HAL
open science

Etude du contrôle de la transcription envahissante par la terminaison de la transcription

Jean-Baptiste Briand

To cite this version:

Jean-Baptiste Briand. Etude du contrôle de la transcription envahissante par la terminaison de la transcription. Biologie moléculaire. Université Paris Sud - Paris XI, 2015. Français. NNT : 2015PA112079 . tel-01236631

HAL Id: tel-01236631

https://theses.hal.science/tel-01236631v1

Submitted on 2 Dec 2015

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 PARIS-SUD

ÉCOLE DOCTORALE 426 :
GÈNES GÉNOMES CELLULES

Laboratoire : *Métabolisme et fonction de l'ARN dans le noyau*

THÈSE DE DOCTORAT

SCIENCES DE LA VIE ET DE LA SANTÉ
BIOLOGIE MOLECULAIRE

par

Jean-Baptiste Briand

Etude du contrôle de la transcription envahissante par la terminaison de la transcription

Date de soutenance : 03/06/2015

Composition du jury :

Directeur de thèse :	Domenico Libri	DR CNRS (Institut Jacques Monod, Paris)
Co-encadrante :	Jessie Colin	Chercheur post-doctorant (Institut Jacques Monod, Paris)
Rapporteurs :	Micheline Fromont	DR CNRS (Institut Pasteur, Paris)
	Antonin Morillon	DR CNRS (Institut Curie, Paris)
Examinateurs :	Daniel Gautheret	Professeur (Université Paris Sud, Orsay)
	Stéphane Marcand	DR CEA (CEA, Fontenay-aux-Roses)

Remerciements

Je remercie tout d'abord Domenico Libri pour m'avoir permis d'effectuer ma thèse dans son équipe, pour son encadrement et pour son soutien pendant toute la durée de mon doctorat. Je remercie aussi tout particulièrement Jessie Colin qui par ses conseils et son encadrement a su m'imposer à la fois la rigueur scientifique et le régime alimentaire nécessaire à la réussite d'une thèse.

Je remercie l'ensemble des membres du laboratoire pour leur aide, leurs conseils et les bons moments passés ensemble. J'ai été très heureux de réaliser mes travaux de thèse dans cette équipe : un fabuleux melting pot qui m'a beaucoup apporté, tant professionnellement qu'humainement. Mes connaissances en biologie, en anglais, en jurons italiens et en spécialités culinaires du monde entier se sont largement développées dans cet environnement agréable. Je remercie notamment Jocelyne Boulay et Tito Candelli dont les travaux ont été essentiels à l'avancé de mes projets de thèse.

Je remercie les membres de mon comité de suivi de thèse pour leurs remarques et leurs avis : mon tuteur Olivier Namy, Frank Feuerbach et Dominique Fourmy. Je remercie également les membres de mon jury de thèse : Micheline Fromont-Racine, Antonin Morillon, Daniel Gautheret et Stéphane Marcand qui ont accepté d'évaluer mes travaux.

Je remercie ma famille et particulièrement mes parents pour leur confiance et leur soutien pendant ces nombreuses années d'étude.

Je remercie mes amis grâce auxquels j'ai pu prendre le recul parfois nécessaire sur mes activités scientifiques.

Je suis particulièrement reconnaissant envers les personnes qui m'ont accordé les financements nécessaires au déroulement de mon doctorat. Ces financements sont provenus du Ministère de la Recherche et de l'Enseignement Supérieur, de la Ligue Contre le Cancer et des fonds du laboratoire.

Liste des abréviations

ORF : Open Reading Frame

NFR/NDR : Nucleosome Free Region / Nucleosome Depleted Region

Sn(o)ARN : Small nucle(ol)ar ARN

ARNr : ARN ribosomique

ARNt : ARN de transfert

CUT : Cryptic Unstable Transcript

SUT : Stable Unanotated Transcrit

XUT : Xrn1-sensitive Unstable Transcript

CTD : Carboxy Terminal Domain

PIC : PreInitiation Complex

GTF : General Transcription Factor

TBP : TATA Binding Protein

TSS: Transcription Start Site

TTS : Transcription Termination Site

CPF-CF : Cleavage and Polyadenylation Factor – Cleavage Factor

PFI : Polyadenylation Factor I

CFI/CFII : Cleavage Factor I et II

APT : Associated with Pta1

NNS : Nrd1-Nab3-Sen1

TRAMP : Trf4/5 Air1/2 Mtr4 Polyadenylation complex

CID : CTD Interacting Domain

NIM : Nrd1 Interaction Motif

PAR-CLIP : PhotoActivatable-Ribonucleoside-enhanced CrossLinking and ImmunoPrecipitation

