°C - Sensibilité 2% formamide	Instabilité de la partie N-terminale de la protéine	OK	OK		OK	OK	
pcf11-13 (D68A, S69A, I70A)	Modification croissance à 24°C, 30°C, thermosensibilité à 37°C - Sensibilité 2% formamide						Défaut terminaison transcription	
pcf11-21 (E199A)	Sauvage à 24°C, 30°C et 37°C - Temps de génération 1h39 - Pas de sensibilité 2% formamide							
pcf11-22 (R102A)	Croissance lente à 24°C et 30°C, thermosensibilité à 37°C - Sensible 2% formamide							
pcf11-23 (R102A; E199A)	Croissance lente à 24°C et 30°C plus pronconée que *pcf11-22*, thermosensibilité à 37°C							

C.2. Discussion et Perspectives

Dans la première partie de cette étude, nous nous sommes intéressé au 2nd NTD, nouvellement identifié (Xu et al., 2015). Le 2nd NTD pour « *2nd N-terminal domain* » de Pcf11 est formé des acides aminés 141 à 233, et est situé entre le CID, domaine d'interaction avec le CTD de l'ARN polymérase II et le motif Q_{20}, répétition de 20 résidus glutamines. La structure de ce domaine a été résolue par nos collaborateurs à l'IECB, à Bordeaux (Xu et al., 2015). Le 2nd NTD est formé de trois hélices α formant un domaine hydrophobe au centre de la molécule. Contrairement au CID, le 2nd NTD n'interagit pas avec le CTD de l'ARN polymérase II mais interagit avec l'ARN de séquence GA, comme le CID (Hollingworth et al., 2006). Le CID et le 2nd NTD agissent de manière coopérative pour l'interaction avec l'ARN (C. Mackereth, *données non publiées*). Aucune homologie de séquence n'a pu être mise en évidence pour ce domaine parmi les homologues de Pcf11 dans le règne des métazoaires ; ce domaine semble spécifique aux levures. Cela n'exclut pas la possibilité d'homologies de structures avec les homologues chez les eucaryotes supérieurs. Une analyse bio-informatique poussée, à partir des données structurales et fonctionnelle obtenues pourrait permettre de mettre en évidence des homologies lointaines éventuelles, non détectées par alignements de séquences comme avec PSI-BLAST. Certains nouveaux outils bio-informatiques, qui utilisent une approche locale et non globale de prédiction de structure ont permis de mettre en évidence certaines homologies de domaines dits « cachés » car non identifiées dans les bases de données de domaines comme PFAM ; c'est notamment le cas pour certaines protéines impliquées dans les voies de réparation de l'ADN (Becker et al., 2006). L'objectif de cette étude est d'essayer de comprendre le rôle du 2nd NTD et du motif Q_{20}, dans l'une ou l'autre des fonctions attribuées à Pcf11. Pour répondre à cette question, nous avons mis en place une étude par génétique inverse, en délétant ou en mutant une partie de ce domaine. Quatorze souches ont été obtenues, dont deux possédant un phénotype de croissance très affectée - Δ(141-233) et Δ(141-253). Les phénotypes et les résultats obtenus pour les différents mutants sont récapitulés dans le *TABLEAU 2.1*.

La délétion du 2nd NTD dans son ensemble (Δ(141-233)) conduit à un phénotype de croissance lente à 24°C et 30°C, et une létalité à 37°C ; ce qui suggère un rôle important du 2nd NTD pour la viabilité cellulaire. Ce domaine est proche du CID domaine essentiel chez *S. cerevisiae* (Sadowski et al., 2003). Nous pouvons nous demander si la délétion du 2nd NTD n'induirait pas une déstabilisation de la structure du CID ; le phénotype observé alors serait

dû à un effet indirect. Les données fonctionnelles obtenues par RMN confirment que le 2nd NTD n'interfère pas avec les fonctions du CID (interaction entre le CID et le CTD et interaction avec l'ARN) et que, malgré une structure comparable, ces deux domaines sont formés indépendamment. L'indépendance du CID vis-à-vis du 2nd NTD pour l'interaction avec le CTD peut être confirmée dans un contexte plus général, à partir de facteurs purifiés chez *S. cerevisiae*. L'analyse par *pull-down* GST-CTD en présence de facteurs purifiés à partir des différentes souches construites dans cette étude nous permettrait d'observer si la délétion du 2nd NTD modifie l'interaction du CID avec le CTD en présence du reste de la protéine et des autres partenaires du complexe CF IA.

En accord avec les résultats obtenus par Sadowski *et al.* (Sadowski et al., 2003) définissant la région C-terminale de Pcf11 comme nécessaire et suffisante pour les activités de maturation en 3', nous avons pu montrer que le 2nd NTD n'est pas impliqué dans l'assemblage du sous-complexe CF IA, ni dans les activités de maturation de l'extrémité 3' des ARNs messagers. Cependant, les analyses par Northern blot et par essai de protection contre la digestion par la S1 Nucléase, dirigés contre snR13 ont permis de montrer que le 2nd NTD était impliqué dans la terminaison de transcription des snoRNAs. Dans les essais de protection contre la digestion par la S1 Nucléase, deux populations d'ARNs sont détectées : une présentant une terminaison de transcription normale, et une population plus abondante, de taille plus grande illustrant un défaut de terminaison. La délétion du 2nd NTD de Pcf11 conduit donc à un défaut de terminaison de transcription, cependant d'autres mécanismes permettent de compenser cette déficience, permettant la terminaison d'une partie des snoRNAs et ainsi d'assurer la viabilité des cellules. Cela pourrait expliquer en partie le phénotype marqué mais non létal observé pour la souche délétée pour le 2nd NTD. Ces résultats sont surprenants puisque plusieurs études ont défini le CID comme nécessaire et suffisant pour la terminaison de transcription des ARNs non codants (Grzechnik et al., 2015; Sadowski et al., 2003). De plus, une terminaison de transcription correcte est décrite comme nécessitant l'interaction avec le CTD. Hors, *in vitro*, l'interaction entre le CID et le CTD de l'ARN polymérase II ne semble pas être perturbé par la délétion du 2nd NTD, qui induit un défaut de terminaison de transcription prononcé. Nos résultats permettent de définir un domaine plus grand, englobant l'ensemble de la région N-terminale de la protéine (1-233) comme nécessaire et suffisant pour la terminaison de transcription des ARNs non codants.

In vitro, le 2nd NTD n'interagit pas avec le CTD de l'ARN polymérase mais interagit avec l'ARN de manière coopérative avec le CID. L'interaction du 2nd NTD avec l'ARN serait

plus forte que l'interaction du CID avec un même ARN (C. Mackereth, *données non publiées*). Durant la terminaison de transcription des snoRNAs, le complexe NNS est recruté par interaction avec la forme P-SER^5 du CTD (via Nrd1) (Vasiljeva et al., 2008). La phosphorylation du résidu SER^2 du CTD permet le recrutement de Pcf11 par interaction avec le CTD (Barilla et al., 2001; Licatalosi et al., 2002) et permet de modifier l'interaction entre le CID de Nrd1 et le CTD. Cela permet la transition entre le complexe NNS de terminaison et celui de maturation/dégradation (Grzechnik et al., 2015; Tudek et al., 2014). Le CID de Pcf11 remplacerait le CID de Nrd1 au niveau du CTD de l'ARN polymérase II, le rendant accessible. Ce qui conduit : (i) au recrutement du complexe TRAMP/Exosome, via le CID de Nrd1 et la formation du complexe NNS de maturation/dégradation, (ii) à la translocation de l'hélicase Sen1 (Martin-Tumasz and Brow, 2015; Porrua and Libri, 2015a) et le décrochage de la polymérase. Les deux fonctions de Pcf11 – interaction avec le CTD et interaction avec l'ARN – seraient deux mécanismes en compétition, faisant intervenir des résidus se chevauchant partiellement (Hollingworth et al., 2006). Les essais de protection contre la digestion par la S1 Nucléase ont permis de mettre en évidence que la délétion du 2nd NTD conduit à un défaut de terminaison de transcription et également une forte accumulation de transcrits de grande taille ; cela illustrerait une stabilité de ces transcrits pouvant être due à un défaut de recrutement du complexe TRAMP/Exosome.

Dans la terminaison de transcription des snoRNAs, dépendante du complexe NNS, lorsque le CID de Nrd1 est remplacé par le CID de Pcf11, le 2nd NTD de Pcf11 prendrait le relais du CID dans l'interaction avec l'ARN. Cela permettrait de compenser la perte de l'interaction entre le CID de Pcf11 et l'ARN. Pcf11 (CID + 2nd NTD) aurait un rôle de *linker* entre l'ARN et la polymérase, et pourrait permettre de maintenir le complexe NNS lorsque le CID de Nrd1 est désengagé de son interaction avec le CTD, et permettre à terme l'action de Sen1. Lors de la délétion du 2nd NTD, l'organisation du complexe NNS lorsque le CID de Nrd1 est remplacé par le CID de Pcf11 ne serait pas maintenue. Cela entrainerait un défaut du recrutement du complexe TRAMP/Exosome et de Sen1, qui dans la voie de terminaison dépendante du complexe NNS permettrait de résoudre les hybrides ARN/ADN (Martin-Tumasz and Brow, 2015; Porrua and Libri, 2013), entraînant un défaut de terminaison de transcription.

Les objectifs de ce projet à court terme sont de valider les résultats obtenus par essai de protection contre la digestion par la S1 Nucléase contre snR13 sur un autre snoRNAs à boite H/ACA comme snR5. Les résultats préliminaires obtenus pour ce snoRNA ont permis de mettre en évidence un défaut de terminaison mais les conditions de détection doivent être

optimisées, en modifiant la quantité de sondes ADN et/ou en modifiant les conditions de digestion par la S1 Nucléase. À plus long terme, pour valider le modèle proposé, plusieurs approches peuvent être utilisées : (i) l'analyse par ChIP de la localisation des différents composants du complexe NNS, en particulier Sen1 dans la souche délétée pour le 2nd NTD (ii) l'étude du rôle du 2nd NTD dans la terminaison de transcription des SUTs et des XUTs, où la terminaison de transcription est préférentiellement dépendante du clivage (Grzechnik et al., 2015; Marquardt et al., 2011; Tudek et al., 2015; Tudek et al., 2014).

La complémentation en *trans* de la délétion du 2nd NTD par la protéine entière ou par la partie N-terminale de la protéine (1-287) permet de restaurer le phénotype de croissance et un profil de protection contre la digestion par la S1 Nucléase comparable à ceux de la souche portant une version sauvage de *PCF11*. Les résultats obtenus diffèrent de ceux décrits par Sadowski *et al.* (Sadowski et al., 2003). Le phénotype de croissance de la souche portant l'allèle *pcf11-13* peut être restauré par l'expression du fragment N-terminal 1-260 de Pcf11. Cependant, les analyses par TRO (« *transcriptional run-on* ») ont montré que cela ne permet pas de rétablir un profil de terminaison de transcription comparable à celui de la souche portant une version sauvage de *PCF11*. L'interaction du CID avec le CTD n'est donc pas totalement indépendante du reste de la protéine. Les résultats obtenus permettent de montrer que la fonction du 2nd NTD serait, elle, indépendante du reste de la protéine. Le phénotype de croissance et le profil de protection contre la digestion par la S1 Nucléase sont comparables à ceux de la souche portant une version sauvage de *PCF11*. L'interaction entre l'ARN et le 2nd NTD pourrait être complémentée en *trans* mais pas l'interaction avec le CTD, qui permettrait de lier les fonctions attribuées à Pcf11. L'étude du mutant R102A (*pcf11-22*) où l'interaction avec le CTD est intacte mais pas l'interaction avec l'ARN nous permettra de mieux comprendre ce résultat.

Les résultats obtenus avec le mutant portant l'allèle *pcf11-13*, utilisé comme témoin positif dans cette étude, sont surprenants. Ce mutant, initialement décrit par Sadowski *et al.,* (Sadowski et al., 2003) comprend trois mutations au niveau de résidus conservés du CID (D68A, S69A, I70A) conduisant à un défaut de terminaison de transcription des ARNs codants et non codants, par un défaut d'interaction avec le CTD de l'ARN polymérase II (Grzechnik et al., 2015; Sadowski et al., 2003). La mutation des résidus DSI conduit à une déstructuration du CID (C. Mackereth, *données non publiées*). Dans les essais de protection contre la digestion par la S1 Nucléase, les fragments protégés de grande taille illustrant un

défaut de terminaison sont détectés en très faible quantité. Nous pouvons imaginer que la présence du CID, même déstructuré, serait suffisante pour assurer une terminaison de transcription, expliquant les résultats obtenus dans les essais de protection contre la digestion par le S1 Nucléase. Les différences entre les résultats obtenus dans cette étude et dans celle publiée par Sadowski *et al.,* peuvent également s'expliquer par les méthodes d'analyse. Dans l'étude précédemment citée, le défaut de terminaison de transcription est mis en évidence par TRO qui permet d'étudier la terminaison en fin de transcription. Dans notre étude, le défaut de terminaison de transcription est étudié par essais de protection contre la digestion par la S1 Nucléase à partir d'ARN totaux ; la population d'ARNs étudiée correspond à l'équilibre entre synthèse et dégradation. Les ARNs de grandes tailles synthétisés en présence des mutations de l'allèle *pcf11-13* peuvent être dégradés et moins visibles dans notre étude en comparaison des analyses par TRO, expliquant les résultats obtenus.

Les mutations de l'allèle *pcf11-13* conduisent à la déstructuration du CID et une inhibition de l'interaction entre le CID et le CTD mais n'induisent pas de phénotype de croissance marqué à 24°C et 30°C. La mutation R102A (*pcf11-22*) n'affecte, elle, pas la structure du CID et n'interfère pas dans l'interaction avec le CTD. Cependant, cette mutation diminue l'interaction entre le CID et l'ARN (C. Mackereth, *données non publiées*), et conduit à un phénotype de croissance lente important en comparaison au phénotype des mutations *pcf11-13* où l'interaction avec le CTD est modifiée – l'interaction entre le CID, déstructuré et l'ARN n'est pas connue. L'interaction entre le CID et l'ARN aurait une fonction prédominante par rapport à l'interaction avec le CTD de l'ARN polymérase II. De plus, la diminution de l'interaction entre l'ARN et le 2nd NTD (mutation E199A - *pcf11-21*) associée à la mutation R102A (*pcf11-23* - R102A, E199A) aggrave le phénotype observé ; l'interaction de la région N-terminale de la protéine avec l'ARN est fortement réduite et serait essentielle. Zhang *et al.,* (Zhang and Gilmour, 2006) ont montré que l'interaction entre Pcf11 et l'ARN permet le désassemblage de la polymérase en fin de transcription. À l'exception des phénotypes observés et des résultats obtenus par RMN, peu de données sont disponibles pour ces mutants. Un modèle de coopération entre le CID et le 2nd NTD peut cependant être proposé. Les deux fonctions de Pcf11 – interaction avec le CTD et interaction avec l'ARN – sont deux mécanismes en compétition, faisant intervenir des résidus se chevauchant partiellement (Hollingworth et al., 2006). La mutation du résidu R102 aurait un effet dominant négatif et induirait une perte de la fonction d'interaction du CID avec l'ARN. Le CID de Pcf11 serait alors « bloqué » dans l'interaction avec le CTD, ne permettant plus l'échange entre interaction avec le CTD et interaction avec l'ARN ; cela ne permettrait plus le

désassemblage de l'ARN polymérase. Le 2nd NTD permettrait de renforcer l'interaction avec le pré-ARNm et pourrait également être nécessaire au recrutement de Sen1 qui permettrait de résoudre les hybrides ARN/ADN. L'étude de la terminaison de transcription de ce mutant sera cruciale pour comprendre le rôle de l'interaction entre l'ARN et le CID, ainsi que le rôle du 2nd NTD.

Pcf11 étant également impliquée dans la voie de terminaison dépendante du complexe de polyadénylation (Grzechnik et al., 2015; Lunde et al., 2010; Sadowski et al., 2003), nous nous sommes intéressé à la terminaison de transcription des ARNs codants. Les analyses par Northern blot et par essai de protection contre la digestion par la S1 Nucléase n'ont pas permis d'étudier avec précision le rôle du 2nd NTD dans la terminaison de transcription des ARNs codants. Plusieurs facteurs sont à prendre en compte : (i) la détection des ARNs ribosomiques par Northern blot n'a pas permis d'obtenir un profil suffisamment résolutif ; (ii) l'abondance et la demi-vie des ARNs utilisés comme cibles (*CYH2* et *CYC1*). À court terme, nous souhaitons diminuer la quantité d'ARNs ribosomiques dans les échantillons, par l'utilisation d'un kit (RiboMinus™ - Fisher Scientific). Cela nous permettrait d'éliminer le bruit de fond et d'obtenir une meilleure résolution afin de valider le rôle du 2nd NTD. Nous souhaitons également modifier les cibles utilisées, en particulier le choix de la cible dans les essais de protection contre le S1 Nucléase. Nous avons choisi d'étudier *CYC1*, utilisé précédemment dans des études de terminaison de transcription par run-on mais peu abondant dans les cellules en comparaison à d'autres ARNs (Dichtl et al., 2002a). *PGK1* ou *ADH1* pourraient être des cibles plus adaptées, car plus abondants et présentants une demi-vie plus importante (45 minutes pour *PGK1*). À plus long terme, nous souhaitons valider les résultats obtenus (Northern blot et essai de protection contre la digestion par la S1 Nucléase) par RT-qPCR (Grzechnik et al., 2015). Cela nous permettra d'avoir une vision globale du rôle du 2nd NTD.

Cependant l'analyse par Northern blot dirigée contre *TRS31* suggère que le 2nd NTD ne serait pas impliquée dans la terminaison de transcription des ARNs codants, terminés par la voie dépendante du complexe de polyadénylation. La délétion du 2nd NTD conduit à l'accumulation d'un fragment correspondant uniquement à snR13 et *TRS31* (1400 nucléotides), due à un défaut de la voie de terminaison des snoRNAs. Un défaut également dans la voie dépendante du complexe de polyadénylation conduirait à la synthèse et la détection d'un fragment de plus grande taille – supérieur à 1400 nucléotides. Nous avons également montré que le 2nd NTD n'induit pas de modifications des étapes de maturation des

ARNs messagers. Les ARNs détectés par Northern blot correspondraient à snR13 et *TRS31* stabilisé par l'ajout d'une queue poly(A) dans la région 3' de *TRS31*. Cependant, l'analyse par Northern blot est effectuée sur des ARNs totaux, et la population d'ARNs étudiée correspond à l'équilibre entre synthèse et dégradation. L'hypothèse de la synthèse d'ARNs de plus grande taille illustrant un défaut partiel de la voie dépendante du complexe de polyadénylation ne peut être totalement exclue. Si les résultats obtenus dans cette étude sont confirmés par Northern blot et RT-qPCR, le 2nd NTD constituerait un domaine unique, impliqué dans la terminaison de transcription dépendante du complexe NNS.

De manière surprenante, les mutations ponctuelles au niveau de la région hydrophobe du 2nd NTD formée par les trois hélices α n'induisent pas de modification du phénotype de croissance aux températures étudiées. Cependant, les mutations L189D et L213D correspondant à l'allèle *pcf11-17* confèrent une sensibilité en présence de 2% de formamide ainsi qu'une instabilité de la partie N-terminale de la protéine lors de la purification par *TAP-tagging*. Les études des activités de maturation en 3' des ARN messagers et de la terminaison de transcription des snoRNAs (Northern blot et essai de protection contre la digestion par la S1 Nucléase) n'ont pas permis de mettre en évidence un effet biologique de cette instabilité. Cette instabilité pourrait correspondre à un artefact causé par les conditions de purification utilisées. Nous avons essayé de répondre à cette question en étudiant la stabilité de Pcf11 en présence de l'allèle *pcf11-17* à partir d'extraits totaux. Les résultats obtenus à ce jour n'ont pas permis de reproduire l'instabilité observée lors de la préparation des facteurs par *TAP-tagging*. La forte instabilité de la région N-terminale de la protéine observé lors de la purification des facteurs serait due aux conditions de purification. La sensibilité uniquement en présence de formamide induite par les mutations L189D et L213D suggère cependant que cette instabilité pourrait être présente *in vivo* mais constituerait un événement minoritaire.

La délétion de l'une des trois hélices α du 2nd NTD n'induit pas de modification du phénotype de croissance ni de défaut dans l'une des fonctions de Pcf11 (maturation de l'extrémité 3' ou terminaison de transcription). La délétion de l'hélice α1 induit une sensibilité en présence de formamide ; cela peut être dû à la modification de l'interaction avec l'ARN faisant intervenir les acides aminés amont de l'hélice α1. L'objectif à plus long terme serait d'étudier l'impact de la délétion de deux des trois hélices du 2nd NTD. La structure du 2nd NTD laisse suggérer un effet de redondance entre les hélices et la présence de deux des trois hélices suffirait pour assurer la structuration du domaine et sa fonction.

La délétion du motif Q_{20} en aval du 2nd NTD ne modifie pas la croissance des souches aux températures étudiées ; cependant le délétion du 2nd NTD et du motif Q_{20} (Δ(141-253)) conduit à une aggravation du phénotype de croissance lente observé lors de la délétion du 2nd NTD. Parmi les protéines impliquées dans la maturation de l'extrémité 3', Nab2 présente une répétition de résidus glutamines (douze résidus) liant les activités d'export (N-terminal) et d'interaction avec l'ARN (domaine central et C-terminal). Comme suggéré par Amrani *et al.*, (Amrani et al., 1997b), le domaine Q_{20} pourrait conférer une flexibilité entre les régions N-terminale et C-terminale de la protéine, impliquées dans deux fonctions distinctes (Sadowski et al., 2003). La délétion du 2nd NTD et du motif Q_{20} peut conduire à une perte de flexibilité de la protéine, et perturber les fonctions indépendantes de la région N-terminale de la protéine. Les répétitions de résidus glutamines sont fréquentes dans de nombreuses protéines, comme les protéines Sp, facteurs généraux de transcription (Suske, 1999) ou les facteurs de transcription liant l'AMPc. Plus éloigné de la terminaison de transcription et de la maturation de l'extrémité 3', l'augmentation de la taille des répétitions de résidus glutamines est associée, chez l'Homme, au développement de maladies neuro-dégénératives (neuf ont été recensées) comme la maladie de Huntington. La mutation CAG au niveau de la protéine Huntingtine augmente la taille de la répétition de glutamines (Snell et al., 1993). Récemment, il a été mis en évidence par FLIM-FRET, que la répétition de résidus glutamines de cette protéine formait un domaine flexible, permettant de lier les régions N-terminale et C-terminale de la protéine (Caron et al., 2013). L'augmentation de la taille modifie les interactions et la conformation de la protéine. Une fonction semblable pourrait être attribuée au motif Q_{20} de Pcf11.

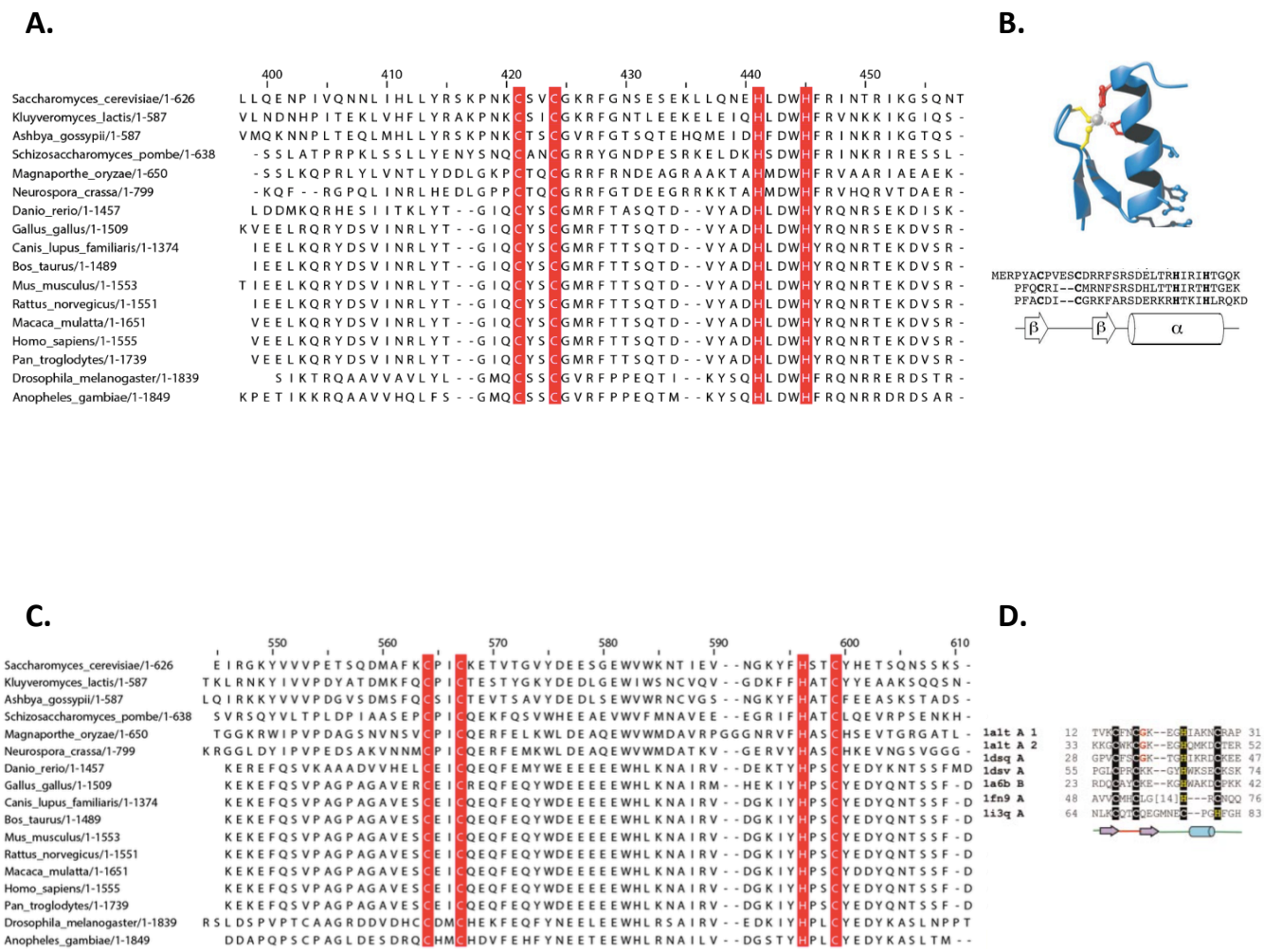


Figure 2.16 : Données préliminaires sur les motifs en doigt de zinc présents dans la partie C-terminale de Pcf11. **(A.)** Alignement de la séquence 398 à 466 de Pcf11 et de ses homologues dans le règne eucaryote (de *K. lactis* à *H. sapiens*). L'alignement est réalisé par ClustalΩ. Les résidus C421, C424, H441 et H445 correspondant au premier motif en doigt de zinc sont conservés et indiqués en rouge. D'après A. Dupin, équipe S Fribourg. **(B.)** (*haut*) Représentation du motif en doigt de zinc de type **Cys_2His_2** de la protéine Zif268. Les chaines latérales des résidus cystéines sont représentées en jaune et les chaines latérales de résidus histidines sont représentés en rouge. L'atome de zinc est représenté en gris. (*bas*) Alignement de séquences du motif en doigt de zinc de type **Cys_2His_2** de la protéine Zif268. Les résidus conservés sont représentés en gras. Les structures secondaires sont indiquées. D'après Pabo *et al.,* 2001. **(C.)** Alignement de la séquence 544 à 611 de Pcf11 et de ses homologues dans le règne eucaryote (de *K. lactis* à *H. sapiens*). L'alignement est réalisé par ClustalΩ. Les résidus C564, C567, H596 et C599 correspondant au second motif en doigt de zinc sont conservés et indiqués en rouge. D'après A. Dupin, équipe S Fribourg. **(D.)** (*haut*) Alignement de séquences du groupe « *gag knuckle* » de type **$Cys_2HisCys$**. Les résidus conservés sont représentés en noir. Les structures secondaires sont indiquées – feuillets β en **rose** et hélice α en **bleu**. D'après Krishna *et al.,* 2003.

D. ÉTUDE DES MOTIFS EN « DOIGT DE ZINC » DE PCF11

D.1. Données préliminaires (S. Fribourg – IECB, Bordeaux)

Dans la région C-terminale de la protéine, Sadowski *et al.,* (Sadowski et al., 2003) ont décrit la présence de deux domaines, encadrant la région d'interaction avec Clp1, contenant des motifs supposés lier un ou plusieurs atomes de zinc. Cependant ces domaines ne sont pas étudiés à ce jour. Nous nous sommes donc intéressé à l'organisation et à la fonction de ces domaines en collaboration avec nos partenaires à l'IECB dans les équipes de S. Fribourg et de C. Mackereth.

L'analyse des séquences a permis de montrer que ces domaines, d'environ 30 nucléotides, sont conservés dans le règne eucaryote en comparaison des autres motifs liant le zinc de Pcf11 (**Figure 2.16A et C** – A. Dupin, équipe de S. Fribourg, Bordeaux). Le premier motif de type **Cys_2His_2** est situé entre les résidus 420 à 450 et présente toutes les caractéristiques théoriques de la classe des motifs en doigt de zinc la mieux caractérisée (Pabo et al., 2001) : ils sont repliés en ββα et sont décrits pas la séquence X_2-**Cys**-$X_{2,4}$-**Cys**-X_{12}-**His**-$X_{3,4,5}$-**His** (**Figure 2.16B**) (Krishna et al., 2003). Ils permettent l'interaction avec des cibles ADN, ARN et protéines. Le second motif (situé entre les résidus 560 à 610) est décrit par les résidus **Cys_2HisCys**, caractéristiques des motifs de type « *gag knuckle* » et qui forment deux feuillets β liés par une boucle suivi d'une hélice α (**Figure 2.16D**) (Krishna et al., 2003). Chaque motif est décrit comme pouvant potentiellement lier un atome de zinc. L'analyse par absorption atomique et par spectrométrie de masse a permis de montrer effectivement la présence de deux atomes de zinc apportés par le dimère Clp1-Pcf11 et portés par les motifs en doigt de zinc de Pcf11 (A. Dupin, équipe de S. Fribourg, Bordeaux et J. Stojko, équipe de S. Cianférani, Strasbourg). Dans le but d'aller plus loin dans l'étude de ces motifs en doigt de zinc, nous avons réalisé une étude fonctionnelle à partir de mutants ponctuels. Les différents mutants obtenus dans cette étude ainsi que les résultats obtenus pour chaque mutant sont récapitulés dans le *TABLEAU 2.2*.