NET-Seq : Native Elongating Transcripts Sequencing

SELEX : Systematic Evolution of Ligands by Exponential enrichment

HDAC : Histone Deacetylase

C-ter : Caboxy-terminal

N-ter : Amino-terminal

uHYP2 : upstream *HYP2*

d*STF1* : downstream *STF1*

Résumé

La terminaison de la transcription est essentielle, aussi bien pour assurer la formation de l'extrémité 3' de transcrits fonctionnels que pour éviter les phénomènes d'interférence transcriptionnelle entre des régions transcrites adjacentes. Ceci est particulièrement important dans un génome compact comme celui de *S. cerevisiae*. La terminaison est aussi l'une des stratégies principales que la cellule emploie pour contrôler et limiter la transcription dite envahissante ou cachée. Chez *S. cerevisiae*, l'ARN polymérase II est responsable de la transcription des ARNm et de nombreuses classes d'ARN non codants tels que les sn(o)ARN et les CUT (Cryptic Unstable Transcripts). Ces derniers représentent une fraction importante des transcrits issus de la transcription cachée.

Il existe deux voies canoniques de terminaison de la transcription par cette polymérase. Elles font intervenir le complexe de clivage et de polyadénylation, CPF-CF, notamment pour la terminaison des ARNm ou le complexe NNS pour la terminaison des sn(o)ARN et des CUT. Au cours de ma thèse j'ai étudié deux aspects de la terminaison de la transcription : 1) l'étude des motifs de recrutement du complexe NNS et 2) l'identification et la caractérisation d'une nouvelle voie de terminaison par le facteur Rap1.

Les complexes CPF-CF et NNS agissent tous les deux en liant le transcrit naissant et l'ARN pol II. Le complexe NNS lie l'ARN naissant grâce à ses sous-unités Nrd1 et Nab3 qui reconnaissent des motifs spécifiques. Cependant, bien que la séquence de ces motifs soit maintenant connue, leur présence ne permet pas de définir de façon certaine un terminateur. En effet, le nombre de ces motifs varie beaucoup d'un terminateur à l'autre. Afin de mieux comprendre la structure des terminateurs ciblés par le complexe NNS et l'organisation des motifs liés par Nrd1 et Nab3, j'ai recherché les séquences impliquées dans la terminaison d'un CUT modèle en réalisant une mutagenèse aléatoire et j'ai identifié par SELEX des motifs de fixation optimale du dimère Nrd1-Nab3.

Un second volet de ma thèse porte sur la caractérisation d'une nouvelle voie de terminaison de la transcription dépendante du facteur Rap1. Rap1 est important

pour la structure des télomères et c'est aussi un facteur de transcription ciblant des centaines de promoteurs. Il active ou réprime l'initiation de la transcription notamment en recrutant des complexes de remodelage de la chromatine sur les promoteurs ciblés. De façon surprenante, le motif de fixation de ce facteur a été identifié dans des séquences capables de terminer la transcription isolées au laboratoire. Mes travaux ont permis de caractériser le mécanisme de terminaison par Rap1 et de distinguer cette voie des voies de terminaison canoniques. Ce facteur, lié à l'ADN, agit comme une barrière en bloquant la progression de l'ARN polymérase II par un mécanisme de « road-block ». Les polymérases ainsi arrêtées sont ciblées par une voie qui permet leur élimination lorsqu'elles sont bloquées par des dégâts sur l'ADN, impliquant leur ubiquitination et vraisemblablement leur dégradation par le protéasome. Les ARN libérés sont polyadénylés par la poly(A)-polymérase Trf4 et dégradés par l'exosome nucléaire. Ce mécanisme de terminaison est utilisé dans un contexte naturel puisque j'ai identifié des transcrits endogènes de *S. cerevisiae* terminés par cette voie. Nous proposons que la terminaison par Rap1 contribue au contrôle de la transcription envahissante. Ce facteur assurerait ainsi au niveau des promoteurs qu'il lie une double fonction de facteur de transcription et de protection de ces promoteurs contre l'interférence transcriptionnelle.

Remerciements 2
Liste des abréviations 3
Résumé 4

I. Introduction 9
I.1. La transcription chez les eucaryotes 9
I.1.1. Diversité et fonctions des ARN 9
I.1.1.1. Les ARN codants 10
I.1.1.2. Les ARN non-codants 10
I.2. La transcription par l'ARN pol II chez *S. cerevisiae* 12
I.2.1. L'ARN pol II 12
I.2.2. L'initiation de la transcription 14
I.2.2.1. Reconnaissance du promoteur 14
I.2.2.2. Formation du PIC 15
I.2.2.3. Structure chromatinienne du promoteur 18
I.2.3. L'élongation de la transcription 20
I.2.3.1. Départ du promoteur 20
I.2.3.2. Elongation de la transcription à travers la chromatine 22
Facteurs de remodelage de la chromatine recrutés lors de l'élongation 23
Modification des histones lors de l'élongation de la transcription 24
L'acétylation des histones 24
La méthylation des histones 24
I.2.3.3. Contrôle de l'arrêt transcriptionnel 27
Facteurs impliqués dans la résolution de l'arrêt transcriptionnel 27
Ubiquitination et dégradation de l'ARN pol II 29
I.2.4. La terminaison de la transcription 32
I.2.4.1. La terminaison par le complexe CPF-CF 32
Les signaux de terminaison sur l'ARN naissant 33
Les facteurs de terminaison du CPF-CF 34
Le CPF 35
Le CFI 38
La poly(A) polymérase Pap1 40
Les protéines de fixation aux extensions poly(A) Pab1 et Nab2 41
Deux mécanismes alternatifs de terminaison 42
I.2.4.2. La terminaison par le complexe Nrd1-Nab3-Sen1 : 45
Le complexe Nrd1-Nab3-Sen1 46
Les signaux de terminaison sur l'ARN naissant 48

Le complexe TRAMP 49
L'Exosome 51
I.2.4.3. Mécanismes de terminaison de la transcription par les ARN polymérases I et III 54
Terminaison de la transcription par l'ARN pol I 54
La terminaison de la transcription par l'ARN pol III 55
I.3. La transcription envahissante 57
I.3.1. Découverte de la transcription envahissante 57
I.3.2. Origines de la transcription envahissante 59
I.3.3. Fonctions de la transcription envahissante 62
I.3.3.1. Régulation de la transcription par des ARN non-codants 62
I.3.3.2. Les transcrits envahissants : une source de nouveaux gènes ? 66
I.3.4. Mécanismes de contrôle de la transcription envahissante 67
I.3.4.1. Contrôle de l'initiation de la transcription 67
I.3.4.2. Contrôle des transcrits envahissants par le couplage de la terminaison et de la dégradation des ARN 69
I.3.4.3. Contrôle post-transcriptionnel des transcrits envahissants 69
I.3.5. Un nouveau mécanisme de contrôle de la transcription envahissante 69
I.3.5.1. Le facteur de transcription Reb1 70
La terminaison par Reb1 71
Un mécanisme de terminaison de secours 72
I.3.5.2. Rap1, un autre facteur de terminaison ? 73

II. Travaux de thèse 78
II.1. Etude de séquences nécessaires à la terminaison par la voie NNS 78
II.1.1. Introduction 78
II.1.2. Résultats et discussion 80
II.1.2.1. Le système rapporteur 80
II.1.2.2. Identification de 5 motifs potentiellement impliqués dans la terminaison du CUT du locus *IMD2* 81
II.1.2.3. Limites de l'expérience et perspectives 83
II.2. Analyse des sites de fixation du dimère Nrd1-Nab3 sur l'ARN par SELEX 85
II.2.1. Introduction 85
II.2.2. Résultats 85
II.2.2.1. Le SELEX Nrd1-Nab3 85
II.2.2.2. Séquençage et analyse des séquences sélectionnées 87
II.2.2.3. Les sites de fixation de Nrd1 et Nab3 ne sont pas positionnés de façon aléatoire. 89