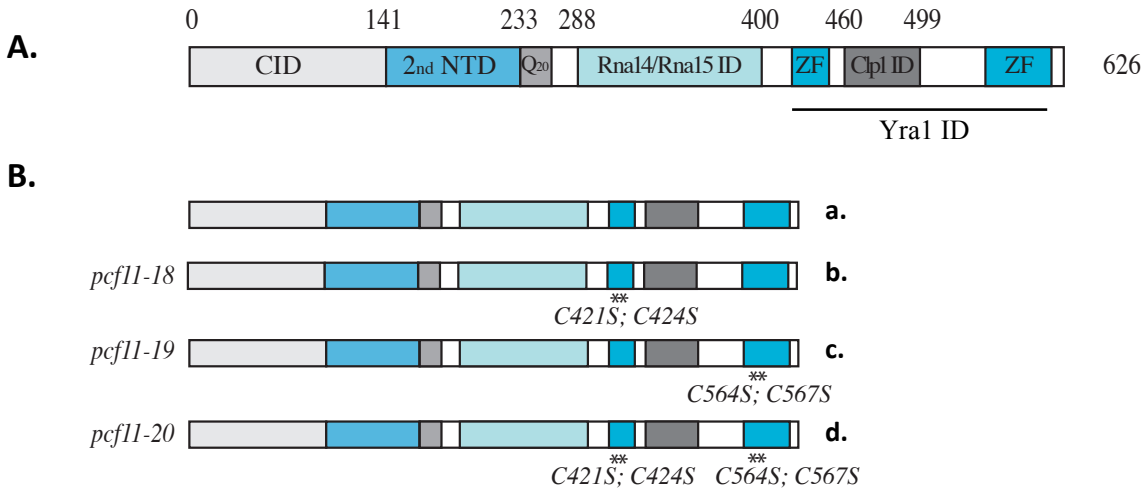


Figure 2.17 : Représentation schématique des différents mutants construits pour l'étude des motifs en doigts de zinc de Pcf11. **(A.)** Représentation schématique de l'organisation des différents domaines de Pcf11. Les différents domaines sont à l'échelle et leur taille est donnée en acides aminés. CID (1-141) : « *CTD Interaction Domain* ». 2nd NTD (141-233) : « *2nd N-terminal domain* ». Q_{20} (233-253) : motif Q_{20} formé d'une répétition de 20 résidus glutamines. Rna14/RNa15 ID (288-400) : région contenant les domaines d'interaction avec Rna14 et Rna15. Clp1 ID (460-499) : région contenant le domaine d'interaction avec Clp1. ZF (420-450 et 560-610) : région comprenant deux motifs en doigt de zinc. Yra1 ID (420-608) : région d'interaction minimale avec Yra1. **(B.)** Représentation des différents mutants utilisés dans cette étude. Les * représentent les mutations ponctuelles ; les résidus mutés sont indiqués en italique. ***a***. *PCF11*. ***b.*** : *pcf11-18* : motif en doigt de zinc type **Cys_2His_2** (C421S, C424S). ***c.*** : *pcf11-19* : motif en doigt de zinc type **$Cys_2HisCys$** (C564S, C567S). ***d.*** *: pcf11-20* : mutations au niveau de deux motifs en doigt de zinc (C421S, C424S, C564S, C567S).

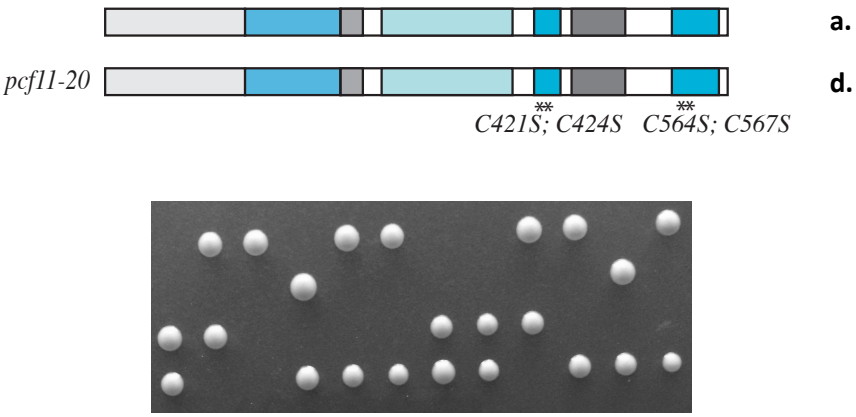


Figure 2.18 : La mutation des deux motifs en doigt de zinc de Pcf11 est létale chez *S. cerevisiae*. (*haut*) Représentation schématique de la mutation de l'allèle *pcf11-20* mutations au niveau de deux motifs en doigt de zinc (C421S, C424S, C564S, C567S). (*bas*) Étude de la ségrégation des spores pour la mutation de deux motifs en doigt de zinc de Pcf11. Les spores sont séparées à l'aide d'un micromanipulateur sur milieu YPD(A) puis incubées à 24°C 3 jours. La présence de cellules est vérifiée au microscope.

D.2. Résultats

D.2.1. Obtention des mutants

D.2.1.1. Obtention des mutants

La première étape de ce projet a été de construire différents mutants pour ces domaines. Les mutants ponctuels utilisés sont présentés dans la **Figure 2.17B**. Dans le premier motif, les résidus Cystéines en position 421 et 424 sont mutés en Sérine (mutant ***pcf11-18*** – b.) ; dans le second motif nous avons modifié les résidus Cystéines en position 564 et 567 en résidus Sérine (mutant ***pcf11-19*** –c.). Nous avons également obtenu un mutant portant des mutations au niveau des résidus Cystéines des deux motifs (C421S, C424S, C564S, C567S - ***pcf11-20*** – d.).

Les mutations correspondant aux mutants ***pcf11-18*** et ***pcf11-19*** ont été introduites sur, respectivement, les plasmides pFL36ΔSX et pFL36 (marqueur *LEU2*). Nous avons transformé la souche YLM233 (*pcf11 ::HIS3* complémentée par pFL38-*PCF11*, marqueur *URA3*) avec les plasmides pFL36ΔSX et pFL36 puis effectué la chasse du plasmide pFL38-*PCF11* sur milieu 5-FOA. Cette approche a pu être utilisée pour les mutants ***pcf11-18*** et ***pcf11-19***. Pour les mutations au niveau des deux motifs correspondant à l'allèle ***pcf11-20***, nous avons utilisé la même approche que celle utilisée pour étudier la double délétion [2^{nd} NTD+Q_{20}]. Le plasmide portant la double mutation ***pcf11-20*** a été transformé dans la souche diploïde YLM232 (MATa/α - *pcf11*::*TRP1*/*PCF11*). Après sporulation, nous avons analysé la ségrégation des spores, et comme présenté dans la **Figure 2.18** obtenu une ségrégation de type 2 : 0 – deux spores viables et deux spores non viables pour la totalité des tétrades analysées. Les spores viables sont Trp -. L'association des mutations au niveau des deux motifs en doigt de zinc (***pcf11-20***) est létale.

D.2.1.2. Caractérisation du phénotype associé

Après obtention des différents mutants, nous avons caractérisé les phénotypes associés aux différentes mutations sur milieu complet YPD(A) à 24°C, 30°C et 37°C et en présence de 2% de formamide. La formamide est non métabolisable par les levures et modifierait les liaisons hydrogènes ; une sensibilité au formamide est décrite comme illustrant une modification de la stabilité des protéines (Hampsey, 1997). Les résultats sont présentés dans la **Figure 2.19A**. Les mutations (C421S, C424S) correspondantes au mutant ***pcf11-18***

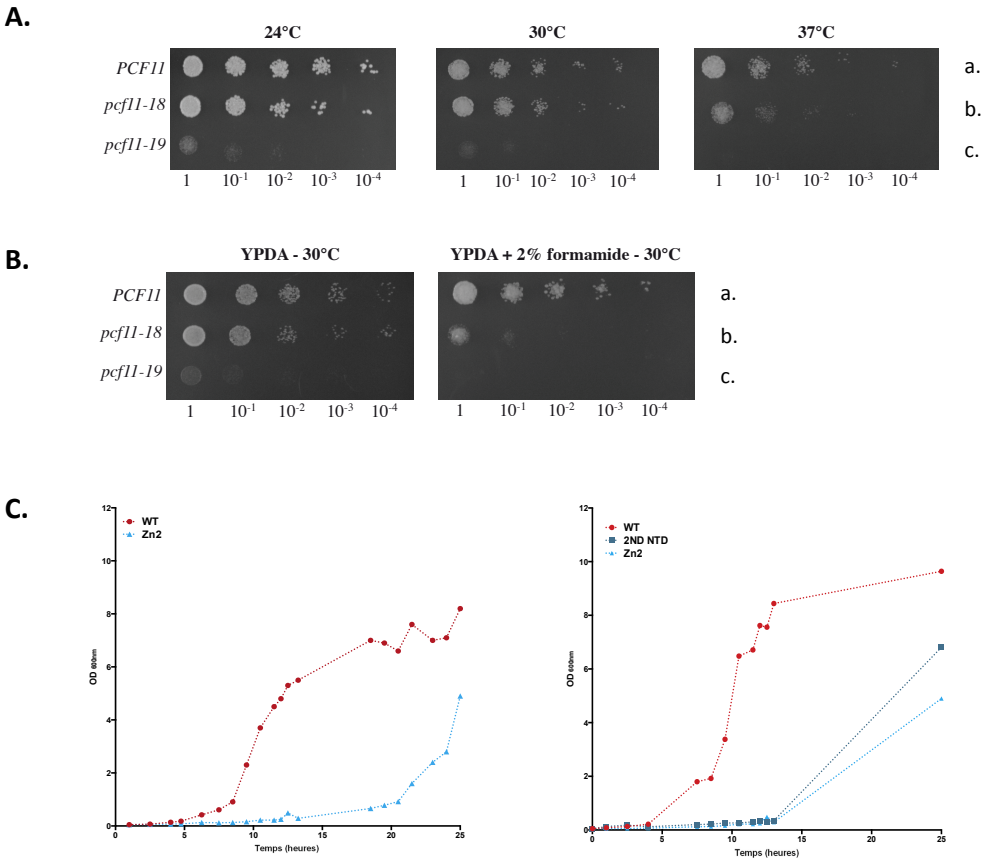


Figure 2.19 : La mutation du second motif en doigt de zinc induit un phénotype de croissance lente marqué. **(A.)** Analyse des phénotypes associés aux mutants des motifs en doigt de zinc de Pcf11 à 24°C, 30°C et 37°C sur milieu YPD(A). ***a.*** témoin positif, souche exprimant une version sauvage de *PCF11*. ***b.*** *pcf11-18* : motif en doigt de zinc type **Cys_2His_2** (C421S, C424S). ***c.*** *pcf11-19* : motif en doigt de zinc type **$Cys_2HisCys$** (C564S, C567S). **(B.)** Phénotypes associés aux mutants des motifs en doigt de zinc de Pcf11 à 30°C sur milieu YPD(A) en absence (*à gauche*) et en présence de 2% de formamide (*à droite*). ***a.*** témoin positif, souche exprimant une version sauvage de *PCF11*. ***b.*** *pcf11-18* : motif en doigt de zinc type **Cys_2His_2** (C421S, C424S). ***c.*** *pcf11-19* : motif en doigt de zinc type **$Cys_2HisCys$** (C564S, C567S). **(C.)** Courbes de croissance en milieu liquide à 30°C pendant 25 heures. La DO à 600nm est donnée en fonction du temps d'incubation en heures. (¢) souche exprimant une version sauvage de *PCF11*. (n) pcf11Δ(141-233) : souche délétée pour le 2nd NTD. (Δ) *pcf11-19* : motif en doigt de zinc type **$Cys_2HisCys$** (C564S, C567S).

localisées au niveau du premier motif en doigt de zinc n'induisent pas de modification de croissance à 24°C et 30°C (**Figure 2.19A**, ligne b). À 37°C, on observe une thermosensibilité en comparaison de la souche portant une version sauvage de *PCF11*. Les mutations du premier motif en doigt de zinc (C421S, C424S – ***pcf11-18***) induisent une sensibilité à 30°C en présence de 2% de formamide (**Figure 2.19B**, ligne b). Les mutations (C564S, C567S) correspondant au mutant ***pcf11-19*** au niveau du second motif en doigt de zinc affectent la croissance de manière important à 24°C et 30°C et conduisent également à une létalité à 37°C (**Figure 2.19A**, ligne c). Les mutations de l'allèle ***pcf11-19*** induisent une sensibilité accrue à la température (30°C) en présence de 2% de formamide (**Figure 2.19B**, ligne c).

Nous avons également réalisé une courbe de croissance à 30°C, en milieu liquide complet, pendant 24 heures du mutant ***pcf11-19*** en comparaison avec la souche portant une version sauvage de *PCF11* et la souche délétée pour le 2nd NTD – **Δ(141-233)** (**Figure 2.19C**). Le temps de génération est d'environ 5h30 pour le mutant ***pcf11-19*** (courbe **Δ – Zn2**), pour un temps de génération d'environ 1h40 pour la souche sauvage (courbe **O - WT**). La mutation au niveau du second motif **Cys$_2$HisCys** induit un retard de croissance plus important que la délétion du 2nd NTD (courbe ■) – après 12 heures de culture, DO_{600nm} de 0,81 pour la souche **Δ(141-233)** contre de 0,22 pour le mutant ***pcf11-19***. En milieu liquide, on peut voir que la souche mutante ***pcf11-19*** présente un délai de croissance jusqu'à 20 heures d'incubation, puis une augmentation de la rapidité de croissance. Cela suggère un envahissement de la culture par des révertants. Pour écarter cette hypothèse, nous avons effectué une nouvelle courbe de croissance, et, après 24 heures, incubé de nouveau les cellules obtenues sur milieu YPD(A) solide et en milieu liquide à une DO_{600nm} initiale de 0,05 en comparaison à la souche sauvage (*données non montrées*). L'allure générale de la courbe est identique à celle obtenue précédemment. Le délai de croissance observé pour la mutation ***pcf11-19*** n'est pas dû à l'apparition de révertants, mais traduirait une adaptation au milieu du mutant.

En conclusion, dans le but d'étudier les motifs en doigt de zinc présents dans la partie C-terminale de Pcf11, nous disposons de trois mutants :

- ***pcf11-18*** (C421S, C424S)
- ***pcf11-19*** (C564S, C567S)
- ***pcf11-20*** (C421S, C424S, C564S, C567S)

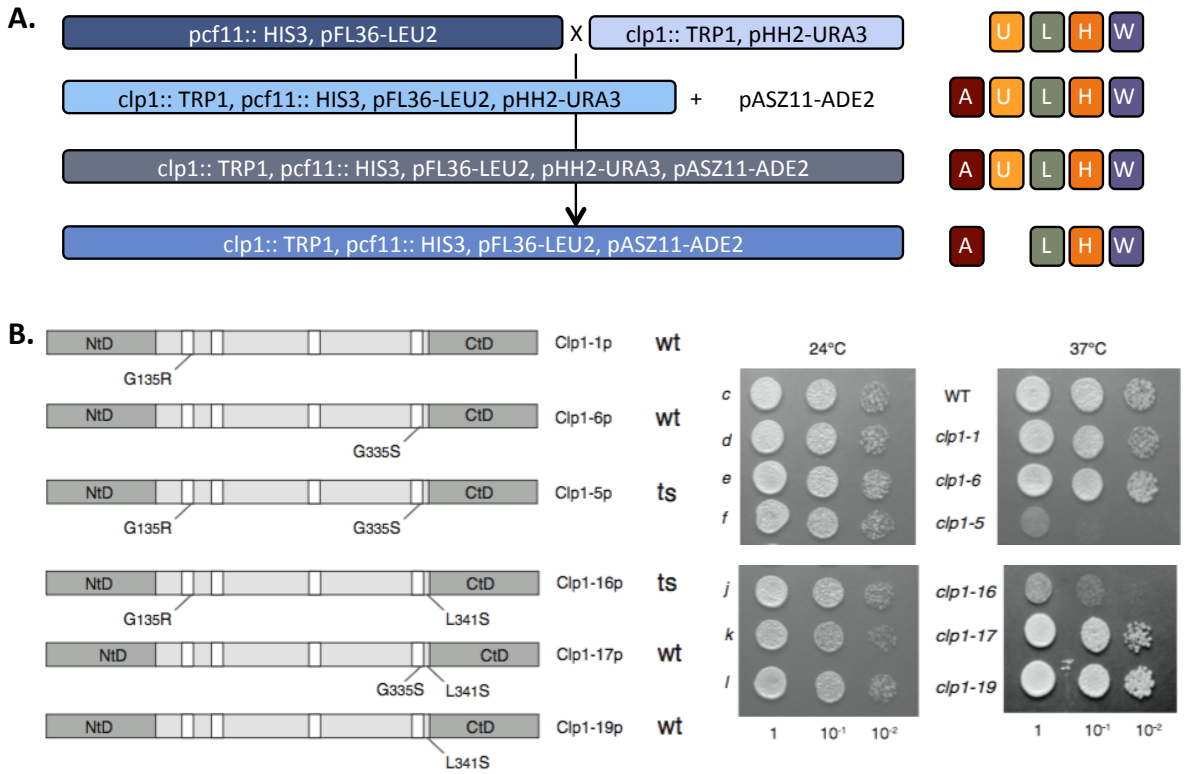


Figure 2.20 : Recherche de létalité synthétique entre les mutations des ZF de Pcf11 et différentes mutations de Clp1. **(A.)** Représentation de la stratégie utilisée pour la recherche de létalité synthétique. Les souches haploïdes portant la délétion chromosomique *pcf11 :HIS3* complémentée par le vecteur d'expression pFL36 (marqueur *LEU2*) portant la version sauvage de *PCF11* ou les allèles *pcf11-18* et *pcf11-19* sont croisées avec la souche portant la délétion chromosomique *clp1::TRP1* complémentée par le vecteur pHH2-*CLP1*, marqueur *URA3*. La sélection se fait sur milieu SC-U-L-H-W. Les souches obtenues sont transformées avec le plasmide pASZ11, marqueur *ADE2*, portant les allèles de *CLP1* testées. La sélection se fait sur milieu SC-A-U-L-H-W. Le plasmide pHH2 portant le version sauvage de CLP1 est chassé sur milieu 5-FOA. La sélection se fait sur milieu SC-A-L-H-W avec ou sans Uracile. **(B.)** Représentation schématique de Clp1 et des six allèles testées. Le phénotype induit par chaque mutation est indiqué (*à droite*). : *clp1-1* (G135R – phénotype sauvage), *clp1-6* (G335S – phénotype sauvage), *clp1-5* (G135R, G335S – thermosensible), *clp1-16* (G135R, L341S – thermosensible), *clp1-17* (G335S, L341S – phénotype sauvage) et *clp1-19* (L341S – phénotype sauvage).

La mutation simultanée des deux motifs est létale chez *S. cerevisiae*, illustrant le rôle majeur de ces deux domaines. La mutation du second motif **$Cys_2HisCys$** (C564S, C567S) conduit à un phénotype de croissance lente important ; la mutation du premier motif **Cys_2His_2** (C421S, C424S) induit une létalité à 37°C et une sensibilité au formamide à 30°C. Les phénotypes des chacune des souches sont récapitulés dans le *TABLEAU 2.2*.

D.2.2. Recherche d'une létalité synthétique avec Clp1

L'analyse, par génétique inverse, a permis d'assigner une fonction à un grand nombre de gènes chez *S. cerevisiae*. La recherche de létalité synthétique est une approche permettant de comprendre les interactions génétiques pouvant exister entre deux gènes et d'attribuer un rôle aux protéines de fonctions inconnues ainsi que d'appréhender la complexité des processus cellulaires (Hartman et al., 2001; Tong et al., 2004). Cela se fait par la recherche de létalité ou de modifications majeures de phénotypes en combinant les mutations au niveau de deux gènes. Si les protéines testées sont impliquées dans les mêmes fonctions biologiques ou dans des voies parallèles, la létalité synthétique indique que les produits de gènes interagissent, de manière directe ou indirecte. À l'inverse, si les deux protéines testées sont impliquées dans deux voies biochimiques distinctes, la létalité synthétique indique que la fonction de l'une des deux protéines peut compenser la déficience de l'autre. Cette approche a pu être utilisée pour caractériser les interactions génétiques entre Rna14, Rna15 et Pap1 (Minvielle-Sebastia et al., 1994), entre Rna14, Rna15 et Pcf11 (Amrani et al., 1997b) ou entre Yra1 et Sub2 (Strasser and Hurt, 2001).

Nous nous sommes intéressé aux interactions génétiques pouvant exister entre les mutants des motifs en doigt de zinc de Pcf11 et certains mutants de Clp1 précédemment décrits (Haddad et al., 2012). Pcf11 et Clp1 étant codées par deux gènes essentiels, nous avons étudié la létalité synthétique pouvant exister en utilisant le principe de la chasse sur 5-FOA (**Figure 2.20A**). La première étape de ce projet a été de construire des souches contenant les mutations de *PCF11* et dans lesquelles *CLP1* était présent sur un plasmide portant le marqueur *URA3*. Pour cela nous avons croisé les souches obtenues précédemment (YLM233 transformée avec les plasmides pFL36ΔSX-***pcf11-18*** (C421S, C424S) et pFL36-***pcf11-19*** (C564S, C567S) – marqueur *LEU2*) avec la souche *clp1::TRP1*, pHH2-*URA3-CLP1* disponible au laboratoire (YLM126).

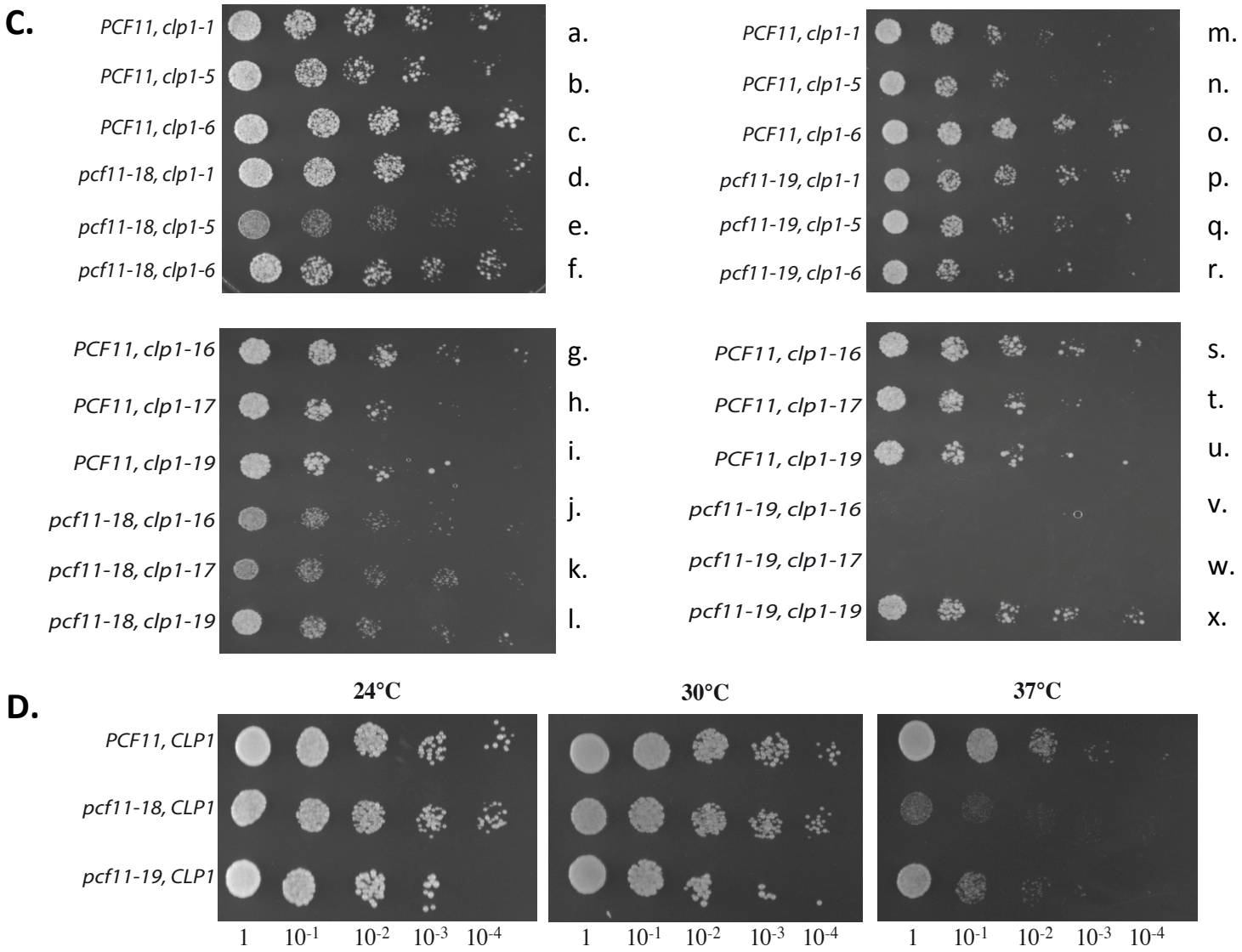


Figure 2.20 : (*Suite*)

(C.) Résultats obtenus sur milieu 5-FOA à 24°C, après 72 heures d'incubation. Les différentes combinaisons sont indiquées pour chaque ligne.

(D.) Phénotype des souches parentales sur milieu YPD(A) à 24°C, 30°C et 37°C.

PCF11, CLP1 :*pcf11::HIS3*, *clp1::TRP1*, pFL36-PCF11, pHH2-*CLP1*.

pcf11-18, CLP1 : *pcf11::HIS3, clp1::TRP1*, pFL36-*pcf11-18*, pHH2-*CLP1*.

pcf11-19, CLP1 : *pcf11::HIS3, clp1::TRP1*, pFL36-*pcf11-19*, pHH2-*CLP1*.

Après sélection des diploïdes sur milieu SC-L-H-U-W, dissection des tétrades et sélection, trois souches ont été obtenues :

- YJG48 (*clp1::TRP1, pcf11::HIS3, pFL36 ΔSX-pcf11-18(C421S, C424S), pHH2-URA3-CLP1*)
- YJG49 (*clp1::TRP1, pcf11::HIS3, pFL36-pcf11-19(C564S, C567S), pHH2-URA3-CLP1*)
- YJG50 (*clp1::TRP1, pcf11::HIS3, pFL36-PCF11, pHH2-URA3-CLP1*)

Ces souches sont ensuite transformées par les plasmides pASZ11 (marqueur *ADE2*) portant les différentes allèles de *CLP1* disponibles au laboratoire que nous souhaitions tester. Les transformants sont sélectionnés sur milieu SC-A-L-U-W-H. La létalité synthétique pouvant exister entre les différents allèles où la croissance plus affectée que celle de chacun des mutants pris individuellement sont testées sur milieu 5-FOA. Cela permet la chasse du plasmide pHH2-*URA3-CLP1*. L'efficacité de la chasse du plasmide portant la version sauvage de *CLP1* est vérifiée par étalement sur milieu SC-A-L-W-H en présence et en absence d'uracile.

Nous nous sommes intéressés à six mutations de Clp1 : ***clp1-1*** (G135R – phénotype sauvage), ***clp1-6*** (G335S – phénotype sauvage), ***clp1-5*** (G135R, G335S – thermosensible), ***clp1-16*** (G135R, L341S – thermosensible), ***clp1-17*** (G335S, L341S – phénotype sauvage) et ***clp1-19*** (L341S – phénotype sauvage) (**Figure 2.20B**).

Les résultats obtenus sont présentés dans la **Figure 2.20C**. La combinaison entre les mutations de l'allèle ***pcf11-18*** (C421S, C424S) et celles des allèles ***clp1-5, clp1-16, clp1-17*** et ***clp1-19*** conduit à une modification du phénotype de croissance (aspect « petite colonies ») (lignes e, j, k et l). Les autres allèles testés (***clp1-1***, ***clp1-6*** et ***clp1-19***) n'induisent pas de modifications du phénotype dans les conditions testées. De plus, les clones obtenus se développent sur milieu SC-A-L-W-H-U ; cela signifie que le plasmide portant la version sauvage de *CLP1* n'est pas correctement éliminé sur 5-FOA. Dans le cas de la combinaison entre l'allèle ***pcf11-18*** (C421S, C424S) et les allèles ***clp1-5*** et ***clp1-16***, la croissance n'est pas plus affectée que celle de chacun des mutants de *PCF11* et *CLP1* pris individuellement. Le phénotype observé pour la combinaison de l'allèle ***pcf11-18*** (C421S, C424S) avec les allèles ***clp1-17*** et ***clp1-19*** pourrait correspondre à un effet cumulatif des mutations. La mutation G335S est située dans le motif YFYG, zone d'interaction avec Pcf11 et la mutation L341S est située à proximité directe de ce motif. Les acides aminés C421 et C424 de Pcf11 seraient potentiellement impliqués, de manière directe ou indirecte, dans l'interaction de Pcf11 avec Clp1, ou un de ses autres partenaires.

Dans le cas de ***pcf11-19*** (C564S, C567S), les résultats obtenus ne correspondent pas à ceux pressentis. En effet, dans les souches construites, *PCF11* est délété au niveau chromosomique, et la seule version qui s'exprime est celle portée par le plasmide pFL36. Les souches obtenues après chasse sur 5-FOA devraient donc présenter, au minimum, le phénotype correspondant à la mutation du second motif **$Cys_2HisCys$** – phénotype de croissance lente important – ce qui n'est pas le cas. Nous avons testé les clones obtenus sur milieu de sélection, pour la présence du plasmide pFL36 – marqueur *LEU2*. Le phénotype des souches parentales (YJG48, YJG49, YJG50) a également été vérifié sur milieu complet à 24°C, 30°C et 37°C (**Figure 2.20D**). Le phénotype de la souche portant l'allèle ***pcf11-18*** (C421S, C424S) est en accord avec les résultats précédents. Cependant, le phénotype de la souche portant l'allèle ***pcf11-19*** (C564S, C567S) n'est pas conforme à celui décrit précédemment, pouvant expliquer les résultats obtenus sur 5-FOA. La souche parentale utilisée possède un plasmide portant le marqueur *LEU2*. Nous pouvons penser que certains événements de suppression ont permis de rétablir un phénotype de croissance sauvage. Cela ne peut cependant pas expliquer les résultats obtenus pour les souches portant les plasmides ***clp1-16*** et ***clp1-17***. L'ensemble de ces expériences pour l'allèle ***pcf11-19*** (C564S, C567S) doit être reproduit.

En conclusion, nous avons pu mettre en évidence une aggravation du phénotype de la souche ***pcf11-18*** (C421S, C424S), en combinant les mutations du premier motif en doigt de zinc **Cys_2His_2** avec les allèles ***clp1-16*** (G135R, L341S), ***clp1-17*** (G335S, L341S) ou ***clp1-19*** (L341S). Les différentes mutations de Clp1 (G335S et L341S) sont situées au niveau des zones d'interaction entre ces deux protéines. Cela pourrait modifier les interactions de Clp1 et Pcf11 avec les autres partenaires du complexe, en modifiant par exemple la flexibilité de la partie C-terminale de la protéine, augmentant l'interaction avec un autre partenaire – de la même manière que la mutation $clp1^{D161A}$ augmente l'affinité de Clp1 pour Rna14 (Holbein et al., 2011). Cela conférerait une conformation « bloquée » au dimère Pcf11-Clp1. Aucune information concernant le second motif **$Cys_2HisCys$** n'a pu être obtenue.