II.3. Rap1 induit la terminaison de la transcription par un mécanisme de « road-block » 92
II.3.2. Résultats 92
II.3.2.1. L'expérience de sélection de terminateurs à partir de séquences aléatoires 93
II.3.2.2. Le facteur de transcription Rap1 est impliqué dans la terminaison de la transcription 95
II.3.2.3. Le site de liaison de Rap1 est nécessaire et suffisant pour la terminaison de la transcription 96
II.3.2.4. La terminaison par Rap1 produit des ARN instables et polyadénylés par Trf4 97
II.3.2.5. Rap1 termine la transcription en se liant à l'ADN 98
II.3.2.6. Le site de liaison de Rap1 n'est pas transcrit dans les terminateurs dépendants de Rap1 99
II.3.2.7. Le mécanisme de désassemblage des ARN pol II bloquées par Rap1 101
II.3.2.8. Rap1 termine la transcription dans le génome de *S. cerevisiae* 102
II.3.2.9. Accumulation d'ARN pol II en amont des sites de fixation de Rap1 dans le génome 109
II.3.3. Discussion 113
II.3.3.1. Un mécanisme de terminaison de la transcription par road-block conservé 114
II.3.3.2. Le mécanisme de la terminaison de la transcription par road-block par Rap1 115
II.3.3.3. La fonction biologique de la terminaison par Rap1 118
II.3.3.4. D'autres road-blockers ? 121
II.3.4. Perspectives 124

III. Matériels et méthodes 126
III. 1. Souches de *S. cerevisiae* utilisées 126
III. 2. Vecteurs utilisés 126
III. 3. Techniques communes aux trois projets 126
III.4. Identification des séquences nécessaires à la terminaison du CUT du locus *IMD2* 128
III. 5. Identification du site de fixation optimal du dimère Nrd1-Nab3 sur l'ARN 129
III. 6. Terminaison road-block par Rap1 132

IV. Références bibliographiques 139

I. Introduction

La transcription coordonnée et simultanée de milliers de gènes et d'ARN non-codants au sein de la cellule nécessite de multiples systèmes de régulation et de contrôle. De nombreux mécanismes assurent un fonctionnement maîtrisé de la synthèse des transcrits à chacune des étapes d'initiation, d'élongation et de terminaison de la transcription. Ces mécanismes sont essentiels puisque des évènements transcriptionnels non maîtrisés peuvent altérer l'expression de gènes et menacer l'intégrité de la cellule.

Cette introduction présente les connaissances actuelles à propos des différentes étapes de la transcription par l'ARN polymérase (pol) II et décrit plus en détail ses deux voies de terminaison connues chez la levure *Saccharomyces cerevisiae*. J'exposerai ensuite mes travaux qui portent sur l'étude d'une de ces deux voies et sur la découverte d'un nouveau mode de terminaison. Ces résultats participent au déchiffrement des signaux de ponctuation du génome ainsi qu'à la caractérisation de nouveaux modes de contrôle de la transcription tout au long du génome.

I.1. La transcription chez les eucaryotes

Chez les eucaryotes la transcription nucléaire est assurée par trois ARN polymérases, les ARN pol I, II et III (figure I.1.). Elles sont structurellement et fonctionnellement très proches (Cramer et al., 2008). Une quatrième et cinquième ARN polymérase ont récemment été découvertes chez les plantes, ce sont des homologues de l'ARN pol II chargées de transcrire de petits ARN régulateurs.
Chacune de ces ARN polymérases produit des classes de transcrits qui leur sont spécifiques pour la plupart.

I.1.1. Diversité et fonctions des ARN

Les ARN sont des molécules fonctionnelles essentielles. Ils couvrent des fonctions extrêmement variées, telles que le transport de l'information génétique, le contrôle de l'expression génique ainsi que des rôles de composants structuraux et enzymatique dans une large variété de processus biologiques. Ces transcrits peuvent être classés en deux catégories, les ARN codants et non-codants.

I.1.1.1. Les ARN codants

Appelés ARN messagers (ARNm), ces transcrits produits par l'ARN pol II contiennent au moins une phase ouverte de lecture (ORF) codant une protéine et sont exportés vers le cytoplasme où ils sont traduits. Les ARNm sont tout d'abord transcrits en ARN pré-messagers dont la séquence est divisée en exon, conservés dans l'ARNm mature, et en introns, généralement supprimés lors de l'épissage à la suite de leur transcription. Les ARNm possèdent à leurs extrémités 5' et 3' des modifications qui sont propres à la transcription par l'ARN pol II.
A l'extrémité 5', une coiffe composée d'une 7-methylguanosine est ajoutée de manière co-transcriptionnelle via une liaison 5' vers 5'. Cette coiffe a un rôle important dans la maturation, l'export et la traduction des ARNm. L'extrémité 3' quant à elle, possède une répétition d'environ 70 adénosines chez *S. cerevisiae* dont la synthèse est couplée à la terminaison. Cette queue poly(A) n'est pas transcrite par l'ARN pol II mais synthétisée par des facteurs spécifiques. Ces facteurs sont décrits dans la section portant sur la polyadénlation et terminaison.