D.2.3. Assemblage du complexe CF IA

Dans le complexe CF IA, Clp1 interagit principalement, et de manière très étroite, avec Pcf11 (Gross and Moore, 2001a; Uetz et al., 2000). Pcf11 interagit quant à lui avec l'hétéro-tétramère Rna14/Rna15. Certaines modifications peuvent abolir cette interaction, et

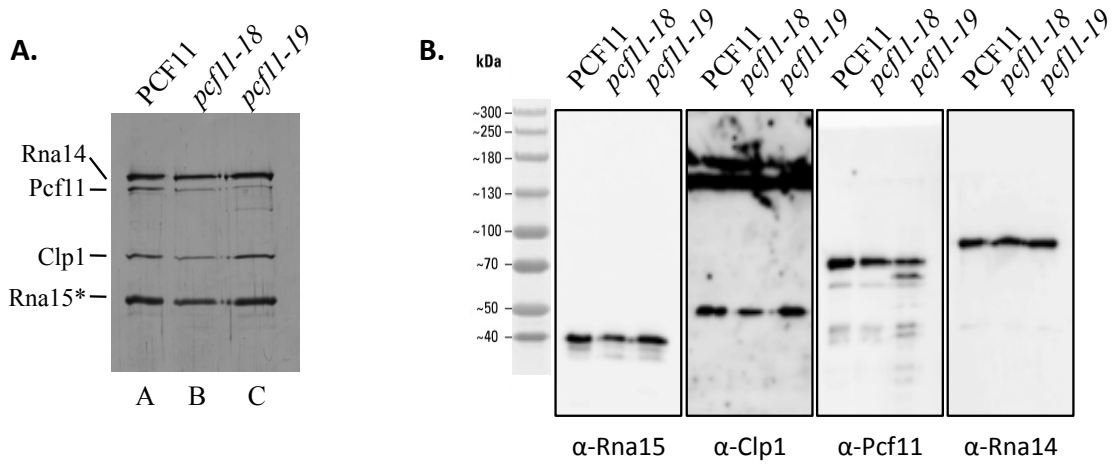


Figure 2.21 : L'assemblage du complexe CF IA n'est pas modifié par l'insertion des mutations au niveau des deux motifs en doigt de zinc de Pcf11. **(A.)** Purification des facteurs du complexe CF IA purifiés par TAP-tagging sur la sous-unité Rna15 (*). Les facteurs sont séparés par SDS-PAGE et coloré au nitrate d'argent. ***A.*** témoin positif portant une version sauvage de *PCF11*. ***B.*** *pcf11-18 :* mutant ponctuel (C421S, C424S) au niveau du premier motif en doigt de zinc. ***C.*** *pcf11-19 :* mutant ponctuel (C564S, C567S) au niveau du second motif en doigt de zinc. **(B.)** Analyse par Western Blot des différentes sous-unités du complexe CF IA d'une souche exprimant une version sauvage de *PCF11* (*à gauche*) et des souches exprimant les allèles *pcf11-18* (C421S, C424S) et *pcf11-19* (C564S, C567S) (*à droite*). Le poids moléculaire apparent des protéines est estimé par rapport au marqueur de taille (*à gauche*) en kDa. Les anticorps utilisés sont indiqués ; la quantité de protéines analysée est normalisée par rapport à Rna15.

les deux partenaires ne sont plus retrouvés lors de la co-purification par *TAP-tagging* sur la sous-unité Rna15 (Haddad et al., 2012). Bien que la région d'interaction entre Pcf11 et Clp1 soit caractérisée (Noble et al., 2007), on peut se demander si l'un ou les deux motifs en doigt de zinc ne participent pas à l'interaction entre Pcf11/Clp1 de manière directe ou indirecte, permettant ainsi de maintenir intact l'assemblage du complexe.

De la même manière que pour l'étude du 2nd NTD, nous avons décidé de purifier le complexe CF IA, par *TAP-tagging* sur la sous-unité Rna15. Pour cela, les souches correspondant aux mutants de *PCF11* obtenus précédemment ont été croisées avec la souche YSD12 ([p*NOP1*]-TAP::*RNA15-TRP1*-Kl) disponible au laboratoire. Après sporulation, dissection des tétrades et sélection, le phénotype des souches obtenues est vérifié sur YPD(A). La purification des facteurs natifs est réalisée sur 2 litres de culture. Les résultats obtenus sont présentés dans la **Figure 2.21A**. Pour le mutant ***pcf11-18*** (C421S, C424S), la coloration à l'argent (**Figure 2.21A**) ainsi que l'analyse par Western blot dirigé contre les quatre sous-unités du complexe (**Figure 2.21B**) nous permettent de conclure que l'assemblage de CF IA n'est pas perturbé par l'insertion des mutations au niveau du premier motif en doigt de zinc **Cys_2His_2**. Pour le mutant ***pcf11-19*** (C564S, C567S), les quatre sous-unités du complexe sont détectées (**Figure 2.21A et B**). Rna15 et Rna14 sont détectés à des niveaux comparables à ceux de la souche portant une version sauvage de *PCF11*. Cependant plusieurs différences sont visibles. Tout d'abord, le rapport entre Clp1 et Pcf11 est modifié : Clp1 est détectée en plus grande quantité par rapport à Pcf11 dans sa forme pleine taille. Ensuite, le profil obtenu pour Pcf11 est différent du profil de la souche sauvage : deux formes sont détectées, une forme pleine taille (72kDa) une forme de plus faible poids moléculaire (≈ 60kDa). Ces deux formes sont détectées à des intensités comparables.

En conclusion, l'assemblage du complexe CF IA n'est pas perturbé par l'insertion au niveau du premier motif (***pcf11-18*** – C421S, C424S) ni du second motif (***pcf11-19*** – C564S, C567S). L'effet potentiel de l'un de ces deux motifs est mineur, les régions d'interaction entre les différentes sous-unités du complexe CF IA ayant été identifiées. Les facteurs issus de la souche ***pcf11-19*** (C564S, C567S) présentent modification du rapport entre Clp1 et Pcf11 (*TABLEAU 2.2*).

D.2.4. Test d'activité de maturation de l'extrémité 3'

Nous avons ensuite étudié l'effet des mutations des domaines en doigt de zinc sur la maturation de l'extrémité 3' des ARN messagers. De la même manière que précédemment, nous avons choisi d'analyser les activités de clivage et de polyadénylation à partir de facteurs CF IA purifiés par *TAP-tagging*. Contrairement à l'étude du 2nd NTD, les complexes CF IA et CPF ne sont pas co-purifiés à partir d'une même souche. Le complexe CF IA est purifié à partir des souches construites précédemment, par croisement avec la souche YSD12 ([p*NOP1*]-TAP::*RNA15-TRP1*-Kl). Le complexe CPF est purifié par une étiquette située sur la sous-unité Fip1, à partir de la souche YJG35, disponible au laboratoire. Cette souche porte une version chromosomique mutée de *RNA15* (***rna15-1***) ; les complexes CF IA endogènes pouvant éventuellement co-purifier sont inactifs en test d'activité de maturation 3'. De plus, la purification du complexe CPF est réalisée en forte salinité (300mM NaCl) afin de minimiser la co-purification de CF IA. L'activité spécifique associée de chaque facteur est ensuite analysée.

Les activités de clivage sont tout d'abord étudiées à 30°C en présence de l'ARN substrat CYC1 radio-marqué et de Hrp1. Cet ARN substrat correspond à la région 3' non traduite de l'ARN du gène *CYC1*, codant pour l'iso-1-cytochrome C. Les résultats obtenus sont présentés dans la **Figure 2.22A**. En présence des facteurs CF IA purifiés à partir d'une souche portant une version sauvage de *PCF11*, le précurseur est clivé au site naturel poly(A) pour former deux fragments aux tailles attendues (186 et 116 nucléotides). Avec le complexe CF IA purifié à partir du mutant ***pcf11-18*** (C421S, C424S), l'ARN précurseur est clivé mais avec une plus faible efficacité en comparaison des facteurs sauvages. Les mutations (C564S, C567S) de la souche ***pcf11-19*** abolissent complètement les activités de clivage à 30°C.

Les activités de polyadénylation sont testées indépendamment du clivage, en présence de l'ARN *CYC1* pré-clivé à son site naturel, et des protéines Hrp1 et Pab1. Si l'activité de polyadénylation est efficace, on observe l'ajout d'environ 70 résidus d'adénosine à l'extrémité 3'-hydroxyle du précurseur. Les résultats obtenus sont présentés dans la **Figure 2.21B**. En comparaison aux facteurs CF IA^{PCF11}, aucune activité de polyadénylation n'est détectée pour chacun des deux mutants testés, ***pcf11-18*** (C421S, C424S) et ***pcf11-19*** (C564S, C567S). Ces résultats sont en accord avec les résultats obtenus lors de l'étude d'autres allèles des sous-unités de CF IA ; les mutants qui présentent un défaut de clivage présentent également un défaut de polyadénylation.

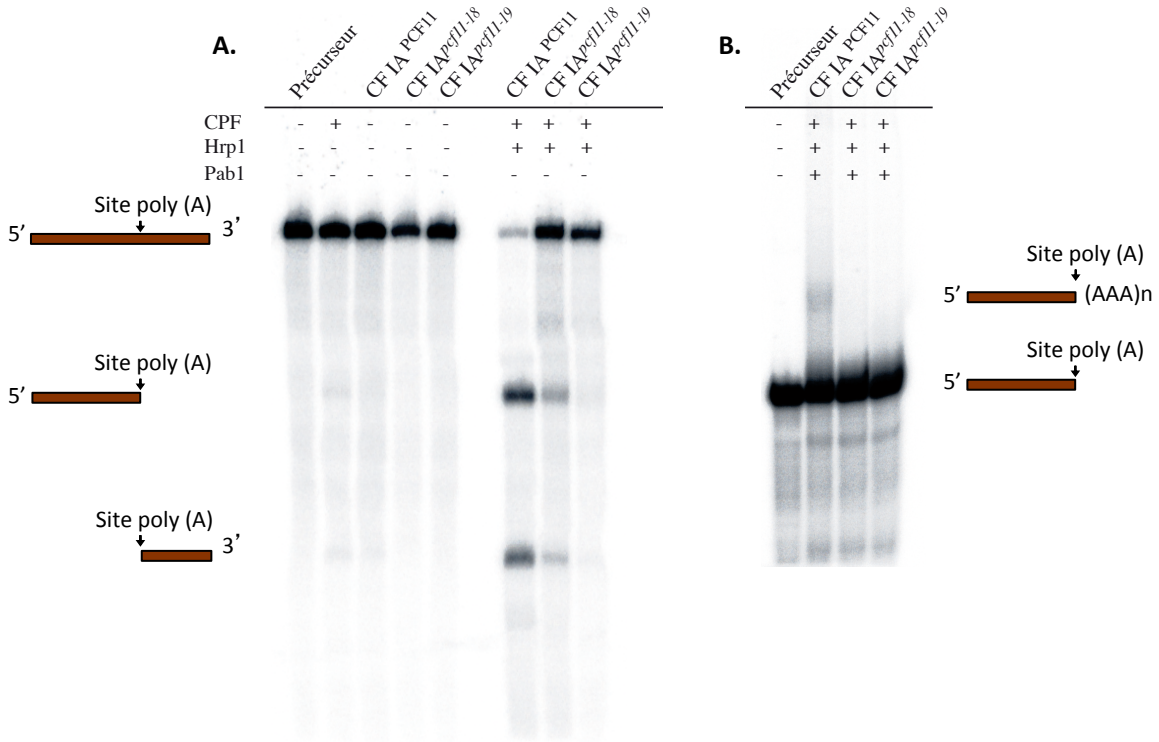


Figure 2.22 : Les motifs en doigt de zinc de Pcf11 sont nécessaires aux activités de maturation de l'extrémité 3' des ARN messagers.

(A.) Essai de reconstitution des activités de clivage en présence de l'ARN CYC1 radiomarqué. Les produits de la réaction sont séparés sur gel de polyacrylamide 6% (acrylamide/bisacrylamide 19 :1) / TBE 1X / 8,3M urée et révélés par autoradiographie. L'organisation des fragments obtenus est présentée *à droite*. L'ajout (+) ou l'absence (-) des protéines nécessaires à la réaction sont indiquées au dessus de chaque ligne. Hrp1 : 50ng.

(B.) Essai de reconstitution des activités de polyadénylation en présence de l'ARN pré-clivé, pré-CYC1, radiomarqué. Les produits de la réaction sont séparés sur gel de polyacrylamide 6% (acrylamide/bisacrylamide 19 :1) / TBE 1X / 8,3M urée et révélés par autoradiographie. L'organisation des fragments obtenus est présentée *à gauche*. L'ajout (+) ou non (-) des protéines nécessaires à la réaction sont indiquées au dessus de chaque ligne. Hrp1 : 50ng. Pab1 : 140ng

En conclusion, les deux mutants disponibles dans l'étude des motifs doigt de zinc de Pcf11 présentent une modification des activités de maturation de l'extrémité 3' des ARN messagers (*TABLEAU 2.2*). La mutation au niveau de premier motif en doigt de zinc **Cys_2His_2** (***pcf11-18*** – C421S, C424S) conduit à une forte diminution de l'activité de clivage et abolit la polyadénylation. La mutation du second motif **$Cys_2HisCys$** (***pcf11-19*** – C564S, C567S) est plus radicale et abolit le clivage et la polyadénylation *in vitro*, permettant d'expliquer le phénotype marqué observé à 24°C et 30°C.

D.2.5. Terminaison de la transcription par l'ARN polymérase II

La partie C-terminale de la protéine a été montrée comme indépendante de la terminaison de transcription par l'ARN polymérase II (Sadowski et al., 2003). Selon cette description, il semble peu vraisemblable que les motifs en doigt de zinc soient impliqués dans la terminaison de transcription. Nous avons analysé la terminaison de transcription par Northern blot et par essai de protection contre la digestion par la S1 Nucléase. À cause des problèmes de détections, les analyses par Northern blot n'ont pas permis d'obtenir des résultats exploitables et de savoir si les deux domaines en doigt de zinc pouvaient jouer un rôle dans la terminaison de transcription. Les essais de protection contre la S1 Nucléase ont cependant permis d'obtenir des résultats analysables.

D.2.5.1. Essai de protection contre la digestion par la S1 Nucléase

Nous nous sommes intéressé à deux snoRNAs : **snR13** (boite C/D – **Figure 2.23A**) et **snR5** (boite H/ACA – **Figure 2.24**).

Les résultats obtenus pour **snR13** sont présentés dans la **Figure 2.23B**. La souche portant une version sauvage de *PCF11* (ligne E et F) et la souche délétée pour le 2nd NTD (ligne G et H) sont utilisées comme témoins. Comme observé précédemment, dans le cas de la souche portant une version sauvage de *PCF11*, seul un fragment protégé de petite taille (75 nucléotides) est détecté en plus de la sonde présente en excès. Pour la souche **Δ(141-233)**, deux fragments protégés sont détectés – un de petite taille et un d'une taille proche de celle de la sonde (450 nucléotides). Pour les ARNs de souches portant les allèles ***pcf11-18*** (C421S, C424S) et ***pcf11-19*** (C564S, C567S) seul un fragment protégé de petite taille (75 nucléotides) est détecté, de faible intensité en comparaison aux résultats obtenus pour la souche portant une version sauvage de *PCF11*. D'autres essais de protection contre la digestion par la S1 Nucléase confirme la détection moins importante de snoRNAs matures. Cependant, aucune

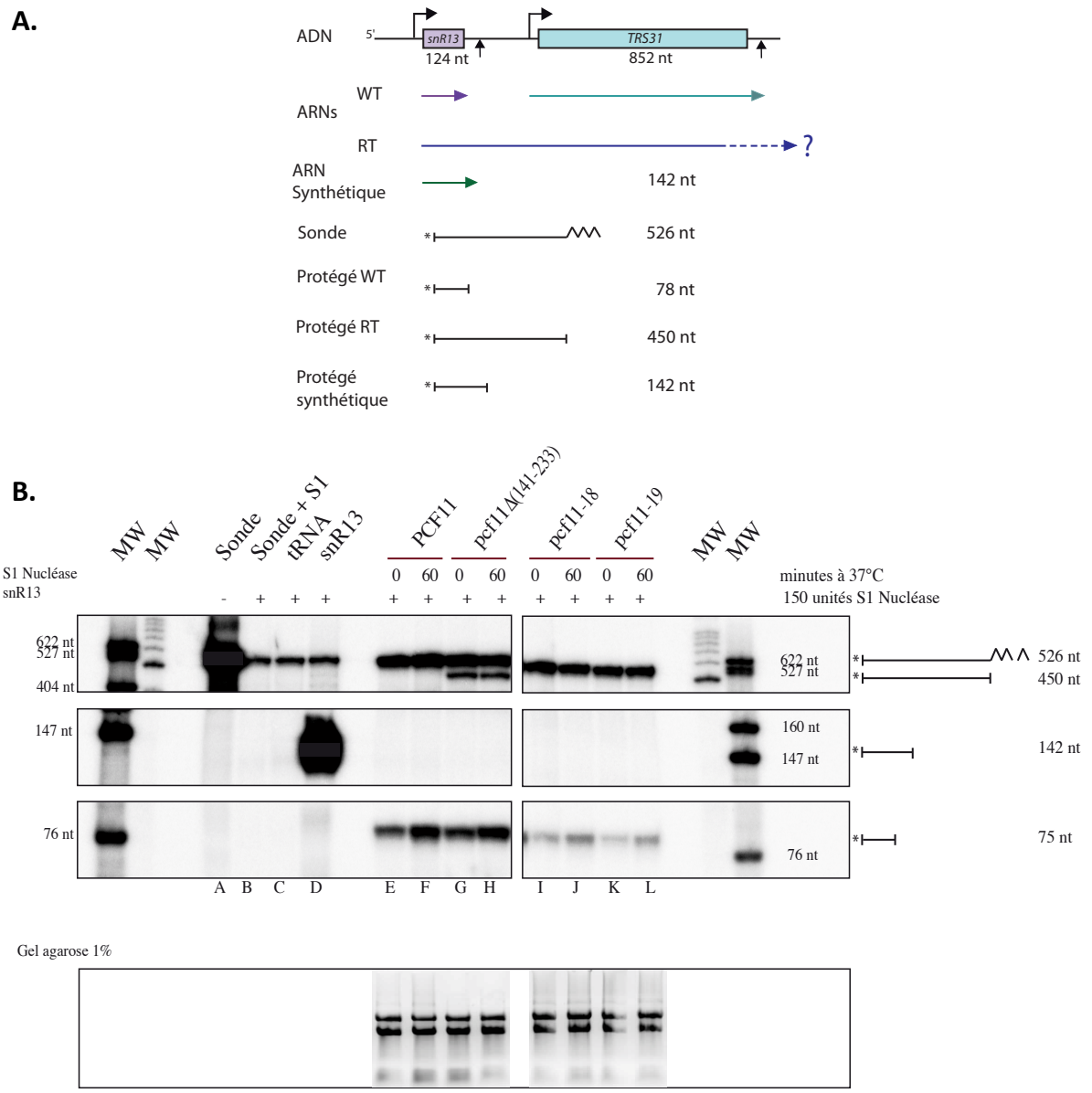


Figure 2.23 : Les motifs en doigt de zinc de Pcf11 ne sont pas impliqués dans la terminaison de la transcription des snoRNAs. **(A.)** Représentation schématique de la zone étudiée. snR13 (124 nucléotides) est représenté en **violet**, TRS31 est représenté en **bleu** (852 nucléotides). Les ARNs de plus grande taille illustrant un défaut de terminaison de transcription sont indiqués en **bleu**. L'ARN synthétique utilisé comme témoin positif est représenté en **vert**. Le marquage de la sonde à l'extrémité 5' est représenté par un * ; la séquence non spécifique est représenté par des zigzags. **(B.)** 30ng de sondes radiomarquées à l'extrémité 5' sont incubés seule (ligne *A*), ou en présence de 150 unités de S1 Nucléase et des ARNs étudiés (ligne *B à R*). *B* : sonde seule. *C* : 2µg de ARNt de levure. *D* : 2µg de transcrit synthétique snR13. L'analyse est effectuée sur 30µg d'ARNs totaux après 0 et 60 minutes d'incubation à 37°C. Les fragments protégés de la digestion par la S1 Nucléase sont séparés sur gel de polyacrylamide 8% et révélés par autoradiographie. Les fragments obtenus sont représentés *à droite*. *E et F* : ARNs totaux extraits d'une souches portant une version sauvage de *PCF11*. *G et H* : ARNs totaux extraits de la souche délétée pour le 2nd NTD Δ(141-233). *I et J* : ARNs totaux extraits de la souche portant l'allèle *pcf11-18* (C421S, C424S). *K et L* : ARNs totaux extraits de la souche portant l'allèle *pcf11-18* (C564S, C567S). *MW* : marqueurs de taille en nucléotides.

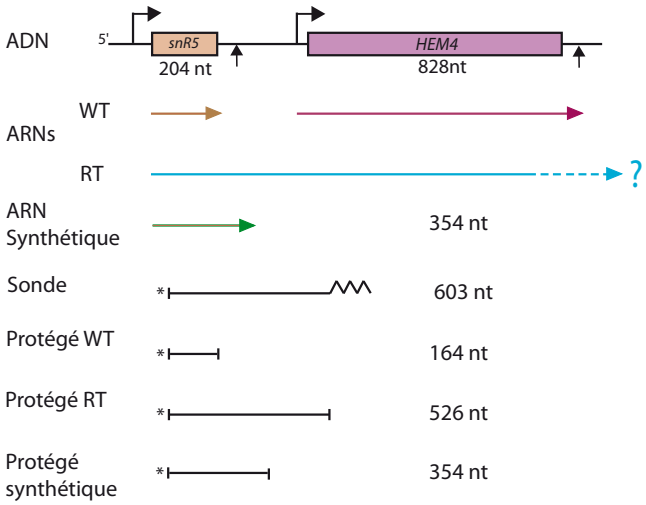


Figure 2.24 : Représentation schématique de la zone étudiée en essai de protection contre la S1 Nucléase. snR5 (204 nucléotides) est représenté en **marron**, TRS31 est représenté en **rose** (828 nucléotides). Les ARNs de plus grande taille illustrant un défaut de terminaison de transcription sont indiqués en **bleu**. L'ARN synthétique utilisé comme témoin positif est représenté en **vert.** Le marquage de la sonde à l'extrémité 5' est représenté par un * ; la séquence non spécifique est représenté par des zigzags.

bande illustrant un défaut de terminaison de transcription n'est détectée après 60 minutes à 37°C. La mutation des motifs en doigt de zinc dans la région C-terminale de Pcf11 ne modifie pas la terminaison de transcription des snoRNAs. La détection en plus faible quantité de snoRNAs matures pourrait être causée par le défaut de maturation de l'extrémité 3' des ARN messagers et illustrer un défaut d'assemblage/recrutement des sous-unités du complexe.

En conclusion, les essais de protection contre la digestion par la S1 Nucléase réalisés sur un snoRNA (boite C/D), nous permettent d'affirmer que les deux motifs **Cys_2His_2** (***pcf11-18*** – C421S, C424S) et **Cys_2HisCys** (***pcf11-19*** – C564S, C567S) situés dans la région C-terminale de Pcf11 n'influent pas sur la terminaison de transcription des snoRNAs. Ces résultats sont en accord avec les résultats d'études précédentes mais doivent être validés par l'étude de **snR5.** Les essais de protection contre la digestion par la S1 Nucléase contre des ARNs codants (*CYC1*) n'ont pour l'instant pas permis d'obtenir de résultats exploitables.

D.3. Discussion et Perspectives

Dans la deuxième partie de cette étude, nous nous sommes intéressés aux motifs liant le zinc présents dans la région C-terminale de Pcf11 (Amrani et al., 1997b). Parmi les protéines du complexe de polyadénylation, Yth1, Mpe1 ou encore Nab2 possèdent des motifs liant le zinc, de type **CCCH**. Les protéines présentant ces motifs de type **CCCH**, considérés comme peu fréquents, sont impliquées dans l'interaction avec l'ARN, la localisation des ARNs et leur stabilité. Les deux motifs retrouvés dans la région C-terminale de Pcf11 présentent une organisation différente. Ces deux motifs sont situés entre les résidus 420-450 et 560-610 et encadrent le domaine d'interaction de Pcf11 avec Clp1. Le premier motif est de type **Cys_2His_2**, replié en ββα ; le second motif est de type **Cys_2HisCys**. Les analyses par spectrométrie de masse et par absorption atomique réalisées par nos collaborateurs à l'IECB (A. Dupin, équipe de S. Fribourg, Bordeaux et J. Stojko, équipe de S. Cianférani, Strasbourg) ont montré que, en accord avec leur structure, ces deux motifs interagissent avec un atome de zinc. Cependant, aucune fonction n'est pour l'instant attribuée à ces deux motifs, conservés dans tout le règne eucaryote. Nous avons décidé de répondre à cette question par génétique inverse. Trois mutants ont été obtenus : deux mutants au niveau de l'un des deux motifs et un troisième mutant comprenant des mutations au niveau des deux motifs. Les phénotypes et les résultats obtenus sont récapitulés dans le *TABLEAU 2.2*.

TABLEAU 2.2 : Récapitulatif des différents mutants obtenus dans l'étude des motifs en doigt de zinc et des phénotypes associés. Les résultats obtenus lors de l'analyse de l'assemblage, des essais de maturation de l'extrémité 3' et de terminaison de transcription sont indiqués. OK : résultats comparable à ceux obtenus pour la souche portant une version sauvage de *PCF11*.

Constructions	**Phénotype**	**CF IA**	**3' processing**		**Terminaison transcription**			
			Clivage	**Polyadénylation**	**Northern CYH2**	**Northern TRS31**	**S1 Nuclease snR13**	**S1 Nuclease snR5**
PCF11	Sauvage à 24°C, 30°C et 37°C - Temps de génération 1h39 - Pas de sensibilité 2% formamide	OK	OK	OK	OK	OK	OK	OK
pcf11-18 (C421A, C424A)	Sauvage à 24°C et 30°C, thermosensibilité à 37°C - Sensible 2% formamide	OK - Clp1/Pcf11	Défaut clivage	Pas de polyadénylation		OK	OK	
pcf11-19 (C564A, C567A)	Croissance lente à 24°C et 30°C, thermosensibilité à 37°C - Temps de génération 5h30 - Sensible 2% formamide	OK - Clp1/Pcf11	Pas clivage	Pas de polyadénylation		OK	OK	
pcf11-20 (C421A, C424A, C564A, C567A)	Létal							

La mutation de deux des quatre résidus formant le premier motif en doigt de zinc (C421S, C424S – *pcf11-18*) induit une thermosensibilité à 37°C. Les mutations C564S et C567S (*pcf11-19*) au niveau du second motif induisent un phénotype de croissance lente à 24°C et 30°C ainsi qu'une thermosensibilité à 37°C. Aucun des autres mutants ponctuels de Pcf11 précédemment décrits dans la littérature (Amrani et al., 1997b; Sadowski et al., 2003) ne présentent de phénotype comparable à celui obtenu en présence de l'allèle *pcf11-19*. De plus, la mutation des résidus Cystéine des motifs en doigt de zinc présents dans d'autres protéines du complexe de polyadénylation ne conduit pas à un phénotype aussi marqué. Les mutations C182G et C185G au niveau de Mpe1, l'allèle *nab2-21* (C437S) et les mutations C415A, C437A et C458A au niveau de Nab2 conduisent à un phénotype « *cold-sensitive* » à 16°C et une croissance comparable à celle d'une souche sauvage à 24°C et 30°C chez *S. cerevisiae* (Hector et al., 2002; Lee and Moore, 2014). Un phénotype « *cold-sensitive* » illustre souvent un défaut d'assemblage de complexe multi-protéique (Hampsey, 1997). Chez l'Homme, les mutations ponctuelles des résidus conservés des motifs liant le zinc conduisent au développement de tumeurs ou de cancers, comme les mutations C305F et C319R de Mdm2 (Lindstrom et al., 2007). Bien qu'une corrélation entre gravité du phénotype et importance des motifs mutés ne puisse être établie chez la levure, les résultats obtenus dans l'étude des motifs liant le zinc de Pcf11 indiquent que ces motifs sont importants pour la viabilité cellulaire chez *S. cerevisiae,* comme les motifs ZF2 et ZF4 de Yth1. La délétion de l'un de ces deux motifs est létale chez la levure (Tacahashi et al., 2003). Cette hypothèse est soutenue par le fait que la mutation des résidus Cystéines des deux motifs en doigt de zinc (*pcf11-20*) soit létale chez *S. cerevisiae*.

La différence de phénotype observée entre les deux mutants peut traduire une indépendance de fonction de ces deux motifs. De nombreux exemples sont décrits dans la littérature, où chaque motif en doigt de zinc présent au sein d'une même région est impliqué dans une fonction différente. Yth1 est une sous-unité du sous-complexe PF I et est formée de cinq motifs liant le zinc de type **CCCH**. Chaque motif présente une fonction différente : les motifs ZF2 et ZF4 sont nécessaires pour l'interaction avec l'ARN, les motifs ZF4 et ZF5 permettent l'interaction avec Fip1, les motifs ZF1 et ZF4 avec l'endonucléase Ysh1/Brr5 et les motifs ZF1 et ZF2 sont nécessaires pour le clivage du pré-ARNm (Tacahashi et al., 2003). Plus éloignée de la polyadénylation des ARNs messagers, chez la souris, la protéine wig-1, dont l'expression est dépendante de p53, contient trois motifs liant le zinc de type **Cys_2His_2**. Le motif ZF1 est essentiel pour l'interaction avec l'ADN double brin ; les motifs ZF2 et ZF3 ne sont pas requis (Mendez-Vidal et al., 2002). La complémentation en *trans* de l'allèle

pcf11-18 avec la partie C-terminale de la protéine portant les mutations de l'allèle *pcf11-19* et inversement, nous permettrait de déterminer si ces deux mutations sont situées sur une même unité fonctionnelle.

En accord avec les résultats obtenus précédemment ayant permis de définir les domaines d'interaction avec les différentes sous-unités du complexe CF IA, la purification par *TAP-tagging* des facteurs a permis de montrer que l'assemblage global du complexe CF IA n'est pas perturbé par l'insertion de mutations au niveau de l'un de ces deux motifs. Cependant, certaines modifications mineures peuvent être observées ; l'interaction entre l'hétéro-dimère Clp1/Pcf11 et l'hétéro-tétramère Rna14/Rna15 semble légèrement perturbée par les mutations au niveau des deux motifs en doigt de zinc. Les modifications observées pourraient être un effet indirect de la mutation des résidus Cystéines ou un effet direct ; les motifs en doigt de zinc pourraient alors permettre la stabilisation de l'interaction entre Pcf11 et Clp1.

Aucune fonction n'est attribuée aux motifs liant le zinc de Pcf11, et nous ne savons pas si ces deux motifs sont impliqués dans une même fonction ou possèdent un rôle indépendant. Les premiers résultats obtenus ont permis de montrer que les motifs en doigt de zinc sont tous les deux impliqués dans la maturation de l'extrémité 3' – clivage et polyadénylation – *in vitro* mais pas dans la terminaison de transcription des snoRNAs – analyse par essais de protection contre la digestion par la S1 Nucléase sur snR13. Ces résultats sont en accord avec les études précédentes définissant la région C-terminale de la protéine comme nécessaire et suffisante pour les activités de maturation de l'extrémité 3' des ARNs messagers et indépendante de la fonction de Pcf11 dans la terminaison de transcription (Sadowski et al., 2003). L'impact sur les activités de maturation peut être corrélé avec le phénotype observé. La mutation du premier motif en doigt de zinc diminue le clivage, en comparaison à la mutation du second motif où le clivage est complètement aboli pouvant permettre d'expliquer le phénotype plus marqué décrit pour ce mutant. Aucune activité de polyadénylation n'est détectée pour chacun des mutants. La fonction des motifs liant le zinc dans la maturation de l'extrémité 3' n'est pour l'instant pas connue, mais ils pourraient intervenir indépendamment dans des interactions protéines-protéines ou des interactions avec le pré-ARNm – comme les motifs liant le zinc dans la région C-terminale de Nab2 (Anderson et al., 1993) – ou être impliqués dans ces deux types de fonction. C'est par exemple le cas du motif ZF4, de type **CCCH** présent dans la région C-terminale de Yth1. Le motif ZF4 est

impliqué aussi bien dans l'interaction avec Fip1 et que dans l'interaction avec le pré-ARNm (Tacahashi et al., 2003). De manière surprenante, le motif ZF4 n'est pas requis pour l'étape de clivage. Lors du clivage, Yth1 via ZF4 permettrait l'interaction avec l'ARN et le positionnement puis le recrutement de Fip1 et le recrutement de Pap1 (Tacahashi et al., 2003).

L'assemblage du complexe CF IA n'étant pas perturbé, les motifs en doigt de zinc permettraient le lien avec les sous-unités du complexe CPF, partenaires de Pcf11. Pcf11 interagit avec : Yhh1/Cft1 (Kyburz et al., 2003) et Mpe1 (Vo et al., 2001) sont requis pour les deux étapes de clivage et de polyadénylation ; Ysh1/Brr5 et Ydh1/Cft2 sont nécessaires pour l'étape de clivage ; Pta1 permet le recrutement du sous-complexe CF II. Seul Swd2, sous-unités du sous-complexe APT et interagissant avec Pcf11 (Cheng et al., 2004) n'est pas requis pour les étapes de clivage et de polyadénylation. Il serait peu probable que le phénotype observé soit dû à un défaut de recrutement de cette sous-unité. En considérant le phénotype observé et les résultats des essais de maturation de l'extrémité 3', seul un défaut d'interaction avec Yhh1/Cft1, Mpe1 ou Pta1 expliquerait l'inhibition du clivage et de la polyadénylation. À court terme, les interactions entre les différents partenaires peuvent être testées par *pull-down* à partir de protéines recombinantes ou de complexes CF IA natifs purifiés chez *S cerevisiae* en présence des différents allèles de *PCF11*.

L'export des mRNPs vers le cytoplasme est dépendant du clivage et de la polyadénylation chez la levure ; les mutations de Rna14, Rna15 et Pcf11 conduisent à un défaut d'export (Brodsky and Silver, 2000; Hammell et al., 2002; Johnson et al., 2009). Les facteurs impliqués dans l'export sont recrutés co-transcriptionnellement, comme Yra1 ou Npl3. Johnson *et al.,* (Johnson et al., 2009) ont montré que Pcf11 permettait le recrutement de Yra1/Aly par interaction avec la région 420 à 608 comprenant les deux motifs en doigt de zinc et le domaine d'interaction avec Clp1. De plus, le recrutement de Yra1 est perturbé en présence d'allèles portant des mutations au niveau de la région C-terminale de la protéine (allèles *pcf11-2* ((Amrani et al., 1997b) – E232G, D280G, C424R, S538G, F562S, S579P) et *pcf11-9* ((Amrani et al., 1997b) – A66D, S190P, R198G, R227G, E354V, K435V)). L'interaction entre Yra1 et Pcf11 est conservée chez l'Homme - Yra1/Pcf11 (420-608) et Aly/hPcf11 (1342-1487) ce qui est en accord avec la conservation de ces motifs dans le règne eucaryote. L'hypothèse proposée est qu'au moins l'un des motifs liant le zinc de Pcf11 permettrait l'interaction et le recrutement de Yra1 ; l'autre motif liant le zinc serait alors impliqué dans la maturation de l'extrémité 3'. Pour répondre à cette question, nous souhaitons

étudier l'interaction entre Pcf11 et Yra1 par *pull down*, à partir de protéines recombinantes et de facteurs natifs, par GST-Yra1 et/ou His-Pcf11.

Johnson *et al.,* en 2011 (Johnson et al., 2011) ont proposé un modèle dans lequel Yra1 permet de masquer le site d'interaction entre Pcf11 avec Clp1, et ainsi moduler l'assemblage du complexe CF IA (**Figure 2.25**). Plusieurs arguments sont en faveur de cette hypothèse : (i) l'ajout de Yra1 conduit à une inhibition du clivage et de la polyadénylation, (ii) la délétion de *YRA1* permet un recrutement plus précoce de Clp1 (iii) le recrutement de Pcf11/Clp1 est inversement proportionnel au recrutement de Yra1/Sub2. Si les motifs en doigt de zinc permettent l'interaction avec Yra1, le mécanisme de compétition entre Yra1 et Clp1 pour l'interaction avec Pcf11 décrit par Johnson *et al.,* ne refléterait qu'un problème d'accessibilité du domaine et non un réel mécanisme de compétition ayant une signification biologique.

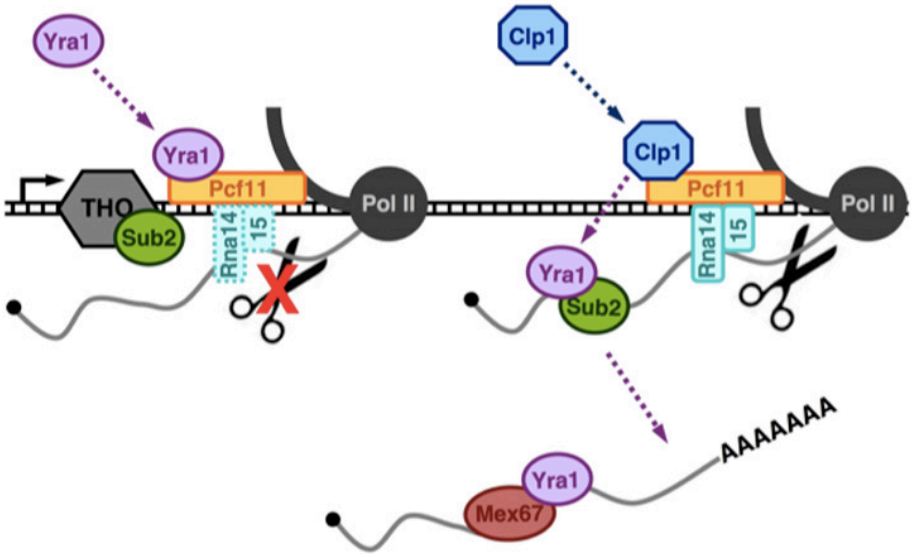


Figure 2.25 : Représentation du modèle d'action de Yra1 dans la maturation de l'extrémité 3'. D'après Johnson *et al.,* 2011. Yra1 (en **violet**) est recruté par interaction avec Pcf11 (en **orange**) lors de la phosphorylation du résidu SER[2] du CTD de l'ARN polymérase II. Le recrutement de Clp1 (en **bleu**) va permettre de libérer Yra1 qui va être recruté par Sub2 et interagir avec l'ARN pour permettre le recrutement de Mex67 et l'export de l'ARN. Le recrutement de Clp1 permet de rendre le complexe CF IA compétent et la maturation de son extrémité 3'.

CHAPITRE 2 : MISE EN EVIDENCE D'UNE ACTIVITE KINASE ASSOCIEE AUX FACTEURS DE MATURATION EN 3' DES ARNS MESSAGERS

CHAPITRE 2 : MISE EN EVIDENCE D'UNE ACTIVITE KINASE ASSOCIEE AUX FACTEURS DE MATURATION EN 3' DES ARNs MESSAGERS

A. INTRODUCTION

A.1. Organisation de Clp1

Clp1 est la dernière protéine du sous-complexe CF IA à avoir été identifiée. Chez *S. cerevisiae*, Clp1 est nécessaire aux activités de maturation de l'extrémité 3' et à la terminaison de transcription des ARNm (CYH2 ou ASC1) et des snoRNAs (Haddad et al., 2012; Holbein et al., 2011). Clp1 est associée à Rna14-Rna15 principalement via son interaction avec Pcf11 (Gross and Moore, 2001a; Noble et al., 2007) et interagit également avec différentes sous-unités du complexe CPF (Yhh1/Cft1, Ysh1/Brr5, Ssu72, Pfs2 et faiblement Ydh1/Cft2). Son organisation est présentée dans la **Figure 3.01A**. La structure de Clp1 a été décrite (Dupin and Fribourg, 2014; Noble et al., 2007) ; cette protéine est formée d'une région centrale et de deux régions, N-terminale et C-terminale, plus petites (**Figure 3.01C**).

La région N-terminale (18 à 102) est formée de sept feuillets β (β1 à β7) repliés en un domaine globulaire hydrophobe (**Figure 3.01B**). Ce domaine est faiblement conservé dans le règne eucaryote. Seuls les résidus 51 à 60 se repliant vers la région centrale de la protéine sont conservés. Le résidu K72 interagit avec le ribose de l'ATP. La région C-terminale (343 à 445) forme un domaine globulaire de sept feuillets β (β15 à β21) et de trois hélices α (α8 à α10) (**Figure 3.01B**). Les feuillets β20 et β21 sont formés de résidus conservés, impliqués dans l'interaction avec Pcf11 (L406, S342, V427).

La région centrale (102 à 343) est formée de sept feuillets β (β8 à β14) entourés par sept hélices α (α1 à α7) (**Figure 3.01B**). Cette structuration est caractéristique des ATPases de la classe SIMIBI (Leipe et al., 2002). La région centrale de Clp1 comprend quatre motifs caractéristiques des NTPases : le motif *P-loop*, les régions « *Switch I* » et « *Switch II* », et un motif « *base binding loop* ». Le motif *P-loop* ou motif Walker A est formé par la séquence GxxxxGKS/T (130-137), et permet l'interaction avec les résidus phosphates de l'ATP (Walker et al., 1982). Les motifs « *Switch I* » de séquence DxxQ (β9- β10 – 161-164), « *Switch II* » de séquence DTPS (β12 - α5 – 251-254) et « *base binding loop* » (308-312) interagissent également avec l'ATP et permettent l'interaction avec le cation divalent Mg^{2+}.

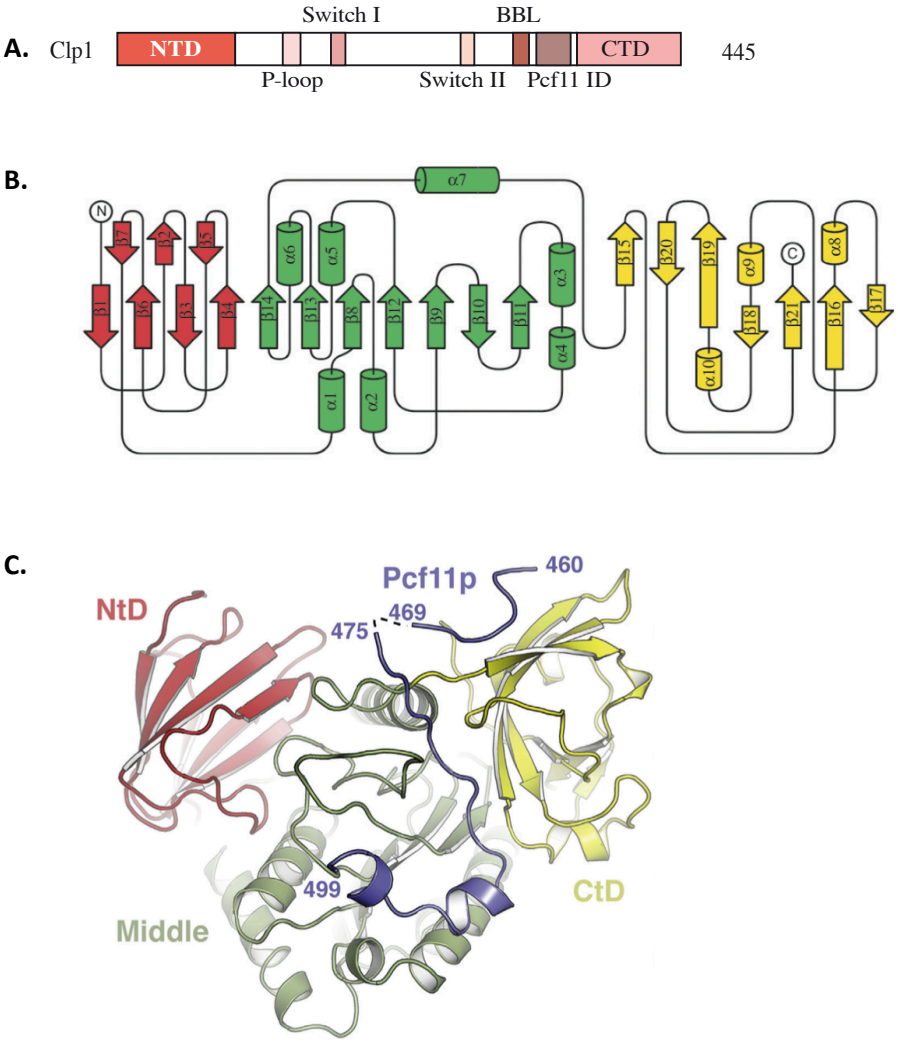


Figure 3.01 : Organisation de Clp1.

(A.) Représentation schématique de Clp1. Les différents domaines sont à l'échelle et leur taille est donnée en acides aminés. NTD (1-102) : domaine N-terminal. CTD (343-445) : domaine C-terminal. Le domaine central est représenté en blanc (102-343). Les différents domaines de la région centrale sont indiqués. P-loop (130-137) :motif P-loop/Walker A. Switch I : motif DxxQ (161-164). Switch II : motif DTPS (251-254). BBL : *base binding loop* (308-312). Pcf11 ID : domaine d'interaction avec Pcf11.

(B.) Représentation linéaire de l'organisation des différentes structure secondaires de Clp1. Les extrémité N-terminale (N) et C-terminale (C) sont indiqués. Le domaine N-terminal est représenté en **rouge**, le domaine central en **vert** et le domaine C-terminal en **jaune**. Les hélices α sont représentées par des cylindres. Les feuillets β sont représentés par des flèches. D'après Noble *et al.,* en 2007 (Noble et al., 2007).

(C.) Structure de Clp1 en présence de la région d'interaction avec Pcf11. Le domaine N-terminal est représenté en **rouge**, le domaine central en **vert** et le domaine C-terminal en **jaune**. Le domaine d'interaction avec Pcf11 (454-563) est représenté en **violet**. D'après Dupin et Fribourg, 2014 (Dupin and Fribourg, 2014).

L'interaction de Clp1 avec Pcf11 fait intervenir des résidus situés dans la région centrale de Clp1, comprenant un motif YFYG (α7 – 332-335) conservé. La mutation de l'ensemble de ces quatre résidus est létale chez *S. cerevisiae*.

A.2. Fonctions attribuées aux homologues de Clp1

Clp1 est une protéine conservée dans le règne eucaryote ainsi que des archées. Elle présente 26% d'identité avec hClp1, son homologue chez l'Homme (de Vries et al., 2000). L'organisation des protéines est également conservée. hClp1 contient un motif Walker A (interaction avec une molécule d'ATP) et Walker B en aval du motif Walker A. Avec hPcf11, hClp1 forme le complexe CF II_m et est essentielle pour les activités de maturation de l'extrémité 3'. hClp1 est une polynucléotide kinase spécifique de l'ARN (RPNK – « *RNA-specific polynucleotide kinase* ») ; à la différence de certaines PNK comme la T4PNK active sur différents substrats, l'activité des RNPK est restreinte aux étapes de maturation des ARNs. La structure de ceClp1, son homologue chez *C. elegans* (48%) en présence d'un substrat ARN, a récemment été décrite (Dikfidan et al., 2014). En 2009, PhoClp1, l'homologue de Clp1 chez *Pyrococcus horikoshii* a été décrit comme présentant une activité de RPNK, thermostable avec une activité optimale entre 55 et 85°C (Jain and Shuman, 2009).

hClp1 est impliquée dans l'épissage des pré-ARNt. Chez les vertébrés, deux voies d'épissage des ARNt sont décrites : une voie dite « *animal* » et une voie dite « *yeast-like* ». L'intron est clivé, permettant la formation d'un phosphate cyclique 2'-3' sur l'exon 5' et d'une extrémité 5'-OH sur l'exon 3'. La voie « *yeast-like* » fait intervenir trois activités portées par le complexe TSEN (« *tRNA Splicing endonuclease* »). (i) une activité de phosphodiestérase cyclique, pour permettre la formation d'une extrémité 2' libre sur l'exon 5' à partir du phosphate cyclique 2'-3' (ii) une activité de polynucléotide kinase – phosphorylation de l'extrémité 5'-OH de l'exon 3' (iii) une activité d'ARN ligase dépendante de l'ATP, pour joindre les deux exons de l'ARNt via le phosphate. Deux études indépendantes (Paushkin et al., 2004; Ramirez et al., 2008) ont montré que hClp1 permettait le transfert d'un phosphate sur l'exon 3' de l'ARN simple brin et que hClp1 était l'ARN kinase impliquée dans cette étape de maturation. Récemment, il a été montré qu'un défaut de cette activité d'ARN kinase, par mutation ponctuelle (R140H), conduit au développement de pathologies neuro-dégénératives graves chez l'homme (Karaca et al., 2014; Schaffer et al., 2014; Weitzer et al., 2015).

L'activité ARN kinase de hClp1 est également impliquée dans la maturation des siRNA (Weitzer and Martinez, 2007a). Pour être incorporé dans le complexe de « *silencing* » RISC, l'ARN guide doit être phosphorylé à son extrémité 5' pas une ARN kinase. Weitzer et Martinez, en 2007 (Weitzer and Martinez, 2007b), ont montré que hClp1 était l'ARN kinase responsable de la phosphorylation de l'ARN double brin. Ils ont également montré que hClp1 était actif sur l'ADN double brin.

Weitzer et Martinez (Weitzer and Martinez, 2007b) ont également proposé un rôle de l'activité ARN kinase de hClp1 dans la maturation de l'extrémité 3'. Dans ce modèle, hClp1 permettrait de maintenir l'extrémité 5'-P nécessaire pour l'activité de Xrn2.

Contrairement à ses homologues hClp1 ou PhoClp1, aucune activité kinase n'a pu être attribuée à la protéine Clp1 de levure seule. *In vitro*, Clp1 produite de manière recombinante ne présente pas d'activité de type polynucléotide kinase. De plus, la surexpression de Clp1 ne permet pas de complémenter les mutants présentant un défaut dans l'étape de phosphorylation des pré-ARNt. À l'inverse, l'expression de hClp1 ne permet pas de complémenter la croissance de la souche *clp1*Δ chez *S. cerevisiae*. Les activités de phosphodiestérase, de polynucléotide kinase et d'ARN ligase sont portées par la protéine Trl1 chez *S. cerevisiae* ; l'expression de hClp1 permet de restaurer le phénotype sauvage du mutant *trl1* (Ramirez et al., 2008). La mutation des résidus du site actif de hClp1 modifie son activité ; la mutation des résidus équivalents au niveau de la zone d'interaction avec l'ATP chez la levure (G135, K136 et T137) ne modifie pas les activités de Clp1 dans la maturation de l'extrémité 3'. La structure de la région centrale de Clp1, et la présence d'une molécule d'ATP laissent suggérer une fonction ATPase à cette protéine. Aucune activité d'hydrolyse de l'ATP n'a pu être attribuée à Clp1, en présence ou en absence de Pcf11 (Noble et al., 2007).

B. RESULTATS PRELIMINAIRES / HYPOTHESES

Dans ce contexte, nous avons fait l'hypothèse que la présence des autres partenaires des complexes CF IA et/ou CPF et d'une molécule d'ATP dans la poche de liaison de Clp1, peut conférer une activité de type protéine kinase à Clp1. L'association des différents partenaires conduirait à un changement de conformation de la protéine permettant ainsi l'activation de son activité kinase et la phosphorylation des protéines cibles.

Pour tester cette hypothèse, les complexes **CF IA** et **[CF IA+CPF]** purifiés chez *S. cerevisiae* sont incubés en présence de [γ^{32}P]-ATP (avec ou sans ATP non marqué),

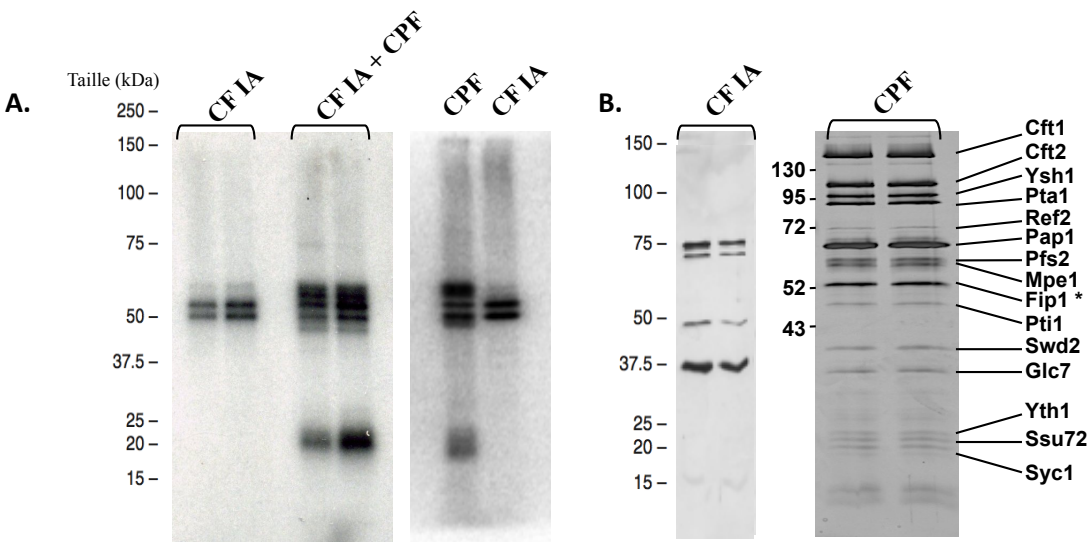


Figure 3.02 : Résultats préliminaires – Une activité kinase est retrouvée associée aux complexes CF IA et CPF.

(A.) Essai d'activité kinase des complexes CF IA, CPF et [CF IA+CPF]. Les facteurs purifiés sont incubés en présence de [γ^{32}P]-ATP puis séparés par SDS-PAGE. Le transfert du γ^{32}P est observé par autoradiographie. Le poids moléculaire apparent des protéines est estimé par rapport au marqueur de taille (*à gauche*) en kDa.

(B.) Séparation des protéines par SDS-PAGE et coloration au nitrate d'argent pour les complexes CF IA (*à gauche*) et CPF (*à droite*).

permettant de visualiser le transfert éventuel d'un phosphate $\gamma^{32}P$ sur une ou plusieurs protéines. Plusieurs allèles de *CLP1* ont été testés, parmi lesquels la version sauvage du gène *CLP1* et les allèles ***clp1-5*** (G135R ; G335S) et ***clp1-15*** (G135R, H267Q, L341S) – décrit par Haddad *et al.,* en 2012 (Haddad et al., 2012). Le résidu G135 est situé au niveau du motif P-loop et interagit avec le phosphate γ de l'ATP. La mutation G135R conduit à un réarrangement local de la zone d'interaction, et ne permet plus la fixation de l'ATP au niveau de la poche de liaison. Le résidu G335 est situé dans le motif YFYG participant à l'interaction avec Pcf11 ; le résidu L341S est situé à proximité de cette zone d'interaction. Les mutations des allèles ***clp1-5*** et ***clp1-15*** entraînent un défaut d'assemblage du complexe CF IA (Clp1 et Pcf11 ne sont plus associées entre elles ni à l'hétéro-tétramère Rna14/Rna15) et un défaut de maturation en 3' *in vitro*.

Dans le cas de facteurs purifiés à partir d'une souche sauvage, les premiers essais montrent le marquage de deux protéines et le marquage plus faible d'une troisième protéine pour le sous-complexe **CF IA** (**Figure 3.02A).** En présence des complexes **[CF IA+CPF]**, plusieurs protéines sont marquées, d'intensité comparable (**Figure 3.02A**). Pour les complexes **[CF IA+CPF]** en présence des allèles ***clp1-5*** et ***clp1-15***, plusieurs protéines sont marquées mais le marquage correspondant au sous-complexe CF IA n'est pas visible. Les allèles ***clp1-5*** et ***clp1-15*** sont décrits comme conduisant à un défaut d'assemblage du complexe CF IA ; Clp1 et Pcf11 ne sont plus associés à Rna14 et Rna15. Nous pouvons donc en déduire que Clp1 et Pcf11 ne sont plus retrouvées associées aux complexes **[CF IA+CPF]** en présence des allèles ***clp1-5*** et ***clp1-15***. Une activité de phosphorylation est donc observée en absence de Clp1. L'activité kinase ne peut, par conséquent, être attribuée à Clp1. Ce contrôle permet également d'affirmer que la protéine kinase co-purifie d'une manière indépendante de Pcf11 et Clp1. Des quantités croissantes du sous-complexe **CF IA** purifiées chez *S. cerevisiae* ou de la protéine recombinante Clp1 ont été ajoutées aux essais, sans modification du signal obtenu. De plus, l'analyse du sous-complexe **CPF** seul présente un signal comparable à celui obtenu après co-purification de deux sous-complexes en présence des allèles ***clp1-5*** et ***clp1-15*** (**Figure 3.02A**).

Nous avions émis l'hypothèse que Clp1 possède une activité de type protéine kinase en présence d'une molécule d'ATP dans sa poche de liaison et de ses partenaires des complexes CF IA et CPF. Les résultats préliminaires indiquent qu'une activité de type protéine kinase est bien retrouvée associée aux complexes mais qu'elle ne peut être attribuée à Clp1.

Nous avons ensuite cherché à caractériser l'activité kinase associée au complexe. Plusieurs paramètres ont ainsi été testés :

- Le signal est conservé en présence de RNase A – l'activité observée est une activité de protéine kinase et non d'ARN kinase.
- L'incubation en présence de 5mM d'EDTA et en absence de $MgCl_2$ (2mM) abolit le marquage – la kinase co-purifiant est Mg^{2+}- dépendante.
- L'ajout d'inhibiteurs de phosphatases (PhosSTOP – Roche) ne modifie pas le profil obtenu – l'activité de la kinase est indépendante des phosphatases présentes dans le complexe de polyadénylation. L'activité kinase observée n'est pas dépendante d'une déphosphorylation initiale de la protéine kinase.
- Le marquage est absent en présence de Lambda Protein Phosphatase – le signal obtenu correspond à la phosphorylation de protéines.

La préparation des complexes **CF IA**, **CPF** et **[CF IA+CPF]** a donc permis la co-purification d'une protéine kinase, Mg^{2+}-dépendante, phosphorylant vraisemblablement plusieurs cibles.

En supposant que les protéines phosphorylées dans les essais préliminaires fassent partie des complexes CF IA et CPF, plusieurs essais d'identification des cibles ont été réalisés à partir de **[CF IA+CPF]** par Western blot, avant et après phosphorylation. Cependant cela n'a pas permis d'identifier précisément les différentes cibles, mais plusieurs candidats potentiels peuvent être proposés (**Figure 3.02B**). Dans le cas de **CF IA**, seule une phosphorylation de Clp1 pourrait correspondre au signal obtenu (50kDa) ; la double bande pouvant traduire un état de phosphorylation différent. L'absence de signal correspondant au complexe CF IA en présence des allèles ***clp1-5*** ou ***clp1-15*** est en accord avec cette hypothèse. Dans le cas de **CPF**, sept protéines peuvent être proposées comme cibles en fonction de leur poids moléculaire : Pfs2 (53kDa), Fip1 (36kDa + 5kDa étiquette CBP), Mpe1 (50kDa), Pti1 (47kDa), Yth1 (26kDa), Syc1 (21kDa) et Ssu72 (23kDa). Il est également possible que les protéines phosphorylées ne fassent pas partie du complexe de polyadénylation. Le marquage observé pourrait alors correspondre à la phosphorylation de protéines co-purifiant en faible quantité lors de la préparation des complexes.

Les résultats préliminaires ont permis de mettre en évidence la présence d'une activité protéine kinase associée au complexe de polyadénylation des ARN messagers. À partir de ces résultats, les objectifs de ce projet sont (i) d'identifier la kinase qui co-purifie lors de la préparation des facteurs, par l'utilisation d'un analogue de l'ATP, le FSBA (5′-

fluorosulfonylbenzoyl 5′-adenosine) couplé à la spectrométrie de masse. Il a été montré que le FSBA modifie spécifiquement, de manière covalente, le site de fixation de l'ATP de nombreuses protéines kinases (Boettcher and Meister, 1980). Cette modification est ensuite détectable par spectrométrie de masse (ii) d'identifier les cibles de la protéine kinase, par immuno-détection des résidus phosphorylés, analyses par spectrométrie de masse ou encore par modification du site actif de la kinase. Cette dernière méthode permet de rendre la protéine kinase sélectivement sensible pour un analogue de l'ATP. Les essais kinase réalisés avec la protéine kinase mutée et l'analogue de l'ATP radiomarqué permettent le marquage spécifique des protéines cibles, pouvant être identifiées par analyse comparative puis par spectrométrie de masse. (iii) À terme, l'objectif est d'essayer de comprendre le rôle de ces modifications post-traductionnelles dans la maturation de l'extrémité 3'. En effet, des études récentes ont décrit la phosphorylation / déphosphorylation comme un des mécanismes de régulation de cette étape – l'activité phosphatase portée par Glc7, permettant de réguler l'action de Pta1 (Nedea et al., 2008; Schreieck et al., 2014).

C. RESULTATS

C.1. Résultats obtenus

Dans un premier temps, l'objectif a été de reproduire les résultats préliminaires afin d'obtenir une quantité de matériel suffisante pour le reste de l'étude – notamment afin d'identifier la protéine kinase. Les sous-complexes **CF IA** ou **[CF IA+CPF]** ont été purifiés par *TAP-tagging* à partir de différentes souches disponibles au laboratoire (**WT** : YSD12, YLM221, YSD39 et ***clp1-15*** : YLM206). Les facteurs obtenus sont analysés par SDS-PAGE et coloration au nitrate d'argent afin de vérifier l'efficacité de la purification et d'estimer leur quantité (**Figure 3.03A**). L'activité associée est ensuite testée en incubant les facteurs pendant 30 minutes à 30°C, en présence d'ATP radio-marqué. Les premiers résultats sont présentés **Figure 3.03B**. Les lignes A, B et C correspondent aux différentes combinaisons des sous-complexes analysés lors des expériences préliminaires – ces facteurs sont utilisés comme témoins positifs lors du reste de cette étude. Les lignes D à I correspondent aux complexes **CF IA** (CF IA^{WT} ou CF $IA^{clp1\text{-}5}$) nouvellement purifiés : un marquage de faible intensité est visible à un poids moléculaire similaire à celui observé précédemment. Nous avons réalisé de nouvelles purifications des sous-complexes **CF IA**, **[CF IA+CPF]** et **CPF**, sans réussir à reproduire les résultats préliminaires.