I.1.1.2. Les ARN non-codants

Cette catégorie contient une très grande variété de transcrits parmi lesquels se trouvent des ARN régulateurs, structuraux, catalytiques et les produits de la transcription envahissante, un processus décrit ultérieurement.

Egalement synthétisés par l'ARN pol II, les ARN régulateurs ont pour fonction de réguler l'expression d'un gène ciblé. Ils peuvent intervenir à tous les stades de la transcription, dans le noyau comme dans le cytoplasme, de manière co- ou post-transcriptionelle (revue : Goodrich and Kugel, 2006). Parmi ces ARN, des petits transcrits de 20 à 30 nucléotides appelés petits ARN interférents (siARN), micros ARN (miARN) et piwi interacting ARN (piARN) sont impliqués dans les mécanismes d'interférence par ARN et provoquent la dégradation de leurs ARNm cibles ou réduisent la transcription de régions précises du génome par la modification locale de la structure chromatinienne. Etonnamment, l'interférence par ARN n'est pas conservée chez la levure *S. cerevisiae* alors qu'elle est retrouvée chez d'autres levures proches dont *Saccharomyces castellii* (Drinnenberg et al., 2009, 2011). D'autres ARN régulateurs plus longs, nommés long non-coding ARN (lncARN) ont

été découverts plus récemment. Cette catégorie de transcrits regroupe une multitude de transcrits dont les deux caractéristiques communes sont d'être des transcrits non-codants et d'une taille supérieure à 200 nucléotides. Leurs modes d'action étant très variés ils ne seront pas décrits ici (revue : Ulitsky and Bartel, 2013).

Il existe plusieurs familles d'ARN structuraux et catalytiques dont les principales sont les ARN ribosomaux (ARNr), les ARN de transfert (ARNt), les petits ARN nucléaires (snARN) et les petits ARN nucléolaires (snoARN). Ces ARN sont des composants essentiels de complexes ribonucléoproteiques impliqués dans la maturation et la traduction des transcrits.
Les ARNr forment la base de la structure et du fonctionnement des ribosomes. Chez les eucaryotes il en existe quatre nommés d'après leurs coefficients de sédimentation en gradient de sucrose. Les ARNr 28S (ou 25S chez *S. cereviae*), 18S et 5,8S sont synthétisés par l'ARN pol I et l'ARNr 5S est lui produit par l'ARN pol III.
L'ARN pol III synthétise également les ARNt qui sont chargés de transporter les acides aminés jusqu'au centre catalytique du ribosome lors de la traduction.
Bien que l'ARN pol III transcrive certains snARN, c'est l'ARN pol II qui transcrit la majorité des snARN et snoARN, des ARN impliqués dans la modification post-transcriptionnelle de certains transcrits. Les snARN sont recrutés au sein de la machinerie d'épissage (revue : Matera and Wang, 2014) et les snoARN guident les modifications de nucléotides sur les ARNr et les snARN (revue : Bratkovič and Rogelj, 2014).

La transcription efficace et régulée de tous ces ARN requiert un fort contrôle de la circulation des ARN polymérases tout au long du génome. Cela est d'autant plus important pour un génome compact et fortement transcrit tel que celui de *S. cerevisiae*, comptant environ 6000 gènes codant des protéines et de très nombreuses autres unités de transcription non-codantes réparties de manière très dense sur les 16 chromosomes. Les gènes ne sont souvent séparés que par quelques centaines de paires de bases et on estime que 85% du génome est transcrit en milieu riche (David et al., 2006).

Le recrutement des ARN polymérases sur l'ADN est étroitement contrôlé. La structure de la chromatine empêche les polymérases d'accéder librement à l'ADN,

leur recrutement s'effectue au niveau de régions d'ADN décompacté. La suite de cette introduction est consacrée à la description des différentes étapes de la transcription par l'ARN pol II chez *S. cerevisiae* et leurs mécanismes de contrôle associés.