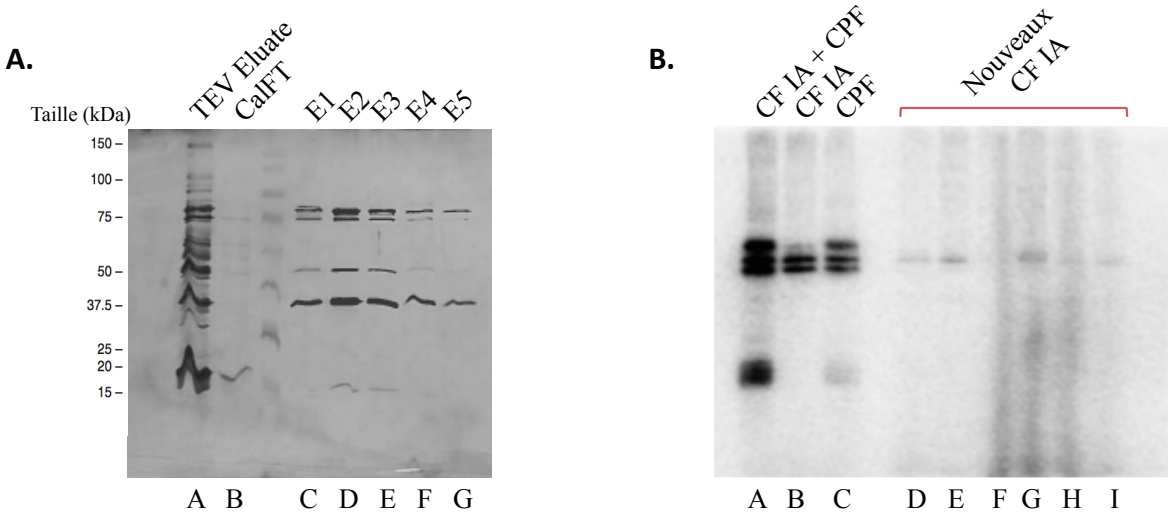


Figure 3.03 : Activité kinase associée aux complexes nouvellement purifiés.
(A.) Analyse des différentes étapes de purification par TAP-tagging du complexe CF IA. Séparation des protéines par SDS-PAGE et coloration au nitrate d'argent. ***A*** : TEV Eluate (1/100ième). ***B*** **:** Calmoduline Flow Through (1/700ième). ***C à G*** : fraction d'élution E1 à E5 (1/30ième). Le poids moléculaire apparent des protéines est estimé par rapport au marqueur de taille (*à gauche*) en kDa.
(B.) Essai d'activité kinase des complexes CF IA nouvellement purifiés. Les facteurs purifiés sont incubés en présence de [γ^{32}P]-ATP puis séparés par SDS-PAGE. Le transfert du γ^{32}P est observé par autoradiographie. ***A, B et C*** : facteurs témoins [CF IA+CPF], CF IA et CPF obtenu dans l'étude préliminaire. ***D à I*** : facteurs nouvellement purifiés - complexe CF IA.

Les différentes solutions testées afin de reproduire les résultats obtenus précédemment sont présentées dans les paragraphes suivants. Nous avons tout d'abord cherché à modifier les protocoles de tests d'activité, puis nous avons tenté d'isoler l'activité kinase lors des différentes étapes de purification et de modifier les protocoles de purification.

C.2. Modification des tests d'activité kinase

Parmi les différentes méthodes de détection des activités kinases, nous avons choisi d'utiliser la méthode de détection par test radioactif, appropriée pour la détection d'activité kinase de protéines avec une haute sensibilité. Les conditions de détection utilisées lors des résultats préliminaires correspondent à des conditions standards et peuvent nécessiter une mise au point. À partir du protocole définit précédemment et des protocoles disponibles dans différentes études, nous avons modifié certains paramètres afin de tenter d'améliorer la détection du signal. Ces différents tests ont été réalisés sur les facteurs témoins présentant une activité efficace puis, par la suite, sur les facteurs nouvellement purifiés, présentant une faible activité.

Les différents paramètres modifiés sont les suivants :

(i) les temps et température d'incubation, pouvant ne pas être optimaux pour permettre une réaction complète, de 30 minutes à 1 heure et de 30°C à 37°C.
(ii) la composition du mix réactionnel, en comparant les réactions en présence de 10mM HEPES-KOH pH 7,9 et 50mM TrisHCl pH 7,9 (He and Moore, 2005)
(iii) modification de la quantité des cofacteurs apportés dans le mix réactionnel pour une concentration finale 10mM $MgCl_2$ et 2mM $CaCl_2$.
(iv) Si lors de la préparation des facteurs, une protéine kinase co-purifie, on peut penser qu'une phosphatase peut co-purifier. Nous avons donc ajouté des inhibiteurs de phosphatase - PhosSTOP – dans le mix réactionnel.
(v) Nous avons également dialysé les fractions obtenues en fin de purification, pour s'affranchir de l'EGTA utilisé lors de l'élution des facteurs.

Les résultats obtenus pour les facteurs témoins CF IA présentant une activité efficace sont présentés **Figure 3.04** ; aucune des modifications apportées au protocole n'induit de modification de l'intensité du signal obtenu pour les échantillons témoins ou pour les facteurs nouvellement purifiés (*données non publiées*). Seule une modification du temps d'incubation modifie le profil de phosphorylation dans l'échantillon témoin. Après 1 heure d'incubation,

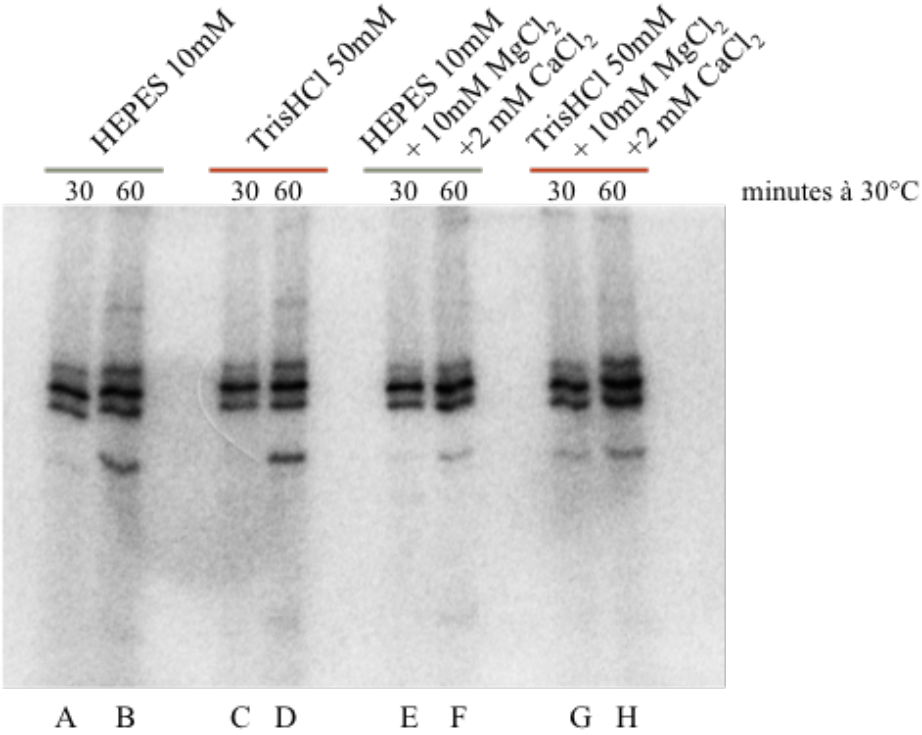


Figure 3.04 : Modification du protocole de tests d'activité kinase des complexes CF IA témoins. Les facteurs témoins sont incubés 30 et 60 minutes à 30°C. Les modifications de la composition du tampon sont indiquées. Les facteurs purifiés sont incubés en présence de [γ^{32}P]-ATP puis séparés par SDS-PAGE. Le transfert du γ^{32}P est observé par autoradiographie. ***A et B*** : 10mM HEPES. ***C et D*** : 50mM TrisHCl, pH 7,9. ***E et F*** : 10mM HEPES, 10mM $MgCl_2$, 2mM $CaCl_2$. ***G et H*** : 50mM TrisHCl, pH 7,9, 10mM $MgCl_2$, 2mM $CaCl_2$.

l'intensité du signal est augmentée, avec une meilleure détection d'une bande de basse masse moléculaire, qui pourrait correspondre à Rna15.

Les différentes modifications apportées au protocole des essais d'activité kinase n'ont pas permis de reproduire le profil de phosphorylation obtenu lors des analyses préliminaires. Par la suite, nous avons décidé d'utiliser les conditions de détection standard.

C.3. Analyse des différentes étapes de purification

Nous nous sommes ensuite intéressés aux différentes étapes de la purification. En effet, dans l'hypothèse où l'activité observée reflète la présence d'une protéine co-purifiant lors de la préparation des facteurs, nous avons cherché à voir si cette protéine kinase, à cause de la faible interaction avec les facteurs, ne pouvait pas être « perdue » dans les étapes intermédiaires de purification.

Les facteurs sont isolés par *TAP-tagging* selon le protocole décrit par Puig *et al.,* en 2001 (Puig et al., 2001) et adapté à notre modèle d'étude. Les protéines sont purifiées grâce à une étiquette située à l'extrémité N-terminale ou C-terminale. Cette dernière est formée de deux domaines d'affinité : CBP (Calmodulin Binding Peptide) et protéine A, séparés par un site de clivage par la TEV protéase (**Figure 3.05A**). Une première étape de purification par affinité est effectuée, en présence de billes IgG-Sépharose, qui vont interagir avec la protéine A (**Figure 3.05B**). Les protéines d'intérêts fixées sont récupérées par action de la TEV protéase. La fraction obtenue correspond au « ***TEV Eluate*** » et est incubée en présence de billes de calmoduline. La fraction non retenue, contenant aussi bien les protéines aspécifiques que les protéines d'intérêts n'ayant pu se fixer aux billes de calmoduline, correspond au « *Calmodulin Flow Through* » ou ***CalFT***. Les protéines d'intérêts sont éluées par ajout d'EGTA.

Nous avons tout d'abord testé uniquement les protéines récupérées à chaque étape en présence d'ATP radiomarqué : le ***TEV Eluate***, le ***CalFT*** et les facteurs obtenus. Les résultats pour une purification de **CF IA** sont présentés dans la **Figure 3.06B** et dans la **Figure 3.06C** pour **[CF IA+CPF]**. Le ***TEV Eluate*** est la première étape de purification et correspond à une fraction très concentrée (ligne A **Figure 3.06A**.). Dans les conditions d'essai, une forte activité kinase est retrouvée associée à cette fraction, mais le profil de phosphorylation est différent de celui observé pour les facteurs purifiés (lignes A et B **Figure 3.06B** ; lignes A/B

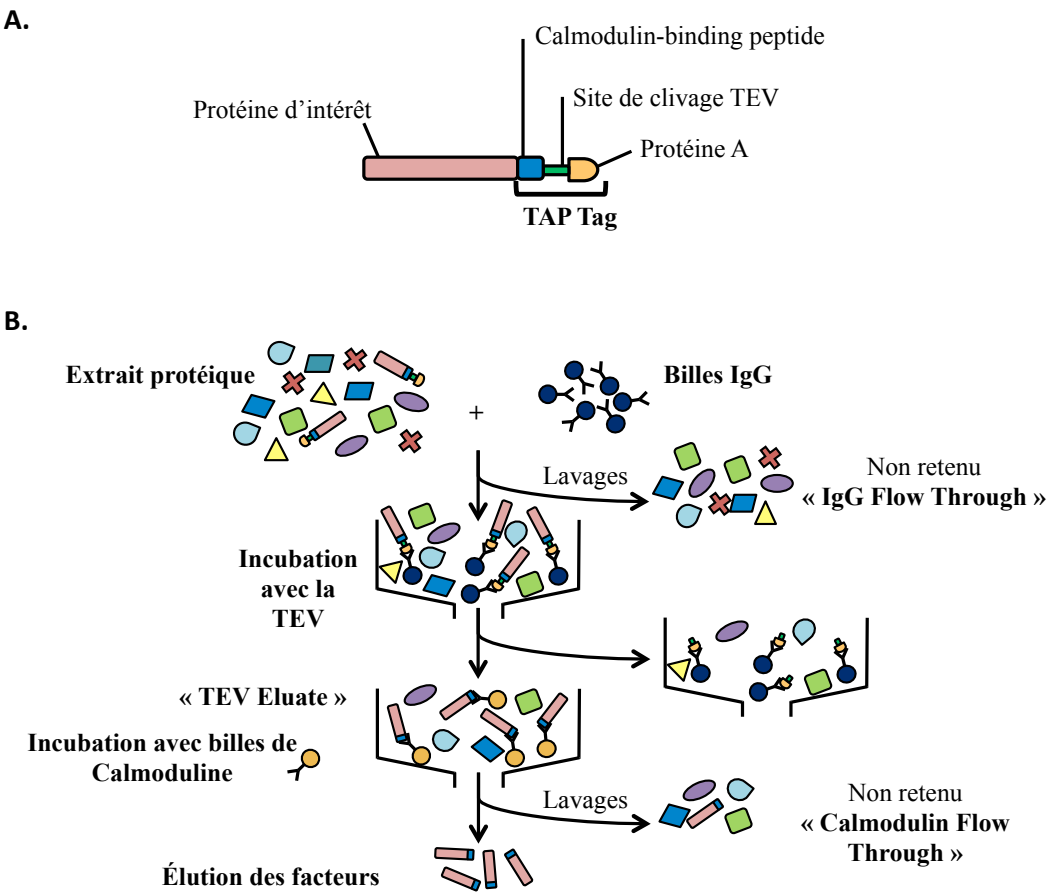


Figure 3.05 : Purification par TAP-tagging. **(A.)** L'étiquette TAP est située en C-terminale de la protéine d'intérêt, représentée en **rose**. L'étiquette TAP est formée d'une protéine A, d'un peptide interagissant avec la calmoduline (« *Calmodulin-binding peptide* ») séparé par un site de clivage par la TEV protéase (en **vert**).
(B.) L'extrait protéique est incubé en présence des billes d'IgG-sépharose (en **bleu**). La fraction non retenue correspond à *l'IgG Flow Through*. Les billes d'IgG sont incubées en présence de 100U de TEV à 4°C. La fraction non retenue contient les protéines d'intérêts et correspond au *TEV Eluate*. Cette fraction est incubée en présence des billes de calmoduline-sépharose (en **jaune**). La fraction non retenue correspond au *Calmodulin Flow Through*. Les facteurs sont élués par ajout d'EGTA.

et C/G **Figure 3.06C**). La fraction non retenue après la seconde étape de purification (***CalFT***) présente une activité identique à celle observée lors de l'incubation du ***TEV Eluate***, mais de plus faible intensité (ligne C **Figure 3.06B** ; lignes D et H **Figure 3.06C**); ce qui est en accord avec la quantité de protéines totales contenues dans chaque fraction (ligne B **Figure 3.06A**). La concentration du *CalFT* permet une augmentation du marquage (ligne F **Figure 3.06B**) ; le profil de phosphorylation n'est cependant pas comparable au profil attendu. Nous pouvons penser que cette activité est due à la phosphorylation de protéines extérieures aux complexes CF IA et CPF, cibles de la kinase *in vivo*. Les facteurs obtenus ne présentent pas ou peu de phosphorylation dans les conditions utilisées (lignes D et E **Figure 3.06B** ; lignes E, F, I et J **Figure 3.06C**). Les résultats obtenus ne dépendent pas des sous-complexes purifiés, **CF IA** (**Figure 3.06B**) ou **[CF IA+CPF]** (**Figure 3.06C**).

En supposant que l'activité protéine kinase est diluée lors de la préparation des facteurs, nous avons ensuite testé l'activité associée en combinant des facteurs ayant une faible activité et le ***CalFT*** présentant une activité, avant et après concentration. Comme observé précédemment, les facteurs obtenus ne possèdent pas ou peu d'activité. L'ajout de *CalFT* aux facteurs obtenus permet d'augmenter le signal et d'obtenir un profil de phosphorylation plus intense. La concentration du *CalFT* permet une augmentation plus importante de l'intensité du signal obtenu.

Lors de la purification des complexes **CF IA** et **CPF**, une activité protéine kinase est retrouvée à chaque étape mais les facteurs obtenus ne présentent pas d'activité. Une protéine présentant une activité kinase est ainsi bien co-purifiée, mais l'interaction entre les partenaires est faible et ne permet pas de l'obtenir en fin de purification - l'activité est retrouvée associée aux fractions d'élution et dans la fraction non retenue. Cela nous amène à penser que le protocole de purification par *TAP-tagging* n'est pas adapté par rapport au protocole initial utilisé dans les études préliminaires.

C.4. Modification du protocole de purification

L'analyse des différentes fractions a permis de mettre en évidence la présence d'une activité kinase dans chaque étape de purification et en particulier dans une fraction non retenue, le ***CalFT***. Le protocole utilisé n'est peut-être pas adapté à la co-purification d'une quantité suffisante de protéine kinase, à cause d'une stringence trop importante. Nous avons donc décidé de modifier différents paramètres de purification :

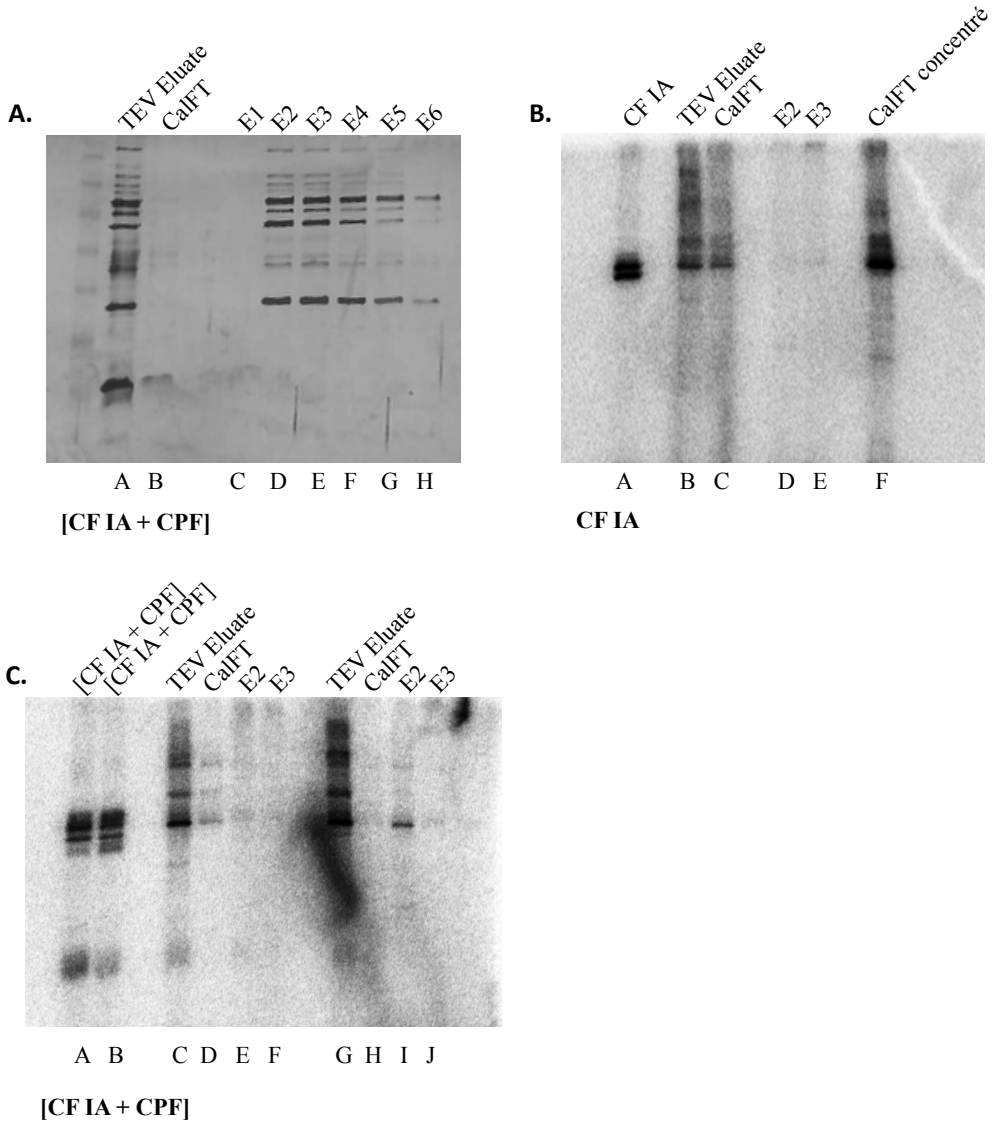


Figure 3.06 : Les différentes étapes de purification présentent une activité kinase.
(A.) Analyse des différentes étapes de purification par TAP-tagging du complexe [CF IA+CPF]. Séparation des protéines par SDS-PAGE et coloration au nitrate d'argent. ***A*** : TEV Eluate (1/100ième). ***B*** **:** Calmoduline Flow Through (1/700ième). ***C à H*** : fraction d'élution E1 à E6 (1/30ième). Le poids moléculaire apparent des protéines est estimé par rapport au marqueur de taille (*à gauche*) en kDa.
(B.) Essai d'activité kinase des différentes étapes de purification du complexe CF IA. 4,4µL de chaque étape de purification sont incubés en présence de [γ^{32}P]-ATP puis séparés par SDS-PAGE. Le transfert du γ^{32}P est observé par autoradiographie. ***A*** : témoin positif complexe CF IA. ***B :*** TEV Eluate (1/100ième). ***C*** **:** Calmoduline Flow Through (1/700ième). ***D et E*** : fractions d'élution E2 et E3 (1/30ième). ***F*** **:** Calmoduline Flow Through concentré.
(C.) Essai d'activité kinase des différentes étapes de purification du complexe [CF IA+CPF] pour deux purifications indépendantes. 4,4µL de chaque étape de purification sont incubés en présence de [γ^{32}P]-ATP puis séparés sur gel SDS-PAGE. Le transfert du γ^{32}P est observé par autoradiographie. ***A et B*** : témoins positifs complexes [CFIA+CPF]. ***C et G :*** TEV Eluate (1/100ième). ***D et H*** **:** Calmoduline Flow Through (1/700ième). ***E, F, I et J*** : fractions d'élution E2 et E3 (1/30ième).

(i) La salinité lors de l'élution des facteurs (100 mM contre 150 mM NaCl dans le protocole initial). Modification de la salinité pour les différents lavages (de 100 mM à 500 mM NaCl).

(ii) Les conditions de stringences, en diminuant la concentration en NP-40 à une concentration finale de 0,01% pendant toutes les étapes de purification et en diminuant les volumes de lavages. Cela a permis d'augmenter la quantité de facteurs obtenus en fin de purification mais pas de rétablir une activité kinase en test *in vitro*.

(iii) Ajout de différents inhibiteurs de phosphatases (fluorure de sodium, β-glycérophosphate et orthovanadate de sodium) lors de la préparation des facteurs – la déphosphorylation de la protéine kinase pouvant conduire à une diminution de son activité et de l'interaction avec les complexes **CF IA** et **CPF**.

Nous avons également cherché à reproduire le protocole initial de la manière la plus fidèle possible. Les paramètres variant par rapport au protocole initial sont : la localisation de l'étiquette pour le *TAP-tagging* dans les souches utilisées, le temps d'incubation en présence de la TEV protéase (3 heures à 16°C initialement contre toute la nuit à 4°C), la quantité d'EGTA et de $CaCl_2$, le temps d'incubation et quantité de billes d'IgG-Sépharose (250µL initialement contre 400µL), ainsi que la DO_{600nm} des cultures de départ. Ces modifications n'ont pas permis de rétablir un profil de phosphorylation semblable à celui obtenu précédemment – une bande de faible intensité reste cependant visible pour **CF IA** et **CPF**.

Nous avons également réalisé une purification à partir de la souche parentale BMA64, ne présentant aucune étiquette sur les différentes sous-unités. L'objectif était d'éliminer la possibilité d'une interaction aspécifique entre la protéine kinase et les billes (IgG ou Calmoduline) utilisées lors de la préparation. Nous avons ensuite concentré les fractions obtenues et analysé chaque étape de la purification. Aucun activité protéine kinase n'est retrouvée, signifiant que la co-purification est spécifiquement associée au complexe de polyadénylation.

En plus du contrôle à partir de la souche parentale, nous avons étudié la possibilité que l'activité de protéine kinase co-purifiant lors de la préparation des facteurs puisse être attribuée à une protéine kinase calmoduline-dépendante (CAMK). L'activité des protéines kinases de la famille des CAMK est permise par la fixation de calcium ou de Calmoduline au niveau d'un domaine situé en C-terminal du domaine catalytique (Nadeau et al., 1997). Nous avons purifié le complexe CF IA grâce à une étiquette formée uniquement par Protéine A – TEV. Les protéines d'intérêt sont récupérées après fixation sur billes d'IgG sépharose et par

action de la TEV protéase. Une fraction unique, le *TEV Eluate* est obtenue. Une activité protéine kinase est retrouvée associée à cette fraction. La kinase est associée spécifiquement au complexe CF IA.

Les modifications apportées au protocole de purification n'ont pas permis de reproduire les résultats préliminaires. L'absence d'activité n'est pas due à un défaut de co-purification entre la protéine kinase et les facteurs lors de la purification par *TAP-tagging*.

C.5. Concentrations des facteurs

En comparant les résultats obtenus – test d'activité et coloration au nitrate d'argent – nous pouvons remarquer une corrélation entre diminution de la quantité de facteurs obtenus et intensité du signal. Nous pouvons donc penser que les différences de profils observés sont dues à un problème de seuil de détection. Ces résultats sont corrélés par les résultats obtenus en associant facteurs purifiés et étapes de purification.

Nous avons donc décidé d'analyser les différents sous-complexes **CF IA** et **[CF IA+CPF]** après concentration progressive des facteurs sur Centricon (10NMWL - Millipore). Les résultats obtenus sont présentés dans la **Figure 3.07A** en comparaison avec le gel de protéines colorées au nitrate d'argent correspondant **Figure 3.07B**. Une augmentation croissante de la quantité de facteurs conduit à une augmentation progressive du marquage, pour chaque facteur testé (**CF IA** et **[CF IA+CPF]**). Pour le complexe **CF IA** (lignes C à G), le marquage visible après concentration des facteurs n'est pas comparable à celui obtenu avec les facteurs témoins. Dans le cas des complexes **[CF IA+CPF]** (lignes H à M), le marquage visible après concentration est comparable au profil des facteurs témoins. Une bande de plus faible poids moléculaire est détectée ainsi que des bandes à des poids moléculaires intermédiaires – environ 50kDa. Cependant, des bandes de plus haut poids moléculaire sont également détectées. Ces dernières peuvent correspondre i) à des bandes de spécificité moins importante et illustrer le marquage de protéines dont la phosphorylation est « forcée » par l'augmentation de la concentration ou ii) à des protéines réellement cibles qui n'étaient pas visibles dans l'étude préliminaire car pas assez abondantes et en dessous du seuil de détection.

En conclusion, une activité de type protéine kinase est retrouvée co-purifiant lors de la préparation des sous-complexes **CF IA**, **CPF** et **[CF IA+CPF]**. Cependant, les quantités obtenues en fin de purification ne sont pas suffisantes pour observer un profil de

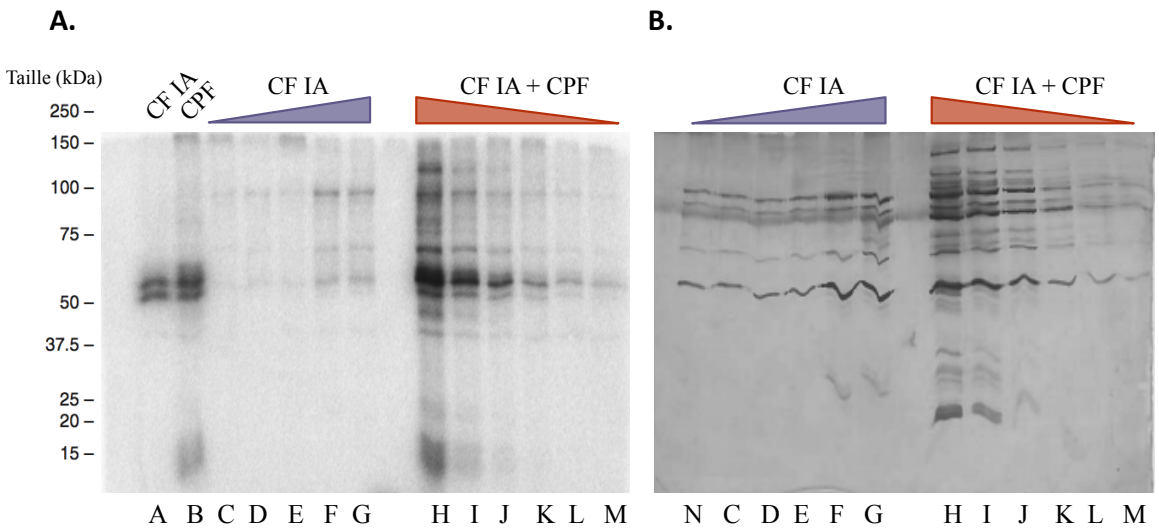


Figure 3.07 : La concentration des facteurs permet de rétablir un marquage comparable à celui des facteurs témoins.

(A.) Essai d'activité kinase après concentration des facteurs pour les complexes **CF IA** et [**CF IA+CPF**]. Les facteurs purifiés sont incubés en présence de [γ^{32}P]-ATP puis séparés par SDS-PAGE. Le transfert du γ^{32}P est observé par autoradiographie. Le poids moléculaire apparent des protéines est estimé par rapport au marqueur de taille (*à gauche*) en kDa. ***A et B*** : témoins positifs complexes CFIA et CPF. ***C à G*** : concentrations croissantes de facteurs du complexe CF IA. ***H à M*** : concentration décroissante de facteurs du complexe [CFIA+CPF].

(B.) Séparation des protéines par SDS-PAGE et coloration au nitrate d'argent des facteurs concentrés pour les complexes **CF IA** et [**CF IA+CPF**]. ***N et C à G*** : concentrations croissantes de facteurs du complexe **CF IA**. ***H à M*** : concentration décroissante de facteurs du complexe [**CF IA+CPF**]. La fraction N n'est pas été analysée dans l'essai d'activité kinase.

phosphorylation « complet ». La concentration des facteurs permet de rétablir un profil similaire, mais conduit à l'apparition de nouvelles cibles, dont la spécificité reste à déterminer.

C.6. Test d'activité ATPase

En parallèle des tests d'activité kinase, ayant écarté une fonction de protéine kinase *in vitro* à Clp1, nous avons étudié la fonction ATPasique putative de cette protéine. Les ATPases sont des enzymes, de la famille des hydrolases, catalysant l'hydrolyse de l'ATP en ADP et en phosphate inorganique. L'analyse de la structure de Clp1 et du domaine d'interaction avec Pcf11, décrite par Noble *et al.,* 2007 (Noble et al., 2007), a permis de montrer que la région de fixation de l'ATP est constituée par un motif « *P-loop* » et des régions appelées « *Switch I et II* », communément retrouvés dans les ATPase/GTPase de la classe des SIMIBI, comme NifH, MinD ou Soj. Ces ATPases ont la particularité de ne présenter une activité d'hydrolyse de l'ATP qu'en présence d'un ou plusieurs partenaires protéiques, permettant un changement de conformation. De la même manière que pour l'activité kinase, nous pouvons nous demander si Clp1 ne présente pas cette particularité. Les essais ATPase à partir de protéines recombinantes (Clp1, Clp1-Pcf11 ou du complexe CF IA) n'ont pas permis de révéler une activité d'hydrolyse de l'ATP à Clp1. Nous avons cherché à étudier l'activité d'hydrolyse de l'ATP à partir de complexes natifs, purifiés à partir de *S. cerevisiae*.

Pour cela nous avons réalisé des tests d'hydrolyse de [γ^{32}P]ATP en ADP + γ^{32}Pi en présence des complexes CF IA portant la version sauvage du gène *CLP1* et l'allèle ***clp1-1***. Les résultats obtenus sont présentés dans la **Figure 3.08A**. La phosphatase alcaline (ou CIP – *Calf Intestinal Alkaline Phosphatase*) est utilisée comme témoin de l'hydrolyse de l'ATP (ligne b.). En présence de CF IA (lignes e., f. et i., j.) ou en présence de l'allèle ***clp1-1*** (lignes c. et d.), aucune hydrolyse d'ATP n'est observée. L'ajout du complexe natif CPF, en présence des différents allèles de Clp1, ne modifie pas le profil obtenu. Cela confirme les résultats obtenus avec les protéines recombinantes et les résultats obtenus lors de l'étude de l'activité kinase pouvant être attribuée à Clp1.

Les protéines kinases sont des protéines qui catalysent le transfert d'un phosphate radioactif sur un résidu d'une protéine cible. Les ATPases sont des enzymes de la famille des hydrolases permettant de l'hydrolyse de l'ATP en ADP et la libération d'un phosphate

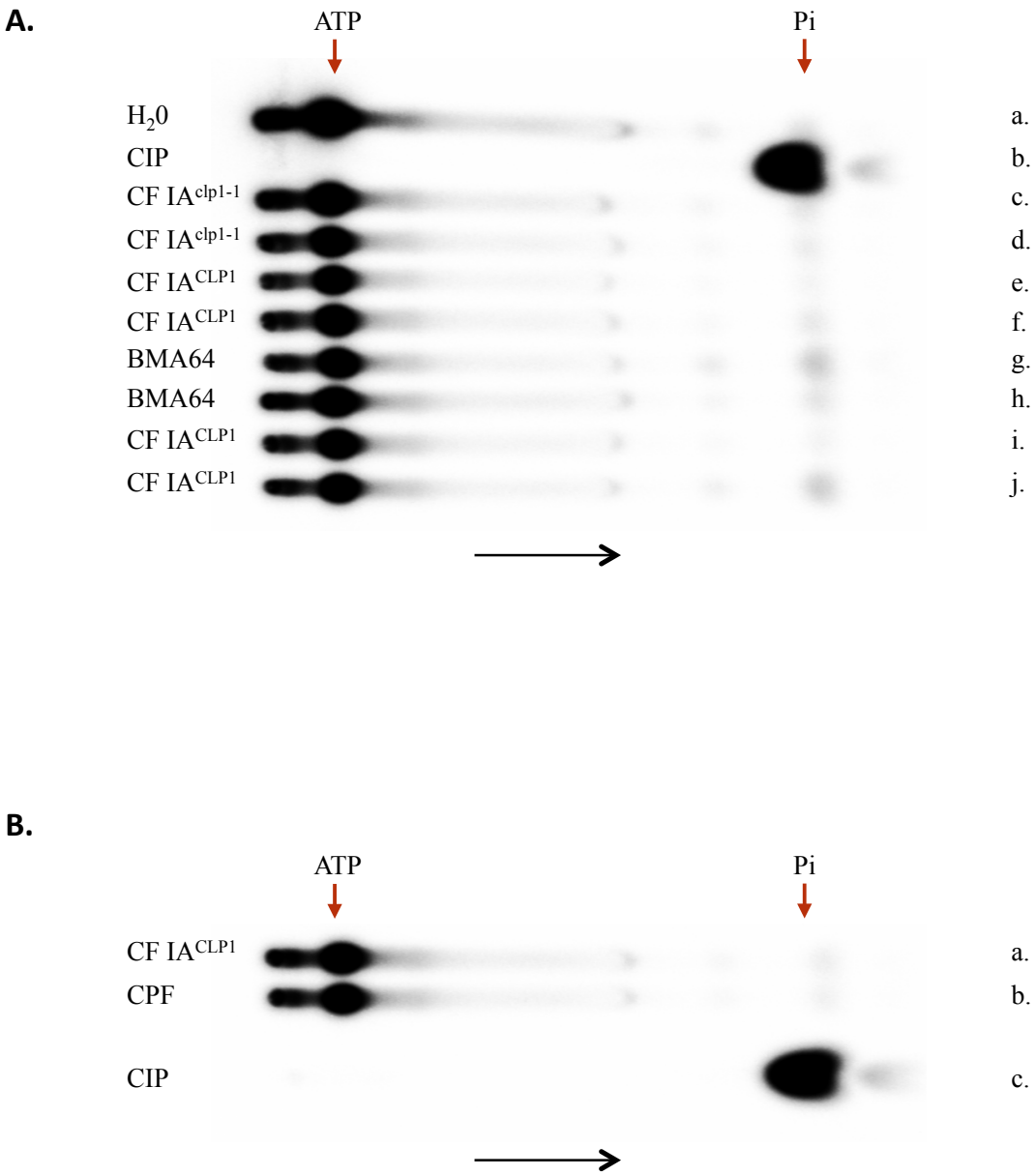


Figure 3.08 : Clp1 ne possède pas d'activité d'hydrolyse de l'ATP. (A.) Essai d'hydrolyse de [γ^{32}P]-ATP en présence des complexes CF IA purifié à partir de souches exprimant une copie sauvage de *CLP1* ou l'allèle *clp1-1*. 1/10ième de la réaction est analysée par chromatographie sur couche mince (CCM). Le γ^{32}P libéré est séparé par 1M d'acide acétique et 0,5M d'acétate de lithium. La migration est effectuée pendant 45 à 60 minutes. Le [γ^{32}P]-ATP et γ^{32}P sont indiqués pas des flèches rouges. Le sens de migration est indiqué. ***a.*** : contrôle négatif. ***b.*** : contrôle positif, phosphatase alcaline (CIP). ***c et d.*** : complexe CF IA^{clp1-1}. ***e, f, i et j.*** : complexe CF IA^{CLP1}. ***g et h.*** : contrôle négatif de la purification, fraction obtenues par purification de la souche parentale BMA64. (B.) Essai d'hydrolyse de [γ^{32}P]-ATP des facteurs témoins présentant une activité protéine kinase. 1/10ième de la réaction est analysée sur chromatographie sur couche mince (CCM). Le γ^{32}P libéré est séparé par 1M d'acide acétique et 0,5M acétate de lithium. La migration est effectuée pendant 45 à 60 minutes. Le [γ^{32}P]-ATP et γ^{32}P sont indiqués pas des flèches rouges. Le sens de migration est indiqué. ***a.*** : facteurs du complexe CF IA. ***b.*** : facteurs du complexe CPF. ***c.*** : contrôle négatif.

inorganique libre et de l'énergie. L'objectif ici, en utilisant de [γ^{32}P]ATP est d'étudier la libération de phosphate inorganique libre. Nous avons également testé l'activité d'hydrolyse de l'ATP des complexes purifiés **CF IA** et **CPF** présentant une activité de protéine kinase. Les résultats obtenus sont présentés dans la **Figure 3.08B**. De la même manière, la phosphatase alcaline est utilisée comme témoin de réaction (ligne c.). Aucune activité d'hydrolyse de l'ATP n'est observée en présence des complexes **CF IA** ou **CPF** (lignes a. et b.) présentant une activité de protéine kinase.

En conclusion, Clp1 ne présente pas d'activité d'hydrolyse de l'ATP comme pouvant être suggéré par son organisation et ce malgré la présence des autres partenaires du complexe de polyadénylation – **CF IA** et **CPF**. De plus, aucune activité n'est détectée en présence de la protéine kinase co-purifiant lors de la préparation des facteurs par *TAP-tagging*.

D. DISCUSSIONS ET PERSPECTIVES

La délétion de Glc7, sérine/thréonine phosphatase associée au sous-complexe CPF conduit à l'accumulation de formes phosphorylées de Pta1 et Fip1 (He and Moore, 2005). Les protéines kinases CK1 et CK2 ont été proposées comme responsables de cette phosphorylation ; cependant aucune donnée *in vivo* n'est disponible. L'activité de certaines protéines kinases, impliquées notamment dans les cascades de signalisation, est régulée par des interactions protéines-protéines. Chez *S. cerevisiae*, il a par exemple été montré que le complexe SWI/SNF, impliqué dans le remodelage de la chromatine, permet d'activer Mec1, protéine kinase qui joue un rôle dans les mécanismes de réparation de l'ADN et le contrôle du cycle cellulaire (Kapoor et al., 2015).

Notre hypothèse de départ dans cette étude est que Clp1, au vu de ces caractéristiques structurales, possède une activité de type protéine kinase uniquement en présence des autres partenaires du/des complexe(s) CF IA et/ou CPF. La présence des différents partenaires conduirait à un changement de conformation, direct ou indirect, permettant l'activation de la protéine et la phosphorylation des protéines cibles, du complexe de polyadénylation. Les résultats préliminaires n'ont pas permis de valider cette hypothèse ni d'attribuer une activité enzymatique à Clp1, comme ses homologues chez l'Homme (Weitzer and Martinez, 2007a) ou chez les archées (Jain and Shuman, 2009). Pourtant, ils ont permis de mettre en évidence une activité protéine kinase, co-purifiant lors de la préparation des sous-complexes CF IA et CPF, indépendamment de la présence de Clp1. Comme l'épissage ou l'export, la maturation de l'extrémité 3' serait régulée par phosphorylation. Les différentes questions biologiques posées par la présence d'une protéine kinase n'ont pu être résolues : la co-purification de l'activité kinase n'ayant pu être reproduite comme obtenue lors des résultats préliminaires. Les différentes modifications apportées aux protocoles de purification et de test d'activité n'ont pas permis de reproduire les résultats préliminaires. Cependant, la concentration des facteurs en fin de purification nous a finalement permis de reproduire un profil de phosphorylation comparable.

À court terme, l'objectif serait d'améliorer les rendements de purification. Les facteurs sont purifiés par *TAP-tagging* ; l'analyse de chaque étape par SDS-PAGE et par essai kinase, a permis de déterminer que l'activité protéine kinase était perdue durant les différentes étapes de purification. Pour améliorer les rendements, nous avons tenté de modifier le protocole de purification, en particulier après la première étape de fixation sur billes d'IgG-Sépharose. La

seconde étape de fixation sur billes de Calmoduline-Sépharose peut être considérée comme une étape limitante et des facteurs peuvent être retrouvés dans le *CalFT*, fraction non retenue. Les résultats obtenus ont également montré que l'élution des facteurs était insuffisante : des facteurs sont retrouvés dans les fractions d'élution E7 à E8. De plus, l'analyse des billes par SDS-PAGE montre que des facteurs sont retrouvés associés aux billes après élution. L'amélioration de ces étapes du protocole de purification (fixation sur billes Calmoduline et élution des facteurs) permettrait d'augmenter les rendements de purification et d'améliorer la détection de l'activité associée aux complexes. Nous souhaitons également étudier l'impact de la $D.O._{600nm}$ sur les rendements de purification et sur l'activité protéine kinase associée. De la même manière que la phosphorylation de Pap1, variable durant le cycle cellulaire (Mizrahi and Moore, 2000), la protéine kinase pourrait être associée transitoirement au complexe de polyadénylation. Nous pouvons également imaginer développer un protocole à une étape de purification, sur billes d'IgG-Sépharose. Le *TEV Eluate*, fraction obtenue après fixation sur billes d'IgG-Sépharose est une fraction relativement concentrée, active en essai kinase. Les premiers essais de purification simplement par une étiquette Protéine A-TEV permettent d'obtenir des facteurs relativement purs et actifs en essai kinase. La mise au point de ce protocole à plus grande échelle permettrait de minimiser la dilution de l'activité à chaque étape de purification, et pourrait permettre d'obtenir une fraction active, plus concentrée, présentant un profil de phosphorylation comparable à celui obtenu lors des résultats préliminaires.

Malgré les problèmes de reproductibilité des résultats, différents contrôles de spécificité ont pu être réalisés. L'activité, associée aux sous-complexes CF IA et [CF IA+CPF], n'est pas portée par une protéine kinase calmoduline-dépendante (CAMK – contrôle avec l'étiquette Protéine A-TEV) ni par une protéine interagissant avec les billes utilisées (IgG ou Calmoduline – contrôle dans la souche parentale BMA64). Le rôle de l'étiquette TAP reste cependant à déterminer ; le polypeptide pourrait présenter une interaction aspécifique avec une protéine kinase. Si l'interaction de la protéine kinase est spécifique au complexe de polyadénylation, la purification par *TAP-tagging* d'une protéine ne faisant pas partie du complexe permettra d'obtenir des fractions inactives en essai kinase. Aucune activité de type protéine kinase ne sera alors purifiée ; et l'ajout des fractions d'élutions obtenues aux fractions de l'étude préliminaire ou aux fractions considérées comme faiblement actives ne modifiera pas le marquage observé.

C'est la première fois qu'une activité de type protéine kinase est retrouvée associée spécifiquement et *in vivo* au complexe de polyadénylation. En effet, dans les études précédentes comme l'étude de la phosphorylation du complexe CPF par CK1 et CK2 (He and Moore, 2005) ou de la phosphorylation de Fip1 par CK2 (Zielinski et al., 2006), la phosphorylation du complexe est réalisée après purification des facteurs, par ajout de la protéine kinase. L'ajout de CK1 ou CK2 en présence du facteur CPF permet la phosphorylation de Yhh1/Cft1 (153kDa), Ysh1/Brr5 (88kDa), Pta1 (88kDa) et Fip1 (36kDa). Cette régulation a bien lieu *in vitro* mais cela ne signifie pas qu'elle possède une signification biologique et que les protéines cibles le soient *in vivo*. La co-purification d'une protéine kinase lors de la préparation des facteurs apporte un nouvel élément de réponse et permettra de valider ou non les différents modèles de régulation proposés.

Le complexe de polyadénylation est composé, chez *S. cerevisiae*, de trois sous-complexes : CF IA, CF IB et CPF. Le sous-complexe CF IB est formé par Hrp1, protéine de 73kDa, ne présentant pas d'homologue dans le système mammifère (Kessler et al., 1996). Nous pouvons nous demander si Hrp1 ne joue pas un rôle dans l'activité protéine kinase retrouvée. Hrp1 est associée au complexe de polyadénylation par interaction avec Rna15 et Rna14, et interagit avec l'ARN, permettant la sélection du site de clivage (Barnwal et al., 2012; Gross and Moore, 2001b; Noble et al., 2004). La présence de Hrp1 pourrait permettre d'augmenter l'association de la protéine kinase avec le complexe. Nous pouvons également imaginer que Hrp1 peut induire une activation de la protéine kinase par un changement de conformation dû à son interaction, directe ou indirecte. Il serait intéressant de tester la présence de Hrp1 dans les fractions « actives » obtenues dans l'étude préliminaire en comparaison aux fractions présentant une activité faible, par Western blot.

Même si Hrp1 n'est pas présent dans les fractions actives, il pourrait être judicieux d'étudier l'effet éventuel de la phosphorylation sur l'activité de Hrp1. Plusieurs sites de phosphorylation sont décrits pour cette protéine (S2, S3, S51, S87, T184, S206, S446, T458, S462) ainsi que deux motifs de reconnaissance pour la protéine CK2 à l'extrémité N-terminale (2 à 5 et 3 à 6 pour CK2) (PhosphoGRID). La phosphorylation de Hrp1 permettrait de réguler le choix du site de clivage par modification de son interaction avec Rna14/Rna15 et/ou l'ARN. De nombreux exemples de modulation de l'interaction avec l'ARN par phosphorylation sont disponibles dans la littérature. L'un des exemples le mieux caractérisé est la régulation de la protéine SFRS1, impliquée dans l'épissage alternatif chez l'Homme SFRS1 est formée de deux motifs liant l'ARN dans sa région N-terminale (RRMs) et d'un

domaine RS (RS1 et RS2) dans sa région C-terminale. La phosphorylation du domaine RS1 par la protéine kinase SRPK1 permet de diminuer son interaction avec l'ARN lors de son association avec les mRNPs, et ainsi un import précoce ; SFRS1 est ensuite associée au *spliceosome* (Thapar, 2015). La phosphorylation de Hrp1 permettrait de réguler spécifiquement son interaction avec l'ARN, sans modifier les interactions avec les autres partenaires protéiques, comme c'est le cas, chez l'Homme, pour la phosphorylation de la protéine $p54^{nrb}$. La phosphorylation du résidu T45 en amont de deux domaines de liaison à l'ARN durant la mitose permet de diminuer la capacité d'interaction de $p54^{nrb}$ avec l'ARN en préservant les interaction protéines-protéines. De plus cette modulation de l'interaction est sélective, l'interaction avec des séquences poly(G) n'est pas dépendante de la phosphorylation (Bruelle et al., 2011).

De la même manière, le rôle éventuel de la phosphorylation sur l'activité des protéines régulant la taille de la queue poly(A), Pab1 et Nab2 peut être étudié. Ces deux protéines présentent des sites de phosphorylation et de reconnaissance pour la CDK1/MAPK et MAPK, respectivement (Pab1 : S63, T102, S125, S249, S332, S405, S565. Nab2 : S2, T178, S180 – PhosphoGRID). Nab2 est impliquée dans le contrôle de la taille de la queue poly(A) (Hector et al., 2002) et également dans l'export des ARNm vers le cytoplasme, en permettant l'interaction des protéines Mex67/Mtr2 avec les mRNPs (Green et al., 2002; Hanada et al., 2013). La phosphorylation de Nab2 pourrait modifier le recrutement des protéines Mex67/Mtr2.

À plus long terme, l'objectif sera d'identifier la protéine kinase responsable de la phosphorylation ainsi que son/ses partenaires(s) au sein du complexe de polyadénylation. Plusieurs stratégies peuvent être mises en place, fondées sur une détection par spectrométrie de masse couplée à l'utilisation d'un analogue de l'ATP, le FSBA. En parallèle des analyses par spectrométrie de masse, nous pouvons nous demander si l'activité protéine kinase observée ne peut être attribuée à l'une des protéines, précédemment décrites (He and Moore, 2005; Kyburz et al., 2003; Mizrahi and Moore, 2000; Zielinski et al., 2006). (i) Les premiers candidats possibles sont les protéines kinases impliquées dans la phosphorylation du CTD, notamment Ctk1/Bur1 permettant la phosphorylation de résidu SER^2 en fin d'élongation (Ahn et al., 2004; Skaar and Greenleaf, 2002). Les fractions d'élutions présentant une activité en essai kinase sont incubées en présence de GST-CTD fixés sur billes GST-Sépharose et de [γ^{32}P]-ATP. Le marquage éventuel du GST-CTD pourra être observé par SDS-PAGE et autoradiographie. Si le CTD est phosphorylé par les fractions actives en essai kinase, les

résidus du CTD phosphorylés peuvent être ensuite identifiés par spectrométrie de masse ou par Western blot. (ii) CK2 peut également être une protéine kinase candidate. CK2 est une sérine/thréonine kinase formée de quatre sous-unités (CKα1, CKα2, CKβ1 et CKβ2) impliquée dans la phosphorylation de Fip1 et Pta1 *in vitro* (He and Moore, 2005; Zielinski et al., 2006). Pcf11 interagirait également avec CKα1 (Uetz et al., 2000). Certaines sous-unités du complexe CPF présentent des motifs de reconnaissance pour cette protéine (Ysh1, Fip1) (PhosphoGRID). Les sous-unités CKα1 et CKα2 sont les deux sous-unités catalytiques de CK2 et peuvent être produites de manière recombinante. Nous pouvons tester si l'incubation des facteurs ne présentant pas d'activité kinase en présence de CKα1 et CKα2 permet de rétablir un marquage comparable à celui obtenu lors des analyses préliminaires. (iii) D'autres protéines peuvent être étudiées de cette même manière, comme CK1 ; des motifs de reconnaissance sont décrits parmi les protéines du complexe de polyadénylation (PhosphoGRID) et CK1 a été montré comme ciblant certains protéines *in vitro* (Yhh1/Cft1, Ysh1/Brr5, Pta1 et Fip1) (He and Moore, 2005).

Dans un second temps, l'objectif sera d'identifier les cibles de cette phosphorylation ainsi que la fonction de ces modifications post-traductionnelles. Plusieurs études ont montré un rôle important de la phosphorylation dans la régulation de l'activité de certaines protéines du complexe de polyadénylation, en particulier Pap1 (He and Moore, 2005; Mizrahi and Moore, 2000; Zielinski et al., 2006). De la même manière que l'identification de la protéine kinase, les cibles peuvent être identifiées par spectrométrie de masse. À partir des résultats préliminaires et si les protéines phosphorylées appartiennent au complexe de polyadénylation, plusieurs protéines candidates peuvent être proposées. La phosphorylation de certains composants du sous-complexe CPF (Yhh1/Cft1 (153kDa), Ysh1/Brr5 (88kDa), Pta1 (88kDa) et Fip1 (36kDa)) a précédemment été décrite mais aucune donnée n'est disponible quant à la phosphorylation du sous-complexe CF IA. Pour le sous-complexe CPF, en ne prenant en compte que la taille des protéines, sept protéines candidates peuvent être proposées : des protéines de poids moléculaire apparent proche de 50kDa et des protéines de plus faible poids moléculaire, dont la phosphorylation n'est pas observée dans l'étude publiée par He et Moore, en 2005. La phosphorylation de protéines de plus haut poids moléculaire - Yhh1/Cft1 (153kDa), Ysh1/Brr5 (88kDa), et Pta1 (88kDa) – n'est pas retrouvée dans nos conditions. La phosphorylation de ces protéines peut être spécifique de protéines kinases CK1 et CK2, qui ne seraient alors pas la/les protéine(s) kinase(s) associée(s) au complexe des polyadénylation ; ou la phosphorylation de ces protéines pourrait ne pas avoir de fonction biologique *in vivo* et

correspondrait à une phosphorylation « forcée » dans les conditions d'essais *in vitro*. Dans le cas où CK1 et CK2 seraient la/les protéine(s) kinase(s) associée(s) au complexe des polyadénylation, nous pouvons également imaginer que dans nos conditions d'étude, leur activité est régulée – par la présence des autres partenaires du complexe de polyadénylation non testés dans les études précédentes, ou par la co-purification d'une protéine régulatrice.

Tout d'abord des protéines de poids moléculaire apparent proche de 50kDa comme Pfs2 (53kDa), Fip1 (36kDa + 5kDa étiquette CBP), Mpe1 (50kDa) et/ou Pti1 (47kDa). Ces résultats sont en accord avec les données *in vitro* précédemment décrites, suggérant que Fip1 est phosphorylée par CK2 et CK1 au sein du complexe CPF (He and Moore, 2005; Zielinski et al., 2006). Plusieurs sites de phosphorylation et des motifs de reconnaissance pour des protéines kinases (CK1, CK2 et MAPK) sont décrits pour Fip1 (S15, S17, S50, S73, S77, S109, S110, T145), Mpe1 (T212, S221), et Pti1 (S177, S190, S267, S268, T271, S272, S274, T276, S279) (PhosphoGRID). Aucun site de phosphorylation ou motif de reconnaissance n'a cependant été décrit pour Pfs2.

La phosphorylation *in vitro* de Fip1 par CK2 a été décrite par Zielinski *et al.*, (Zielinski et al., 2006). Fip1, sous sa forme phosphorylée, modifie l'activité de Pap1 et son interaction avec le complexe CPF. Si la protéine kinase co-purifiant lors de la préparation des facteurs est CK2, les facteurs associés à une activité kinase présenteront une activité de polyadénylation réduite. Si la phosphorylation de Fip1 conduit à une diminution de l'association de Pap1 avec le reste du complexe CPF, bien que non quantitatif, l'analyse par Western blot permettrait d'illustrer cette baisse d'association.

Mpe1 est un médiateur de l'ubiquitination de Pap1 et interagit avec le pré-ARNm, via son domaine « *zinc knuckle* » (Lee and Moore, 2014; Vo et al., 2001). La phosphorylation de Mpe1 permettrait de moduler son rôle dans l'ubiquitination de Pap1, par un changement de conformation ou une diminution de son association au complexe CPF – Mpe1 interagit avec Pfs2 et Fip1 (Gavin et al., 2002). L'ubiquitination de Pap1 peut être étudiée par l'utilisation d'anticorps dirigés spécifiquement contre l'ubiquitine. Cependant, ne connaissant pas la fonction de l'ubiquitination de Pap1 (dégradation, modification de sa localisation, etc.), l'ajout d'inhibiteurs du protéasome pourrait permettre d'augmenter le signal détecté et de comprendre la fonction de l'ubiquitination de Pap1. L'association de Mpe1 au complexe CPF peut être, de la même manière que pour Fip1, étudiée par Western blot.

Pti1 est supposée, avec Pta1, permettre le recrutement de Glc7 (He and Moore, 2005). La phosphorylation de Pti1 pourrait permettre de réguler l'association de Glc7 avec le reste du

complexe de polyadénylation. L'association de Pti1 (phosphorylée ou non) et Glc7 peut être étudiée par pull down.

Ensuite, de manière plus surprenante et inattendue, les résultats obtenus ont permis de décrire la phosphorylation de protéines de plus faible poids moléculaire, pouvant correspondre à Yth1 (26kDa), Syc1 (21kDa) et/ou Ssu72 (23kDa). Aucun mécanisme de régulation par phosphorylation *in vitro* n'a été décrit pour ces protéines. Plusieurs sites de phosphorylation et des motifs de reconnaissance pour la protéine kinase CK1 sont décrits pour Ssu72 (T41, S43, Y68). Cependant, aucun site de phosphorylation ou motif de reconnaissance n'a été décrit pour Yth1 ou Syc1 (PhosphoGRID). Il semble donc peu probable que ces protéines soient la cible de phosphorylation ; les analyses par spectrométrie de masse permettront de confirmer cette hypothèse. Ssu72 est une phosphatase impliquée dans la déphosphorylation des résidus SER^5 et SER^7 en fin d'élongation (Ghazy et al., 2009; Krishnamurthy et al., 2004; Meinhart et al., 2003; Rosado-Lugo and Hampsey, 2014). L'activité de phosphatase de Ssu72 n'est pas nécessaire aux activités de maturation de l'extrémité 3' (Krishnamurthy et al., 2004). Chez l'Homme, il a été montré que l'activité de la phosphatase 2A pouvait être régulée par phosphorylation (Chen et al., 1992). De la même manière, la phosphorylation de Ssu72 pourrait permettre de réguler à la baisse son activité de phosphatase lorsque Ssu72 est associée au reste du complexe de polyadénylation ; et promouvoir son rôle dans la maturation de l'extrémité 3'. Cela permettrait de rester en phase d'élongation et pourrait ainsi participer au maintien de la terminaison de transcription.

Pour la première fois, une activité protéine kinase est retrouvée associée spécifiquement au complexe CF IA. De la même manière que pour le complexe CPF, en ne prenant en compte que la taille des protéines, seule la protéine Clp1 (50kDa) apparaît comme candidat potentiel. La phosphorylation de Clp1 pourrait permettre de réguler l'assemblage du complexe, comme Yra1 permet de moduler les interactions entre Pcf11 et Clp1. Cependant, aucun site de phosphorylation ou motif de reconnaissance n'est décrit pour Clp1 (PhosphoGRID). D'un point de vue structural, Clp1 présente les caractéristiques de certaines ATPases de la classe SIMIBI, caractérisées notamment par la fixation d'une molécule d'ATP au niveau d'un motif *P-loop* (de Vries et al., 2000; Dupin and Fribourg, 2014; Noble et al., 2007). Nous pouvons donc penser que le marquage observé peut être dû à la fixation d'une molécule de [γ^{32}P]-ATP au niveau de son site de liaison. Parmi les sous-unités du complexe CF IA, seule Rna14 est décrite comme présentant des sites de phosphorylation potentiels et de

reconnaissance pour des protéines kinases (T4, Y478, Y485– PhosphoGRID). L'analyse par spectrométrie de masse permettra d'identifier l'origine du marquage observé – phosphorylation d'une protéine cible ou fixation d'une molécule de [γ^{32}P]-ATP au niveau de Clp1 – ainsi que la cible.

CONCLUSION GENERALE & PERSPECTIVES

CONCLUSION GENERALE

Dans les cellules eucaryotes, les ARNs messagers sont synthétisés dans le noyau par l'ARN polymérase II. La synthèse et la maturation des nouveaux transcrits (ou pré-ARNm) pour obtenir un ARN messager mature sont des processus qui requiert plusieurs étapes impliquant de nombreuses interactions protéines-acides nucléiques. Les différentes étapes englobent aussi bien l'initiation de la transcription, l'élongation, la terminaison et la modification du transcrit naissant par l'ajout d'une coiffe (m^7GpppN) à l'extrémité 5', l'épissage des introns non codants et l'ajout d'une queue poly(A) à l'extrémité 3'. La maturation de l'extrémité 3' regroupe deux étapes : le clivage endonucléolytique du transcrit au niveau d'un site spécifique et l'ajout d'une queue poly(A) sur le fragment en amont du site de clivage ; le fragment en aval étant rapidement dégradé. La taille de la queue poly(A) est espèce spécifique et permet de moduler la stabilité des ARNm, leur export vers le cytoplasme, leur traduction et leur durée de vie. La polyadénylation est fortement conservée aussi bien chez les eucaryotes supérieurs que chez les levures et la plupart des facteurs impliqués présentent des homologues dans le système mammifère.

Des nombreux facteurs *cis* et *trans* sont impliqués dans ces deux étapes de la maturation de l'extrémité 3'. Cependant, on sait peu de choses sur la coopération pouvant exister entre les différentes protéines afin d'assurer une maturation correcte. Chez la levure *S. cerevisiae*, deux complexes multi-protéiques, regroupant vingt et une sous-unités, sont nécessaires : le « *Cleavage and Polyadenylation Factor* » (CPF), le « *Cleavage Factor IA* » (CF IA). D'autres facteurs sont nécessaires pour le positionnement de la machinerie comme Hrp1p/Nab4p (ou « *Cleavage Factor IB* » - CF IB), ainsi que Pab1p et Nab2p permettant le contrôle de la taille de la queue poly(A).

Dans une première partie de ce projet, nous nous sommes intéressés à Pcf11, une des sous-unités du complexe CF IA. Pcf11 est une protéine multifonctionnelle impliquée dans la maturation de l'extrémité 3', l'export ainsi que la terminaison de transcription des ARNs codants et non codants. L'objectif était d'essayer de comprendre le rôle de deux des sous-domaines de cette protéine : le 2nd NTD – région N-terminale – et les motifs liant le zinc – région C-terminale. L'ensemble des résultats de cette première partie ont été obtenus en

collaboration avec deux équipes à l'IECB : l'équipe de C. Mackereth (RMN) et l'équipe de S. Fribourg (Cristallographie).

L'étude par génétique inverse adoptée pour l'étude du 2nd NTD nous a permis d'obtenir quinze mutants, dont les phénotypes associés et les résultats obtenus sont regroupés dans le *TABLEAU 2.1*. La délétion du 2nd NTD conduit à un phénotype de croissance lente, causé par un défaut de la terminaison de transcription des ARNs non codants. L'assemblage du complexe CF IA ainsi que les activités de maturation de l'extrémité 3' des ARNs messagers ne sont pas perturbées. Lors de la terminaison de transcription des ARNs non codants, dépendante du complexe NNS, le 2nd NTD permettrait le maintien du complexe NNS lorsque le CID de Pcf11 est engagé dans l'interaction avec le CTD de l'ARN polymérase II ; le 2nd NTD permettrait la translocation de Sen1 nécessaire au décrochage de la polymérase. La fonction du 2nd NTD dans la terminaison de transcription des ARNs codants, dépendant du complexe de polyadénylation n'a pas pu être directement étudiée. Les résultats obtenus permettent cependant de suggérer que la terminaison de transcription des ARNs codants est indépendante du 2nd NTD de Pcf11. Par l'analyse de la terminaison de transcription des XUTs et SUTs, en absence du 2nd NTD, nous espérons pouvoir mieux comprendre le rôle de ce domaine. Si les résultats sont confirmés par les différentes analyses, le 2nd NTD présenterait une fonction unique, non redondante à celle décrite pour le CID mais complémentaire à ce domaine. La région considérée comme nécessaire et suffisante pour la terminaison de transcription des snoRNAs serait alors étendue à l'ensemble de la région N-terminale de la protéine, des résidus 1 à 233.

L'étude des motifs liant le zinc, par génétique inverse, n'a pas permis d'obtenir de réponses claires quant au rôle de ces deux motifs, conservés dans le règne eucaryote. Nous avons pu montrer que la mutation des résidus Cystéines de l'un de ces deux motifs conduit à une modification des activités de maturation de l'extrémité 3' mais nous ne connaissons pas leur fonction précise. L'assemblage du complexe CF IA ainsi que la terminaison de transcription des snoRNAs ne sont pas perturbées. La différence de phénotype observée (*TABLEAU 2.2*) laisse suggérer que ces motifs peuvent avoir un fonctionnement indépendant. Les motifs liant le zinc à l'extrémité C-terminale de Pcf11 ferait le lien entre les complexes CF IA et CPF, afin d'assurer une maturation correcte. L'étude de ces interactions par *pull down* nous permettra d'identifier les partenaires et ainsi de comprendre le rôle de ces motifs et d'améliorer notre connaissance des relations pouvant exister entre les différentes sous-unités du complexe de polyadénylation. La conservation de ces motifs dans le règne eucaryote laisse à penser que ces relations pourrait être conservés dans le système mammifère.

Dans une seconde partie de ce projet, nous nous somme intéressé au complexe de polyadénylation dans son ensemble et la présence d'une activité de phosphorylation associée aux sous-complexes CF IA et CPF. Notre hypothèse de départ était que Clp1, au vu de ces caractéristiques structurales, possède une activité de type protéine kinase uniquement en présence des autres partenaires du/des complexe(s) CF IA et/ou CPF. Les résultats préliminaires nous ont permis de mettre en évidence une activité protéine kinase, co-purifiant lors de la préparation des sous-complexes CF IA et CPF, indépendamment de la présence de Clp1. Dès le début de ce projet, nous avons rencontré des difficultés de reproductibilité, pouvant être attribuée à un protocole de purification non optimisé. La concentration des facteurs obtenus en fin de purification a permis de rétablir un profil de phosphorylation comparable. La protéine kinase associée au complexe ainsi que les cibles de cette phosphorylation reste à être identifié. Comme l'épissage ou l'export, la maturation de l'extrémité 3' serait spécifiquement régulée par phosphorylation, par association d'une protéine kinase. La phosphorylation des sous-unités du complexe permettrait de moduler leur assemblage ou l'accessibilité du pré-ARNm.

Bibliographie

BIBLIOGRAPHIE

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. Molecular cell *13*, 67-76.

Al Husini, N., Kudla, P., and Ansari, A. (2013). A role for CF1A 3' end processing complex in promoter-associated transcription. PLoS genetics *9*, e1003722.

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. The EMBO journal *18*, 5399-5410.

Alzu, A., Bermejo, R., Begnis, M., Lucca, C., Piccini, D., Carotenuto, W., Saponaro, M., Brambati, A., Cocito, A., Foiani, M.*, et al.* (2012). Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. Cell *151*, 835-846.

Amanchy, R., Kalume, D.E., Iwahori, A., Zhong, J., and Pandey, A. (2005). Phosphoproteome analysis of HeLa cells using stable isotope labeling with amino acids in cell culture (SILAC). J Proteome Res *4*, 1661-1671.

Amrani, N., Minet, M., Le Gouar, M., Lacroute, F., and Wyers, F. (1997a). Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro. Molecular and cellular biology *17*, 3694-3701.

Amrani, N., Minet, M., Wyers, F., Dufour, M.E., Aggerbeck, L.P., and Lacroute, F. (1997b). PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I. Molecular and cellular biology *17*, 1102-1109.

Anderson, J.T., Wilson, S.M., Datar, K.V., and Swanson, M.S. (1993). NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. Molecular and cellular biology *13*, 2730-2741.

Aranda, A., and Proudfoot, N. (2001). Transcriptional termination factors for RNA polymerase II in yeast. Molecular cell *7*, 1003-1011.

Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Molecular cell *23*, 841-851.

Arndt, K.M., and Reines, D. (2015). Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. Annual review of biochemistry *84*, 381-404.

August, J.T., Ortiz, P.J., and Hurwitz, J. (1962). Ribonucleic acid-dependent ribonucleotide incorporation. I. Purification and properties of the enzyme. The Journal of biological chemistry *237*, 3786-3793.

Barabino, S.M., Hubner, W., Jenny, A., Minvielle-Sebastia, L., and Keller, W. (1997). The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. Genes & development *11*, 1703-1716.

Barabino, S.M., Ohnacker, M., and Keller, W. (2000). Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs. The EMBO journal *19*, 3778-3787.

Bard, J., Zhelkovsky, A.M., Helmling, S., Earnest, T.N., Moore, C.L., and Bohm, A. (2000). Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. Science *289*, 1346-1349.

Barilla, D., Lee, B.A., and Proudfoot, N.J. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America *98*, 445-450.

Barnwal, R.P., Lee, S.D., Moore, C., and Varani, G. (2012). Structural and biochemical analysis of the assembly and function of the yeast pre-mRNA 3' end processing complex CF I. Proceedings of the National Academy of Sciences of the United States of America *109*, 21342-21347.

Bartkowiak, B., Mackellar, A.L., and Greenleaf, A.L. (2011). Updating the CTD Story: From Tail to Epic. Genet Res Int *2011*, 623718.

Batisse, J., Batisse, C., Budd, A., Bottcher, B., and Hurt, E. (2009). Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure. The Journal of biological chemistry *284*, 34911-34917.

Becker, E., Meyer, V., Madaoui, H., and Guerois, R. (2006). Detection of a tandem BRCT in Nbs1 and Xrs2 with functional implications in the DNA damage response. Bioinformatics *22*, 1289-1292.

Bernstein, J., and Toth, E.A. (2012). Yeast nuclear RNA processing. World J Biol Chem *3*, 7-26.

Bird, G., Zorio, D.A., and Bentley, D.L. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. Molecular and cellular biology *24*, 8963-8969.

Birnboim, H.C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic acids research *7*, 1513-1523.

Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. (1998). Coupling termination of transcription to messenger RNA maturation in yeast. Science *280*, 298-301.

Blum, E., Carpousis, A.J., and Higgins, C.F. (1999). Polyadenylation promotes degradation of 3'-structured RNA by the Escherichia coli mRNA degradosome in vitro. The Journal of biological chemistry *274*, 4009-4016.

Boettcher, B.R., and Meister, A. (1980). Covalent modification of the active site of carbamyl phosphate synthetase by 5'-p-fluorosulfonylbenzoyladenosine. Direct evidence for two functionally different ATP-binding sites. The Journal of biological chemistry *255*, 7129-7133.

Bonneaud, N., Minvielle-Sebastia, L., Cullin, C., and Lacroute, F. (1994). Cellular localization of RNA14p and RNA15p, two yeast proteins involved in mRNA stability. Journal of cell science *107 (Pt 4)*, 913-921.

Boube, M., Joulia, L., Cribbs, D.L., and Bourbon, H.M. (2002). Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. Cell *110*, 143-151.

Braberg, H., Jin, H., Moehle, E.A., Chan, Y.A., Wang, S., Shales, M., Benschop, J.J., Morris, J.H., Qiu, C., Hu, F.*, et al.* (2013). From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. Cell *154*, 775-788.

Braun, I.C., Rohrbach, E., Schmitt, C., and Izaurralde, E. (1999). TAP binds to the constitutive transport element (CTE) through a novel RNA-binding motif that is sufficient to promote CTE-dependent RNA export from the nucleus. The EMBO journal *18*, 1953-1965.

Brodsky, A.S., and Silver, P.A. (2000). Pre-mRNA processing factors are required for nuclear export. Rna *6*, 1737-1749.

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

Brown, J.W., Marshall, D.F., and Echeverria, M. (2008). Intronic noncoding RNAs and splicing. Trends Plant Sci *13*, 335-342.

Bruelle, C., Bedard, M., Blier, S., Gauthier, M., Traish, A.M., and Vincent, M. (2011). The mitotic phosphorylation of p54(nrb) modulates its RNA binding activity. Biochem Cell Biol *89*, 423-433.

Bucheli, M.E., He, X., Kaplan, C.D., Moore, C.L., and Buratowski, S. (2007). Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI. Rna *13*, 1756-1764.

Butler, J.S., and Platt, T. (1988). RNA processing generates the mature 3' end of yeast CYC1 messenger RNA in vitro. Science *242*, 1270-1274.

Calvo, O., and Manley, J.L. (2003). Strange bedfellows: polyadenylation factors at the promoter. Genes & development *17*, 1321-1327.

Cao, G.J., and Sarkar, N. (1992). Identification of the gene for an Escherichia coli poly(A) polymerase. Proceedings of the National Academy of Sciences of the United States of America *89*, 10380-10384.

Caponigro, G., and Parker, R. (1996). Mechanisms and control of mRNA turnover in Saccharomyces cerevisiae. Microbiol Rev *60*, 233-249.

Caron, N.S., Desmond, C.R., Xia, J., and Truant, R. (2013). Polyglutamine domain flexibility mediates the proximity between flanking sequences in huntingtin. Proceedings of the National Academy of Sciences of the United States of America *110*, 14610-14615.

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

Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., 3rd, Ule, J., Manley, J.L., and Shi, Y. (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. Genes & development *28*, 2370-2380.

Chanfreau, G., Noble, S.M., and Guthrie, C. (1996). Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). Science *274*, 1511-1514.

Chavez, S., Beilharz, T., Rondon, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera, A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. The EMBO journal *19*, 5824-5834.

Chen, J., Martin, B.L., and Brautigan, D.L. (1992). Regulation of protein serine-threonine phosphatase type-2A by tyrosine phosphorylation. Science *257*, 1261-1264.

Cheng, H., He, X., and Moore, C. (2004). The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3. Molecular and cellular biology *24*, 2932-2943.

Chinchilla, K., Rodriguez-Molina, J.B., Ursic, D., Finkel, J.S., Ansari, A.Z., and Culbertson, M.R. (2012). Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in Saccharomyces cerevisiae. Eukaryot Cell *11*, 417-429.

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Molecular cell *40*, 179-204.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes & development *11*, 2755-2766.

Colin, J., Candelli, T., Porrua, O., Boulay, J., Zhu, C., Lacroute, F., Steinmetz, L.M., and Libri, D. (2014). Roadblock termination by reb1p restricts cryptic and readthrough transcription. Molecular cell *56*, 667-680.

Collart, M.A., and Panasenko, O.O. (2012). The Ccr4--not complex. Gene *492*, 42-53.

Condon, C. (2003). RNA processing and degradation in Bacillus subtilis. Microbiol Mol Biol Rev *67*, 157-174, table of contents.

Connelly, S., and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes & development *2*, 440-452.

Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., and Corden, J.L. (2000). A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. Genetics *154*, 557-571.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863-1876.

Danckwardt, S., Hentze, M.W., and Kulozik, A.E. (2008). 3' end mRNA processing: molecular mechanisms and implications for health and disease. The EMBO journal *27*, 482-498.

Daneholt, B. (2001). Assembly and transport of a premessenger RNP particle. Proceedings of the National Academy of Sciences of the United States of America *98*, 7012-7017.

de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H., and Keller, W. (2000). Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. The EMBO journal *19*, 5895-5904.

del Olmo, M., Mizrahi, N., Gross, S., and Moore, C.L. (1997). The Uba2 and Ufd1 proteins of Saccharomyces cerevisiae interact with poly(A) polymerase and affect the polyadenylation activity of cell extracts. Mol Gen Genet *255*, 209-218.

Dheur, S., Nykamp, K.R., Viphakone, N., Swanson, M.S., and Minvielle-Sebastia, L. (2005). Yeast mRNA Poly(A) tail length control can be reconstituted in vitro in the absence of Pab1p-dependent Poly(A) nuclease activity. The Journal of biological chemistry *280*, 24532-24538.

Dheur, S., Vo le, T.A., Voisinet-Hakil, F., Minet, M., Schmitter, J.M., Lacroute, F., Wyers, F., and Minvielle-Sebastia, L. (2003). Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation. The EMBO journal *22*, 2831-2840.

Dichtl, B., Aasland, R., and Keller, W. (2004). Functions for S. cerevisiae Swd2p in 3' end formation of specific mRNAs and snoRNAs and global histone 3 lysine 4 methylation. Rna *10*, 965-977.

Dichtl, B., Blank, D., Ohnacker, M., Friedlein, A., Roeder, D., Langen, H., and Keller, W. (2002a). A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. Molecular cell *10*, 1139-1150.

Dichtl, B., Blank, D., Sadowski, M., Hubner, W., Weiser, S., and Keller, W. (2002b). Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. The EMBO journal *21*, 4125-4135.

Dichtl, B., and Keller, W. (2001). Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. The EMBO journal *20*, 3197-3209.

Dikfidan, A., Loll, B., Zeymer, C., Magler, I., Clausen, T., and Meinhart, A. (2014). RNA specificity and regulation of catalysis in the eukaryotic polynucleotide kinase Clp1. Molecular cell *54*, 975-986.

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

Dominski, Z., and Marzluff, W.F. (2007). Formation of the 3' end of histone mRNA: getting closer to the end. Gene *396*, 373-390.

Dominski, Z., Yang, X.C., Purdy, M., and Marzluff, W.F. (2003). Cloning and characterization of the Drosophila U7 small nuclear RNA. Proceedings of the National Academy of Sciences of the United States of America *100*, 9422-9427.

Dupin, A.F., and Fribourg, S. (2014). Structural basis for ATP loss by Clp1p in a G135R mutant protein. Biochimie.

Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. Trends in genetics : TIG *24*, 280-288.

Epshtein, V., Dutta, D., Wade, J., and Nudler, E. (2010). An allosteric mechanism of Rho-dependent transcription termination. Nature *463*, 245-249.

Ezeokonkwo, C., Zhelkovsky, A., Lee, R., Bohm, A., and Moore, C.L. (2011). A flexible linker region in Fip1 is needed for efficient mRNA polyadenylation. Rna *17*, 652-664.

Fatica, A., Morlando, M., and Bozzoni, I. (2000). Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. The EMBO journal *19*, 6218-6229.

Feaver, W.J., Svejstrup, J.Q., Henry, N.L., and Kornberg, R.D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell *79*, 1103-1109.

Feng, Z.H., Wilson, S.E., Peng, Z.Y., Schlender, K.K., Reimann, E.M., and Trumbly, R.J. (1991). The yeast GLC7 gene required for glycogen accumulation encodes a type 1 protein phosphatase. The Journal of biological chemistry *266*, 23796-23801.

Filipowicz, W., and Pogacic, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. Current opinion in cell biology *14*, 319-327.

Fischer, T., Strasser, K., Racz, A., Rodriguez-Navarro, S., Oppizzi, M., Ihrig, P., Lechner, J., and Hurt, E. (2002). The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. The EMBO journal *21*, 5843-5852.

Fribourg, S., and Conti, E. (2003). Structural similarity in the absence of sequence homology of the messenger RNA export factors Mtr2 and p15. EMBO Rep *4*, 699-703.

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m(6)A RNA methylation. Nature reviews. Genetics *15*, 293-306.

Gaillard, H., and Aguilera, A. (2014). Cleavage Factor I Links Transcription Termination to DNA Damage Response and Genome Integrity Maintenance in Saccharomyces cerevisiae. PLoS genetics *10*, e1004203.

Ganem, C., Devaux, F., Torchet, C., Jacq, C., Quevillon-Cheruel, S., Labesse, G., Facca, C., and Faye, G. (2003). Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. The EMBO journal *22*, 1588-1598.

Garcia-Gimeno, M.A., Munoz, I., Arino, J., and Sanz, P. (2003). Molecular characterization of Ypi1, a novel Saccharomyces cerevisiae type 1 protein phosphatase inhibitor. The Journal of biological chemistry *278*, 47744-47752.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M.*, et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Ghazy, M.A., Gordon, J.M., Lee, S.D., Singh, B.N., Bohm, A., Hampsey, M., and Moore, C. (2012). The interaction of Pcf11 and Clp1 is needed for mRNA 3'-end formation and is modulated by amino acids in the ATP-binding site. Nucleic acids research *40*, 1214-1225.

Ghazy, M.A., He, X., Singh, B.N., Hampsey, M., and Moore, C. (2009). The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3'-end processing. Molecular and cellular biology *29*, 2296-2307.

Ghosh, A., and Lima, C.D. (2010). Enzymology of RNA cap synthesis. Wiley Interdiscip Rev RNA *1*, 152-172.

Gietz, R.D., and Woods, R.A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol *350*, 87-96.

Gilbert, W., and Guthrie, C. (2004). The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA. Molecular cell *13*, 201-212.

Gilbert, W., Siebel, C.W., and Guthrie, C. (2001). Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. Rna *7*, 302-313.

Gilmartin, G.M., Schaufele, F., Schaffner, G., and Birnstiel, M.L. (1988). Functional analysis of the sea urchin U7 small nuclear RNA. Molecular and cellular biology *8*, 1076-1084.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M.*, et al.* (1996). Life with 6000 genes. Science *274*, 546, 563-547.

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

Gordon, J.M., Shikov, S., Kuehner, J.N., Liriano, M., Lee, E., Stafford, W., Poulsen, M.B., Harrison, C., Moore, C., and Bohm, A. (2011). Reconstitution of CF IA from overexpressed subunits reveals stoichiometry and provides insights into molecular topology. Biochemistry *50*, 10203-10214.

Graber, J.H., Cantor, C.R., Mohr, S.C., and Smith, T.F. (1999). Genomic detection of new yeast pre-mRNA 3'-end-processing signals. Nucleic acids research *27*, 888-894.

Grant, R.P., Marshall, N.J., Yang, J.C., Fasken, M.B., Kelly, S.M., Harreman, M.T., Neuhaus, D., Corbett, A.H., and Stewart, M. (2008). Structure of the N-terminal Mlp1-binding domain of the Saccharomyces cerevisiae mRNA-binding protein, Nab2. Journal of molecular biology *376*, 1048-1059.

Green, D.M., Marfatia, K.A., Crafton, E.B., Zhang, X., Cheng, X., and Corbett, A.H. (2002). Nab2p is required for poly(A) RNA export in Saccharomyces cerevisiae and is regulated by arginine methylation via Hmt1p. The Journal of biological chemistry *277*, 7752-7760.

Gromak, N., West, S., and Proudfoot, N.J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. Molecular and cellular biology *26*, 3986-3996.

Gross, S., and Moore, C. (2001a). Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. Proceedings of the National Academy of Sciences of the United States of America *98*, 6080-6085.

Gross, S., and Moore, C.L. (2001b). Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation. Molecular and cellular biology *21*, 8045-8055.

Grzechnik, P., Gdula, M.R., and Proudfoot, N.J. (2015). Pcf11 orchestrates transcription termination pathways in yeast. Genes & development *29*, 849-861.

Gu, M., and Lima, C.D. (2005). Processing the message: structural insights into capping and decapping mRNA. Current opinion in structural biology *15*, 99-106.

Gudipati, R.K., Neil, H., Feuerbach, F., Malabat, C., and Jacquier, A. (2012). The yeast RPL9B gene is regulated by modulation between two modes of transcription termination. The EMBO journal *31*, 2427-2437.

Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nature structural & molecular biology *15*, 786-794.

Guo, Z., and Sherman, F. (1996). 3'-end-forming signals of yeast mRNA. Trends in biochemical sciences *21*, 477-481.

Gwizdek, C., Iglesias, N., Rodriguez, M.S., Ossareh-Nazari, B., Hobeika, M., Divita, G., Stutz, F., and Dargemont, C. (2006). Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. Proceedings of the National Academy of Sciences of the United States of America *103*, 16376-16381.

Haddad, R., Maurice, F., Viphakone, N., Voisinet-Hakil, F., Fribourg, S., and Minvielle-Sebastia, L. (2012). An essential role for Clp1 in assembly of polyadenylation complex CF IA and Pol II transcription termination. Nucleic acids research *40*, 1226-1239.

Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche Nel, H., Gostan, T., Vandermoere, F., Robert, M.C., Cusack, S.*, et al.* (2013). CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. Nature structural & molecular biology *20*, 1358-1366.

Hammell, C.M., Gross, S., Zenklusen, D., Heath, C.V., Stutz, F., Moore, C., and Cole, C.N. (2002). Coupling of termination, 3' processing, and mRNA export. Molecular and cellular biology *22*, 6441-6457.

Hampsey, M. (1997). A review of phenotypes in Saccharomyces cerevisiae. Yeast *13*, 1099-1133.

Hanada, T., Weitzer, S., Mair, B., Bernreuther, C., Wainger, B.J., Ichida, J., Hanada, R., Orthofer, M., Cronin, S.J., Komnenovic, V.*, et al.* (2013). CLP1 links tRNA metabolism to progressive motor-neuron loss. Nature *495*, 474-480.

Hani, J., Schelbert, B., Bernhardt, A., Domdey, H., Fischer, G., Wiebauer, K., and Rahfeld, J.U. (1999). Mutations in a peptidylprolyl-cis/trans-isomerase gene lead to a defect in 3'-end formation of a pre-mRNA in Saccharomyces cerevisiae. The Journal of biological chemistry *274*, 108-116.

Hartman, J.L.t., Garvik, B., and Hartwell, L. (2001). Principles for the buffering of genetic variation. Science *291*, 1001-1004.

Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., McKinney, K.M., Eaton, M.L.*, et al.* (2015). BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. Molecular cell *57*, 636-647.

He, X., Khan, A.U., Cheng, H., Pappas, D.L., Jr., Hampsey, M., and Moore, C.L. (2003). Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. Genes & development *17*, 1030-1042.

He, X., and Moore, C. (2005). Regulation of yeast mRNA 3' end processing by phosphorylation. Molecular cell *19*, 619-629.

Hector, R.E., Nykamp, K.R., Dheur, S., Anderson, J.T., Non, P.J., Urbinati, C.R., Wilson, S.M., Minvielle-Sebastia, L., and Swanson, M.S. (2002). Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. The EMBO journal *21*, 1800-1810.

Helmling, S., Zhelkovsky, A., and Moore, C.L. (2001). Fip1 regulates the activity of Poly(A) polymerase through multiple interactions. Molecular and cellular biology *21*, 2026-2037.

Higgs, D.R., and Weatherall, D.J. (1983). Alpha-thalassemia. Curr Top Hematol *4*, 37-97.

Hirose, Y., and Manley, J.L. (1997). Creatine phosphate, not ATP, is required for 3' end cleavage of mammalian pre-mRNA in vitro. The Journal of biological chemistry *272*, 29636-29642.

Hirose, Y., and Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93-96.

Holbein, S., Scola, S., Loll, B., Dichtl, B.S., Hubner, W., Meinhart, A., and Dichtl, B. (2011). The P-loop domain of yeast Clp1 mediates interactions between CF IA and CPF factors in pre-mRNA 3' end formation. PloS one *6*, e29139.

Hollingworth, D., Noble, C.G., Taylor, I.A., and Ramos, A. (2006). RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11. Rna *12*, 555-560.

Holstege, F.C., Fiedler, U., and Timmers, H.T. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. The EMBO journal *16*, 7468-7480.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes & development *26*, 2119-2137.

Hsin, J.P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. Science *334*, 683-686.

Huez, G., Marbaix, G., Hubert, E., Cleuter, Y., Leclercq, M., Chantrenne, H., Devos, R., Soreq, H., Nudel, U., and Littauer, U.Z. (1975). Readenylation of polyadenylate-free globin messenger RNA restores its stability in vivo. Eur J Biochem *59*, 589-592.

Hughes, F.J., and Romanos, M.A. (1993). E1 protein of human papillomavirus is a DNA helicase/ATPase. Nucleic acids research *21*, 5817-5823.

Hurt, J.A., and Silver, P.A. (2008). mRNA nuclear export and human disease. Dis Model Mech *1*, 103-108.

Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A.H., Dargemont, C., and Stutz, F. (2010). Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes & development *24*, 1927-1938.

Jain, R., and Shuman, S. (2009). Characterization of a thermostable archaeal polynucleotide kinase homologous to human Clp1. Rna *15*, 923-931.

Jia, H., Wang, X., Liu, F., Guenther, U.P., Srinivasan, S., Anderson, J.T., and Jankowsky, E. (2011). The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. Cell *145*, 890-901.

Jimeno, S., Rondon, A.G., Luna, R., and Aguilera, A. (2002). The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. The EMBO journal *21*, 3526-3535.

Johnson, S.A., Cubberley, G., and Bentley, D.L. (2009). Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11. Molecular cell *33*, 215-226.

Johnson, S.A., Kim, H., Erickson, B., and Bentley, D.L. (2011). The export factor Yra1 modulates mRNA 3' end processing. Nature structural & molecular biology *18*, 1164-1171.

Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Molecular cell *12*, 5-14.

Kapoor, P., Bao, Y., Xiao, J., Luo, J., Shen, J., Persinger, J., Peng, G., Ranish, J., Bartholomew, B., and Shen, X. (2015). Regulation of Mec1 kinase activity by the SWI/SNF chromatin remodeling complex. Genes & development *29*, 591-602.

Karaca, E., Weitzer, S., Pehlivan, D., Shiraishi, H., Gogakos, T., Hanada, T., Jhangiani, S.N., Wiszniewski, W., Withers, M., Campbell, I.M.*, et al.* (2014). Human CLP1 Mutations Alter tRNA Biogenesis, Affecting Both Peripheral and Central Nervous System Function. Cell *157*, 636-650.

Katahira, J. (2012). mRNA export and the TREX complex. Biochim Biophys Acta *1819*, 507-513.

Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. (2004). Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. The EMBO journal *23*, 616-626.

Kessler, M.M., Henry, M.F., Shen, E., Zhao, J., Gross, S., Silver, P.A., and Moore, C.L. (1997). Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. Genes & development *11*, 2545-2556.

Kessler, M.M., Zhao, J., and Moore, C.L. (1996). Purification of the Saccharomyces cerevisiae cleavage/polyadenylation factor I. Separation into two components that are required for both cleavage and polyadenylation of mRNA 3' ends. The Journal of biological chemistry *271*, 27167-27175.

Kim Guisbert, K.S., Li, H., and Guthrie, C. (2007). Alternative 3' pre-mRNA processing in Saccharomyces cerevisiae is modulated by Nab4/Hrp1 in vivo. PLoS biology *5*, e6.

Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. Nature structural & molecular biology *17*, 1279-1286.

Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004a). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. The EMBO journal *23*, 354-364.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004b). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517-522.

Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct pathways for snoRNA and mRNA termination. Molecular cell *24*, 723-734.

Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell *109*, 145-148.

Kleiman, F.E., and Manley, J.L. (2001). The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. Cell *104*, 743-753.

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & development *14*, 2452-2460.

Krishna, S.S., Majumdar, I., and Grishin, N.V. (2003). Structural classification of zinc fingers: survey and summary. Nucleic acids research *31*, 532-550.

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004). Ssu72 Is an RNA polymerase II CTD phosphatase. Molecular cell *14*, 387-394.

Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. Genes & development *26*, 1891-1896.

Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. Nature reviews. Molecular cell biology *12*, 283-294.

Kuhn, A.N., and Kaufer, N.F. (2003). Pre-mRNA splicing in Schizosaccharomyces pombe: regulatory role of a kinase conserved from fission yeast to mammals. Curr Genet *42*, 241-251.

Kyburz, A., Sadowski, M., Dichtl, B., and Keller, W. (2003). The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. Nucleic acids research *31*, 3936-3945.

Lackford, B., Yao, C., Charles, G.M., Weng, L., Zheng, X., Choi, E.A., Xie, X., Wan, J., Xing, Y., Freudenberg, J.M*., et al.* (2014). Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. The EMBO journal *33*, 878-889.

Lanzotti, D.J., Kaygun, H., Yang, X., Duronio, R.J., and Marzluff, W.F. (2002). Developmental control of histone mRNA and dSLBP synthesis during Drosophila embryogenesis and the role of dSLBP in histone mRNA 3' end processing in vivo. Molecular and cellular biology *22*, 2267-2282.

Lee, J., and Bedford, M.T. (2002). PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. EMBO Rep *3*, 268-273.

Lee, S.D., and Moore, C.L. (2014). Efficient mRNA polyadenylation requires a ubiquitin-like domain, a zinc knuckle, and a RING finger domain, all contained in the Mpe1 protein. Molecular and cellular biology *34*, 3955-3967.

Leeper, T.C., Qu, X., Lu, C., Moore, C., and Varani, G. (2010). Novel protein-protein contacts facilitate mRNA 3'-processing signal recognition by Rna15 and Hrp1. Journal of molecular biology *401*, 334-349.

Lei, E.P., Krebber, H., and Silver, P.A. (2001). Messenger RNAs are recruited for nuclear export during transcription. Genes & development *15*, 1771-1782.

Lei, E.P., and Silver, P.A. (2002). Intron status and 3'-end formation control cotranscriptional export of mRNA. Genes & development *16*, 2761-2766.

Leipe, D.D., Wolf, Y.I., Koonin, E.V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. Journal of molecular biology *317*, 41-72.

Li, Y., and Kiledjian, M. (2010). Regulation of mRNA decapping. Wiley Interdiscip Rev RNA *1*, 253-265.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. Molecular cell *9*, 1101-1111.

Lima, C.D., Wang, L.K., and Shuman, S. (1999). Structure and mechanism of yeast RNA triphosphatase: an essential component of the mRNA capping apparatus. Cell *99*, 533-543.

Lindstrom, M.S., Jin, A., Deisenroth, C., White Wolf, G., and Zhang, Y. (2007). Cancer-associated mutations in the MDM2 zinc finger domain disrupt ribosomal protein interaction and attenuate MDM2-induced p53 degradation. Molecular and cellular biology *27*, 1056-1068.

Lingner, J., Radtke, I., Wahle, E., and Keller, W. (1991). Purification and characterization of poly(A) polymerase from Saccharomyces cerevisiae. The Journal of biological chemistry *266*, 8741-8746.

Lopez, P.J., and Seraphin, B. (1999). Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition. Rna *5*, 1135-1137.

Losh, J.S., King, A.K., Bakelar, J., Taylor, L., Loomis, J., Rosenzweig, J.A., Johnson, S.J., and van Hoof, A. (2015). Interaction between the RNA-dependent ATPase and poly(A) polymerase subunits of the TRAMP complex is mediated by short peptides and important for snoRNA processing. Nucleic acids research *43*, 1848-1858.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nature structural & molecular biology *17*, 1195-1201.

Luo, W., Johnson, A.W., and Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes & development *20*, 954-965.

MacKellar, A.L., and Greenleaf, A.L. (2011). Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II. The Journal of biological chemistry *286*, 36385-36395.

Mahadevan, S., Raghunand, T.R., Panicker, S., and Struhl, K. (1997). Characterisation of 3' end formation of the yeast HIS3 mRNA. Gene *190*, 69-76.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci *65*, 1099-1122.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature *444*, 953-956.

Mann, M., Ong, S.E., Gronborg, M., Steen, H., Jensen, O.N., and Pandey, A. (2002). Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol *20*, 261-268.

Marquardt, S., Hazelbaker, D.Z., and Buratowski, S. (2011). Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. Transcription *2*, 145-154.

Martin, G., and Keller, W. (1996). Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. The EMBO journal *15*, 2593-2603.

Martin-Tumasz, S., and Brow, D.A. (2015). Saccharomyces cerevisiae Sen1 Helicase Domain Exhibits 5'- to 3'-Helicase Activity with a Preference for Translocation on DNA Rather than RNA. The Journal of biological chemistry *290*, 22880-22889.

Matthews, J.M., and Sunde, M. (2002). Zinc fingers--folds for many occasions. IUBMB Life *54*, 351-355.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012a). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science *336*, 1723-1725.

Mayer, A., Schreieck, A., Lidschreiber, M., Leike, K., Martin, D.E., and Cramer, P. (2012b). The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I. Molecular and cellular biology *32*, 1321-1331.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357-361.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature *430*, 223-226.

Meinhart, A., Silberzahn, T., and Cramer, P. (2003). The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase. The Journal of biological chemistry *278*, 15917-15921.

Mendez-Vidal, C., Wilhelm, M.T., Hellborg, F., Qian, W., and Wiman, K.G. (2002). The p53-induced mouse zinc finger protein wig-1 binds double-stranded RNA with high affinity. Nucleic acids research *30*, 1991-1996.

Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic acids research *38*, 2757-2774.

Minvielle-Sebastia, L., Beyer, K., Krecic, A.M., Hector, R.E., Swanson, M.S., and Keller, W. (1998). Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. The EMBO journal *17*, 7454-7468.

Minvielle-Sebastia, L., Preker, P.J., and Keller, W. (1994). RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. Science *266*, 1702-1705.

Minvielle-Sebastia, L., Preker, P.J., Wiederkehr, T., Strahm, Y., and Keller, W. (1997). The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. Proceedings of the National Academy of Sciences of the United States of America *94*, 7897-7902.

Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Molecular cell *41*, 21-32.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. Cell *91*, 457-466.

Mizrahi, N., and Moore, C. (2000). Posttranslational phosphorylation and ubiquitination of the Saccharomyces cerevisiae Poly(A) polymerase at the S/G(2) stage of the cell cycle. Molecular and cellular biology *20*, 2794-2802.

Mohanty, B.K., and Kushner, S.R. (2000). Polynucleotide phosphorylase functions both as a 3' - 5' exonuclease and a poly(A) polymerase in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America *97*, 11966-11971.

Mohanty, B.K., and Kushner, S.R. (2011). Bacterial/archaeal/organellar polyadenylation. Wiley Interdiscip Rev RNA *2*, 256-276.

Mohanty, B.K., Maples, V.F., and Kushner, S.R. (2004). The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli. Mol Microbiol *54*, 905-920.

Moreno-Morcillo, M., Minvielle-Sebastia, L., Fribourg, S., and Mackereth, C.D. (2011). Locked tether formation by cooperative folding of Rna14p monkeytail and Rna15p hinge domains in the yeast CF IA complex. Structure *19*, 534-545.

Morlando, M., Greco, P., Dichtl, B., Fatica, A., Keller, W., and Bozzoni, I. (2002). Functional analysis of yeast snoRNA and snRNA 3'-end formation mediated by uncoupling of cleavage and polyadenylation. Molecular and cellular biology *22*, 1379-1389.

Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., and Washburn, M.P. (2009). Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Molecular cell *34*, 168-178.

Murthy, K.G., and Manley, J.L. (1992). Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. The Journal of biological chemistry *267*, 14804-14811.

Myers, J.K., Morris, D.P., Greenleaf, A.L., and Oas, T.G. (2001). Phosphorylation of RNA polymerase II CTD fragments results in tight binding to the WW domain from the yeast prolyl isomerase Ess1. Biochemistry *40*, 8479-8486.

Nadeau, O.W., Sacks, D.B., and Carlson, G.M. (1997). The structural effects of endogenous and exogenous Ca2+/calmodulin on phosphorylase kinase. The Journal of biological chemistry *272*, 26202-26209.

Nakazato, H., Venkatesan, S., and Edmonds, M. (1975). Polyadenylic acid sequences in E. coli messenger RNA. Nature *256*, 144-146.

Napolitano, G., Lania, L., and Majello, B. (2014). RNA polymerase II CTD modifications: how many tales from a single tail. J Cell Physiol *229*, 538-544.

Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C.L., and Greenblatt, J. (2003). Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. The Journal of biological chemistry *278*, 33000-33010.

Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. (2008). The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. Molecular cell *29*, 577-587.

Nino, C.A., Herissant, L., Babour, A., and Dargemont, C. (2013). mRNA nuclear export in yeast. Chem Rev *113*, 8523-8545.

Noble, C.G., Beuth, B., and Taylor, I.A. (2007). Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. Nucleic acids research *35*, 87-99.

Noble, C.G., Hollingworth, D., Martin, S.R., Ennis-Adeniran, V., Smerdon, S.J., Kelly, G., Taylor, I.A., and Ramos, A. (2005). Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nature structural & molecular biology *12*, 144-151.

Noble, C.G., Walker, P.A., Calder, L.J., and Taylor, I.A. (2004). Rna14-Rna15 assembly mediates the RNA-binding capability of Saccharomyces cerevisiae cleavage factor IA. Nucleic acids research *32*, 3364-3375.

O'Connor, J.P., and Peebles, C.L. (1992). PTA1, an essential gene of Saccharomyces cerevisiae affecting pre-tRNA processing. Molecular and cellular biology *12*, 3843-3856.

O'Hara, E.B., Chekanova, J.A., Ingle, C.A., Kushner, Z.R., Peters, E., and Kushner, S.R. (1995). Polyadenylylation helps regulate mRNA decay in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America *92*, 1807-1811.

Ohnacker, M., Barabino, S.M., Preker, P.J., and Keller, W. (2000). The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex. The EMBO journal *19*, 37-47.

Ozsolak, F., Kapranov, P., Foissac, S., Kim, S.W., Fishilevich, E., Monaghan, A.P., John, B., and Milos, P.M. (2010). Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell *143*, 1018-1029.

Pabo, C.O., Peisach, E., and Grant, R.A. (2001). Design and selection of novel Cys2His2 zinc finger proteins. Annual review of biochemistry *70*, 313-340.

Panse, V.G., Hardeland, U., Werner, T., Kuster, B., and Hurt, E. (2004). A proteome-wide approach identifies sumoylated substrate proteins in yeast. The Journal of biological chemistry *279*, 41346-41351.

Park, N.J., Tsao, D.C., and Martinson, H.G. (2004). The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II carboxyl-terminal repeat domain. Molecular and cellular biology *24*, 4092-4103.

Paulson, A.R., and Tong, L. (2012). Crystal structure of the Rna14-Rna15 complex. Rna *18*, 1154-1162.

Paushkin, S.V., Patel, M., Furia, B.S., Peltz, S.W., and Trotta, C.R. (2004). Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3' end formation. Cell *117*, 311-321.

Pearson, E., and Moore, C. (2014). The evolutionarily conserved Pol II flap loop contributes to proper transcription termination on short yeast genes. Cell Rep *9*, 821-828.

Pedelini, L., Marquina, M., Arino, J., Casamayor, A., Sanz, L., Bollen, M., Sanz, P., and Garcia-Gimeno, M.A. (2007). YPI1 and SDS22 proteins regulate the nuclear localization and function of yeast type 1 phosphatase Glc7. The Journal of biological chemistry *282*, 3282-3292.

Pendurthi, U.R., Alok, D., and Rao, L.V. (1997). Binding of factor VIIa to tissue factor induces alterations in gene expression in human fibroblast cells: up-regulation of poly(A) polymerase. Proceedings of the National Academy of Sciences of the United States of America *94*, 12598-12603.

Perez Canadillas, J.M., and Varani, G. (2003). Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. The EMBO journal *22*, 2821-2830.

Perez-Canadillas, J.M. (2006). Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. The EMBO journal *25*, 3167-3178.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes & development *20*, 2922-2936.

Piekna-Przybylska, D., Decatur, W.A., and Fournier, M.J. (2007). New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. Rna *13*, 305-312.

Porrua, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nature structural & molecular biology *20*, 884-891.

Porrua, O., and Libri, D. (2015a). Characterization of the mechanisms of transcription termination by the helicase Sen1. Methods Mol Biol *1259*, 313-331.

Porrua, O., and Libri, D. (2015b). Transcription termination and the control of the transcriptome: why, where and how to stop. Nature reviews. Molecular cell biology *16*, 190-202.

Preker, P.J., and Guthrie, C. (2006). Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA. Rna *12*, 994-1006.

Preker, P.J., and Keller, W. (1998). The HAT helix, a repetitive motif implicated in RNA processing. Trends in biochemical sciences *23*, 15-16.

Preker, P.J., Kim, K.S., and Guthrie, C. (2002). Expression of the essential mRNA export factor Yra1p is autoregulated by a splicing-dependent mechanism. Rna *8*, 969-980.

Preker, P.J., Lingner, J., Minvielle-Sebastia, L., and Keller, W. (1995). The FIP1 gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase. Cell *81*, 379-389.

Preker, P.J., Ohnacker, M., Minvielle-Sebastia, L., and Keller, W. (1997). A multisubunit 3' end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor. The EMBO journal *16*, 4727-4737.

Prescott, E.M., and Proudfoot, N.J. (2002). Transcriptional collision between convergent genes in budding yeast. Proceedings of the National Academy of Sciences of the United States of America *99*, 8796-8801.

Proudfoot, N., and O'Sullivan, J. (2002). Polyadenylation: a tail of two complexes. Curr Biol *12*, R855-857.

Proudfoot, N.J. (1976). Sequence analysis of the 3' non-coding regions of rabbit alpha- and beta-globin messenger RNAs. Journal of molecular biology *107*, 491-525.

Proudfoot, N.J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. Trends in biochemical sciences *14*, 105-110.

Proudfoot, N.J., and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. Nature *263*, 211-214.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods *24*, 218-229.

Qu, X., Lykke-Andersen, S., Nasser, T., Saguez, C., Bertrand, E., Jensen, T.H., and Moore, C. (2009). Assembly of an export-competent mRNP is needed for efficient release of the 3'-end processing complex after polyadenylation. Molecular and cellular biology *29*, 5327-5338.

Qu, X., Perez-Canadillas, J.M., Agrawal, S., De Baecke, J., Cheng, H., Varani, G., and Moore, C. (2007). The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3'-end processing. The Journal of biological chemistry *282*, 2101-2115.

Ramirez, A., Shuman, S., and Schwer, B. (2008). Human RNA 5'-kinase (hClp1) can function as a tRNA splicing enzyme in vivo. Rna *14*, 1737-1745.

Reeder, R.H., and Lang, W. (1994). The mechanism of transcription termination by RNA polymerase I. Mol Microbiol *12*, 11-15.

Riezman, H., Hase, T., van Loon, A.P., Grivell, L.A., Suda, K., and Schatz, G. (1983). Import of proteins into mitochondria: a 70 kilodalton outer membrane protein with a large carboxy-terminal deletion is still transported to the outer membrane. The EMBO journal *2*, 2161-2168.

Rondon, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. Molecular cell *36*, 88-98.

Rosado-Lugo, J.D., and Hampsey, M. (2014). The Ssu72 phosphatase mediates the RNA polymerase II initiation-elongation transition. The Journal of biological chemistry *289*, 33916-33926.

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

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

Ruby, S.W., and Abelson, J. (1991). Pre-mRNA splicing in yeast. Trends in genetics : TIG *7*, 79-85.

Russnak, R., Nehrke, K.W., and Platt, T. (1995). REF2 encodes an RNA-binding protein directly involved in yeast mRNA 3'-end formation. Molecular and cellular biology *15*, 1689-1697.

Russo, P., Li, W.Z., Hampsey, D.M., Zaret, K.S., and Sherman, F. (1991). Distinct cis-acting signals enhance 3' endpoint formation of CYC1 mRNA in the yeast Saccharomyces cerevisiae. The EMBO journal *10*, 563-571.

Ryan, C.M., Militello, K.T., and Read, L.K. (2003). Polyadenylation regulates the stability of Trypanosoma brucei mitochondrial RNAs. The Journal of biological chemistry *278*, 32753-32762.

Ryan, K., and Bauer, D.L. (2008). Finishing touches: post-translational modification of protein factors involved in mammalian pre-mRNA 3' end formation. The international journal of biochemistry & cell biology *40*, 2384-2396.

Ryan, K., Calvo, O., and Manley, J.L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. Rna *10*, 565-573.

Sachs, A., and Wahle, E. (1993). Poly(A) tail metabolism and function in eucaryotes. The Journal of biological chemistry *268*, 22955-22958.

Sadowski, M., Dichtl, B., Hubner, W., and Keller, W. (2003). Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. The EMBO journal *22*, 2167-2177.

Schaffer, A.E., Eggens, V.R., Caglayan, A.O., Reuter, M.S., Scott, E., Coufal, N.G., Silhavy, J.L., Xue, Y., Kayserili, H., Yasuno, K.*, et al.* (2014). CLP1 Founder Mutation Links tRNA Splicing and Maturation to Cerebellar Development and Neurodegeneration. Cell *157*, 651-663.

Scharl, E.C., and Steitz, J.A. (1994). The site of 3' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. The EMBO journal *13*, 2432-2440.

Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-wide mapping of yeast RNA polymerase II termination. PLoS genetics *10*, e1004632.

Schmitt, M.E., Brown, T.A., and Trumpower, B.L. (1990). A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. Nucleic acids research *18*, 3091-3092.

Schonemann, L., Kuhn, U., Martin, G., Schafer, P., Gruber, A.R., Keller, W., Zavolan, M., and Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. Genes & development *28*, 2381-2393.

Schreieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L.A. (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. Nature structural & molecular biology *21*, 175-179.

Schwer, B., and Shuman, S. (2011). Deciphering the RNA polymerase II CTD code in fission yeast. Molecular cell *43*, 311-318.

Seoane, S., Lamas-Maceiras, M., Rodriguez-Torres, A.M., and Freire-Picos, M.A. (2009). Involvement of Pta1, Pcf11 and a KlCYC1 AU-rich element in alternative RNA 3'-end processing selection in yeast. FEBS letters *583*, 2843-2848.

Setyono, B., and Greenberg, J.R. (1981). Proteins associated with poly(A) and other regions of mRNA and hnRNA molecules as investigated by crosslinking. Cell *24*, 775-783.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Molecular cell *33*, 365-376.

Shuman, S. (2002). What messenger RNA capping tells us about eukaryotic evolution. Nature reviews. Molecular cell biology *3*, 619-625.

Siddiqui, N., and Borden, K.L. (2012). mRNA export and cancer. Wiley Interdiscip Rev RNA *3*, 13-25.

Singh, V., Azad, G.K., Sariki, S.K., and Tomar, R.S. (2015). Flocculation in Saccharomyces cerevisiae is regulated by RNA/DNA helicase Sen1p. FEBS letters.

Sinturel, F., Brechemier-Baey, D., Kiledjian, M., Condon, C., and Benard, L. (2012). Activation of 5'-3' exoribonuclease Xrn1 by cofactor Dcs1 is essential for mitochondrial function in yeast. Proceedings of the National Academy of Sciences of the United States of America *109*, 8264-8269.

Sittman, D.B., Graves, R.A., and Marzluff, W.F. (1983). Structure of a cluster of mouse histone genes. Nucleic acids research *11*, 6679-6697.

Skaar, D.A., and Greenleaf, A.L. (2002). The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p. Molecular cell *10*, 1429-1439.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Molecular cell *42*, 794-805.

Snell, R.G., MacMillan, J.C., Cheadle, J.P., Fenton, I., Lazarou, L.P., Davies, P., MacDonald, M.E., Gusella, J.F., Harper, P.S., and Shaw, D.J. (1993). Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. Nat Genet *4*, 393-397.

Sparks, K.A., and Dieckmann, C.L. (1998). Regulation of poly(A) site choice of several yeast mRNAs. Nucleic acids research *26*, 4676-4687.

Srividya, I., Tirupataiah, S., and Mishra, K. (2012). Yeast transcription termination factor Rtt103 functions in DNA damage response. PloS one *7*, e31288.

Stark, M.J. (1996). Yeast protein serine/threonine phosphatases: multiple roles and diverse regulation. Yeast *12*, 1647-1675.

Steinmetz, E.J., and Brow, D.A. (2003). Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. Molecular and cellular biology *23*, 6339-6349.

Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. Nature *413*, 327-331.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Molecular cell *24*, 735-746.

Strasser, K., and Hurt, E. (2001). Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. Nature *413*, 648-652.

Stumpf, G., and Domdey, H. (1996). Dependence of yeast pre-mRNA 3'-end processing on CFT1: a sequence homolog of the mammalian AAUAAA binding factor. Science *274*, 1517-1520.

Sun, Z.W., and Hampsey, M. (1996). Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo. Molecular and cellular biology *16*, 1557-1566.

Suntharalingam, M., Alcazar-Roman, A.R., and Wente, S.R. (2004). Nuclear export of the yeast mRNA-binding protein Nab2 is linked to a direct interaction with Gfd1 and to Gle1 function. The Journal of biological chemistry *279*, 35384-35391.

Suske, G. (1999). The Sp-family of transcription factors. Gene *238*, 291-300.

Swida, U., Thuroff, E., Steinert, E., and Kaufer, N.F. (1988). A non-conserved sequence in the 5'region of the CYH2 intron from Saccharomyces cerevisiae controls splicing efficiency of the pre-mRNA. Yeast *4*, 209-217.

Tacahashi, Y., Helmling, S., and Moore, C.L. (2003). Functional dissection of the zinc finger and flanking domains of the Yth1 cleavage/polyadenylation factor. Nucleic acids research *31*, 1744-1752.

Takagaki, Y., MacDonald, C.C., Shenk, T., and Manley, J.L. (1992). The human 64-kDa polyadenylylation factor contains a ribonucleoprotein-type RNA binding domain and unusual auxiliary motifs. Proceedings of the National Academy of Sciences of the United States of America *89*, 1403-1407.

Takagaki, Y., and Manley, J.L. (1998). Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. Molecular cell *2*, 761-771.

Takagaki, Y., and Manley, J.L. (2000). Complex protein interactions within the human polyadenylation machinery identify a novel component. Molecular and cellular biology *20*, 1515-1525.

Takagaki, Y., Seipelt, R.L., Peterson, M.L., and Manley, J.L. (1996). The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell *87*, 941-952.

Teixeira, A., Tahiri-Alaoui, A., West, S., Thomas, B., Ramadass, A., Martianov, I., Dye, M., James, W., Proudfoot, N.J., and Akoulitchev, A. (2004). Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. Nature *432*, 526-530.

Thapar, R. (2015). Structural basis for regulation of RNA-binding proteins by phosphorylation. ACS Chem Biol *10*, 652-666.

Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic acids research *33*, 201-212.

Tian, B., and Manley, J.L. (2013). Alternative cleavage and polyadenylation: the long and short of it. Trends in biochemical sciences *38*, 312-320.

Tollervey, D. (2004). Molecular biology: termination by torpedo. Nature *432*, 456-457.

Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M.*, et al.* (2004). Global mapping of the yeast genetic interaction network. Science *303*, 808-813.

Tudek, A., Candelli, T., and Libri, D. (2015). Non-coding transcription by RNA polymerase II in yeast: Hasard or necessite? Biochimie.

Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R.*, et al.* (2014). Molecular basis for coordinating transcription termination with noncoding RNA degradation. Molecular cell *55*, 467-481.

Ueno, T., Takahashi, K., Matsuguchi, T., Ikejiri, K., Endo, H., and Yamamoto, M. (1989). Multiple polyadenylation sites in a large 3'-most exon of the rat insulin-like growth factor II gene. Biochim Biophys Acta *1009*, 27-34.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P.*, et al.* (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623-627.

Ursic, D., Chinchilla, K., Finkel, J.S., and Culbertson, M.R. (2004). Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in

transcription, transcription-coupled DNA repair and RNA processing. Nucleic acids research *32*, 2441-2452.

van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Ne, P., Loeillet, S.*, et al.* (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114-117.

Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS biology *3*, e189.

Varani, G. (1997). A cap for all occasions. Structure *5*, 855-858.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nature structural & molecular biology *15*, 795-804.

Vethantham, V., Rao, N., and Manley, J.L. (2007). Sumoylation modulates the assembly and activity of the pre-mRNA 3' processing complex. Molecular and cellular biology *27*, 8848-8858.

Vethantham, V., Rao, N., and Manley, J.L. (2008). Sumoylation regulates multiple aspects of mammalian poly(A) polymerase function. Genes & development *22*, 499-511.

Vitaliano-Prunier, A., Babour, A., Herissant, L., Apponi, L., Margaritis, T., Holstege, F.C., Corbett, A.H., Gwizdek, C., and Dargemont, C. (2012). H2B ubiquitylation controls the formation of export-competent mRNP. Molecular cell *45*, 132-139.

Vo, L.T., Minet, M., Schmitter, J.M., Lacroute, F., and Wyers, F. (2001). Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. Molecular and cellular biology *21*, 8346-8356.

Wahle, E., and Kuhn, U. (1997). The mechanism of 3' cleavage and polyadenylation of eukaryotic pre-mRNA. Prog Nucleic Acid Res Mol Biol *57*, 41-71.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. The EMBO journal *1*, 945-951.

Walsh, E.P., Lamont, D.J., Beattie, K.A., and Stark, M.J. (2002). Novel interactions of Saccharomyces cerevisiae type 1 protein phosphatase identified by single-step affinity purification and mass spectrometry. Biochemistry *41*, 2409-2420.

Wang, S.W., Asakawa, K., Win, T.Z., Toda, T., and Norbury, C.J. (2005). Inactivation of the pre-mRNA cleavage and polyadenylation factor Pfs2 in fission yeast causes lethal cell cycle defects. Molecular and cellular biology *25*, 2288-2296.

Ward, A.C. (1990). Single-step purification of shuttle vectors from yeast for high frequency back-transformation into E. coli. Nucleic acids research *18*, 5319.

Washburn, R.S., and Gottesman, M.E. (2015). Regulation of transcription elongation and termination. Biomolecules *5*, 1063-1078.

Weitzer, S., Hanada, T., Penninger, J.M., and Martinez, J. (2015). CLP1 as a novel player in linking tRNA splicing to neurodegenerative disorders. Wiley Interdiscip Rev RNA *6*, 47-63.

Weitzer, S., and Martinez, J. (2007a). hClp1: a novel kinase revitalizes RNA metabolism. Cell cycle *6*, 2133-2137.

Weitzer, S., and Martinez, J. (2007b). The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs. Nature *447*, 222-226.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522-525.

West, S., and Proudfoot, N.J. (2008). Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. Nucleic acids research *36*, 905-914.

Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*.

Wood, A.J., Schulz, R., Woodfine, K., Koltowska, K., Beechey, C.V., Peters, J., Bourc'his, D., and Oakey, R.J. (2008). Regulation of alternative polyadenylation by genomic imprinting. Genes & development *22*, 1141-1146.

Xiang, K., Manley, J.L., and Tong, L. (2012). An unexpected binding mode for a Pol II CTD peptide phosphorylated at Ser7 in the active site of the CTD phosphatase Ssu72. Genes & development *26*, 2265-2270.

Xing, D., Zhao, H., Xu, R., and Li, Q.Q. (2008). Arabidopsis PCFS4, a homologue of yeast polyadenylation factor Pcf11p, regulates FCA alternative processing and promotes flowering time. The Plant journal : for cell and molecular biology *54*, 899-910.

Xu, C., and Henry, M.F. (2004). Nuclear export of hnRNP Hrp1p and nuclear export of hnRNP Npl3p are linked and influenced by the methylation state of Npl3p. Molecular and cellular biology *24*, 10742-10756.

Xu, X., Perebaskine, N., Minvielle-Sebastia, L., Fribourg, S., and Mackereth, C.D. (2015). Chemical shift assignments of a new folded domain from yeast Pcf11. Biomol NMR Assign.

Yehudai-Resheff, S., and Schuster, G. (2000). Characterization of the E.coli poly(A) polymerase: nucleotide specificity, RNA-binding affinities and RNA structure dependence. Nucleic acids research *28*, 1139-1144.

Yue, Y., Liu, J., and He, C. (2015). RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. Genes & development *29*, 1343-1355.

Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., and Greenleaf, A.L. (1988). The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. Proceedings of the National Academy of Sciences of the United States of America *85*, 3698-3702.

Zenklusen, D., and Stutz, F. (2001). Nuclear export of mRNA. FEBS letters *498*, 150-156.

Zhang, H., Rigo, F., and Martinson, H.G. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Molecular cell *59*, 437-448.

Zhang, Z., Fu, J., and Gilmour, D.S. (2005). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. Genes & development *19*, 1572-1580.

Zhang, Z., and Gilmour, D.S. (2006). Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. Molecular cell *21*, 65-74.

Zhao, J., Hyman, L., and Moore, C. (1999a). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev *63*, 405-445.

Zhao, J., Kessler, M., Helmling, S., O'Connor, J.P., and Moore, C. (1999b). Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor. Molecular and cellular biology *19*, 7733-7740.

Zhao, J., Kessler, M.M., and Moore, C.L. (1997). Cleavage factor II of Saccharomyces cerevisiae contains homologues to subunits of the mammalian Cleavage/ polyadenylation specificity factor and exhibits sequence-specific, ATP-dependent interaction with precursor RNA. The Journal of biological chemistry *272*, 10831-10838.

Zhelkovsky, A., Helmling, S., and Moore, C. (1998). Processivity of the Saccharomyces cerevisiae poly(A) polymerase requires interactions at the carboxyl-terminal RNA binding domain. Molecular and cellular biology *18*, 5942-5951.

Zhelkovsky, A., Tacahashi, Y., Nasser, T., He, X., Sterzer, U., Jensen, T.H., Domdey, H., and Moore, C. (2006). The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in Saccharomyces cerevisiae. Rna *12*, 435-445.

Zielinski, R., Hellman, U., Kubinski, K., and Szyszka, R. (2006). Fip1--an essential component of the Saccharomyces cerevisiae polyadenylation machinery is phosophorylated by protein kinase CK2. Mol Cell Biochem *286*, 191-197.

TITRE : Étude fonctionnelle des sous-domaines de Pcf11 : rôle du 2nd NTD dans la terminaison de transcription des snoRNAs et des motifs liant le zinc dans les activités de maturation de l'extrémité 3' des ARN messagers.

RESUME : Chez les eucaryotes, la maturation de l'extrémité 3' des ARNs messagers a lieu lors de la transcription et regroupe deux étapes : le clivage endonucléolytique du transcrit au niveau d'un site spécifique et l'ajout d'une queue poly(A) sur le fragment en amont du site de clivage. Chez *S. cerevisiae*, le complexe de polyadénylation est formé par 20 protéines, regroupées principalement en deux sous-complexes : CF IA et CPF. Nous nous intéressons plus spécifiquement à Pcf11, sous-unité du complexe CF IA.

Pcf11 est formé de sept sous-domaines, mais la fonction d'une grande partie de la protéine n'est pour l'instant pas connue. Par exemple, aucune fonction n'est associée à la région située entre le domaine d'interaction avec le CTD de l'ARN polymérase II (CID) et une répétion de 20 résidus glutamines. Récemment, la structure de ce domaine, appelé 2nd NTD a été décrite. Pour essayer de comprendre la fonction du 2nd NTD et des motifs liant le zinc encadrant le domaine d'interaction avec Clp1, nous avons mis en place une stratégie systématique de mutagénèse, soit par délétions soit par mutations ponctuelles.

Le 2nd NTD est formé de trois hélices α et interagit avec l'ARN. La délétion de ce domaine conduit à un phénotype de croissance lente chez la levure et un défaut de terminaison de transcription des snoRNAs. Malgré une similarité de structure et de fonction, le 2nd NTD présenterait une fonction indépendante.

La fonction des motifs liant le zinc n'est pour l'instant pas connue. Cependant, la mutation de l'un de ces deux motifs conduit à un défaut de clivage et de polyadénylation *in vitro*. La mutation des deux motifs est létale chez la levure.

TITLE : Functional analysis of Pcf11 sub-domains : role of the 2nd NTD in transcription termination of snRNAs and zinc finger motifs in 3'-end processing of mRNAs.

ABSTRACT : In eukaryotes, poly(A) tails are added to nuclear pre-mRNA 3'-ends in the two steps of cleavage and polyadenylation. This co-transcriptional processing requires the activity of a large protein complex comprising at least 20 different polypeptides in yeast organized primarily into the two factors CF IA and CPF. We are interested in the functional characterization of Pcf11, a CF IA subunit.

The Pcf11 protein is organized into seven different domains, but here is still a large portion of the polypeptide that has not yet been characterized. For example the region from the end of the CTD interaction domain (CID) to an uninterrupted stretch of 20 glutamine residues has no known function. Recently, the structure of this region, called the 2nd NTD have been characterized. To gain insight into the function of the 2nd NTD and the two zinc fingers motif surounding the Clp1 interaction domain, we have employed a systematic strategy of mutagenesis, either by deletion or via point mutations.

The 2nd NTD is a folded domain composed of three α-helice. The deletion of this domain induced a severe defect of growth in yeast and impaired transcription termination of snoRNAs. Despite it's similarity in structure and function with the CID, the 2nd NTD seems to act like an independant RNA binding domain.

We don't know yet the real function of the two zinc fingers motif at the C-terminal region of Pcf11, but the mutation of Cystein residus into serine of one of the two motifs impaired cleavage and polyadenylation. The mutation of the first motif is less harmful than the mutation of the second motif. The simultaneous mutation is lethal in yeast.

MOTS CLES : ARNs messagers, Maturation en 3', Pcf11, Terminaison transcription, motif liant le zinc, *S. cerevisiae*.

INSERM U869 – LABORATOIRE ARNA – ARN : REGULATIONS NATURELLES ET ARTIFICIELLES
Université de Bordeaux – Bat 3A, 1er étage, 146 rue Léo Saignat, 33076 Bordeaux, France