I.2. La transcription par l'ARN pol II chez *S. cerevisiae*

I.2.1. L'ARN pol II

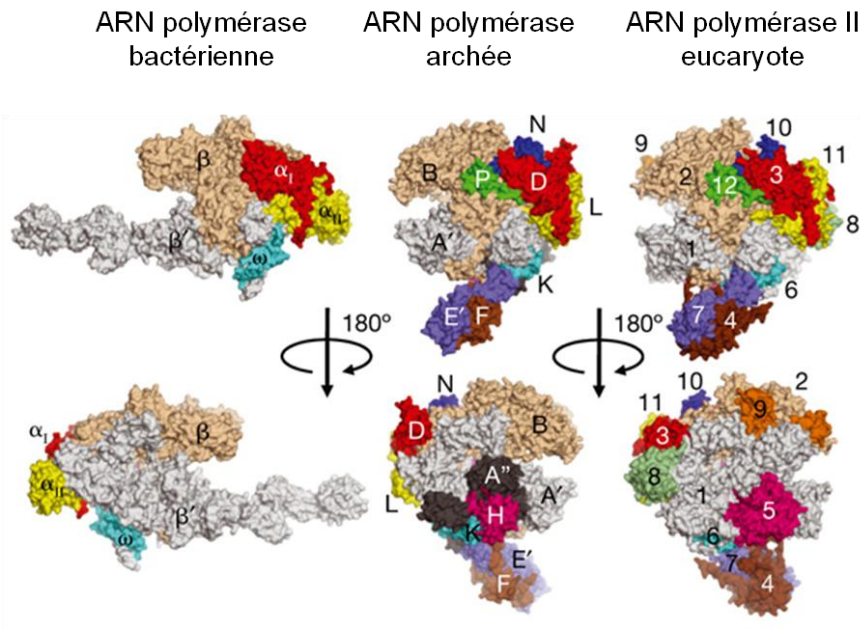

Figure 1.1. Structures d'ARN polymérases bactérienne, archée et eucaryote.
Représentation de la surface des ARN polymérases bactérienne (à gauche, *Thermus aquaticus*), Archée (au centre, *Sulfolobus solfataricus*) et eucaryote (à droite, ARN pol II de *S. cerevisiae,* sans le domaine C-terminal de Rpb1p). Les sous-unités sont colorées et nommées, les sous-unités orthologues sont représentées de la même couleur. Figure adaptée de Hirata et al., 2008.

Présentes dans les trois domaines du vivant, les ARN polymérases sont des complexes multi-protéiques dont l'activité enzymatique permet la formation de liaisons phosphodiester entre ribonucléotides menant à la formation de molécules d'ARN linéaires. De manière très marquante les ARN polymérases provenant des eucaryotes, des procaryotes et des archées sont structurellement très proches (figure 1.1). Bien que le nombre de sous-unités varie de l'ARN polymérase bactérienne (un cœur formé de 5 sous-unités) à l'ARN pol II eucaryote (12 sous-unités) on observe la même structure en « pince de crabe » parmi tous ces complexes (Hirata et al., 2008).

Chez les eucaryotes, les ARN pol I, II et III sont constituées respectivement de 14, 12 et 17 sous-unités, 10 d'entre elles composent leurs cœurs catalytiques. Les sous-unités de l'ARN pol II sont nommées de Rpb1 à 12 dans l'ordre décroissant de la longueur de ces 12 polypeptides chez *S. cerevisiae*. Rpb1 possède un long domaine carboxy-terminal (CTD) essentiel et non structuré (revue : Corden, 2013). Ce domaine spécifique de l'ARN pol II est constitué de répétition en tandem de 7 acides aminés Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7 dont le nombre de répétitions est variable selon les organismes, de 26 chez *S. cerevisiae* à 52 chez l'Homme. Le CTD s'étend à partir du cœur de l'ARN pol II formant une « queue » qui permet l'interaction avec de nombreux facteurs impliqués dans les étapes de l'initiation, de l'élongation et de la terminaison de la transcription. Le CTD participe également au recrutement de facteurs de maturation des ARN naissants. Les acides aminés du CTD sont sujets à de très nombreuses modifications post-traductionnelles telles que des glycosylations, ubiquitinations, méthylations et les plus étudiées sont les phosphorylations (revue : Hsin and Manley, 2012). Comme décrit dans la figure 1.2. le profil de phosphorylation des sérines du CTD varie au cours du cycle de transcription ce qui permet à l'ARN pol II d'interagir spécifiquement avec les facteurs requis pour chacune des étapes de la transcription. Le CTD est hypophosphorylé lors de la formation du complexe de préinitiation (PIC), les sérines 5 et 7 sont ensuite phosphorylées lors des étapes précoces de la transcription, ces phosphorylations sont ensuite perdues au profit de celle de la sérine 2 au cours de l'élongation.

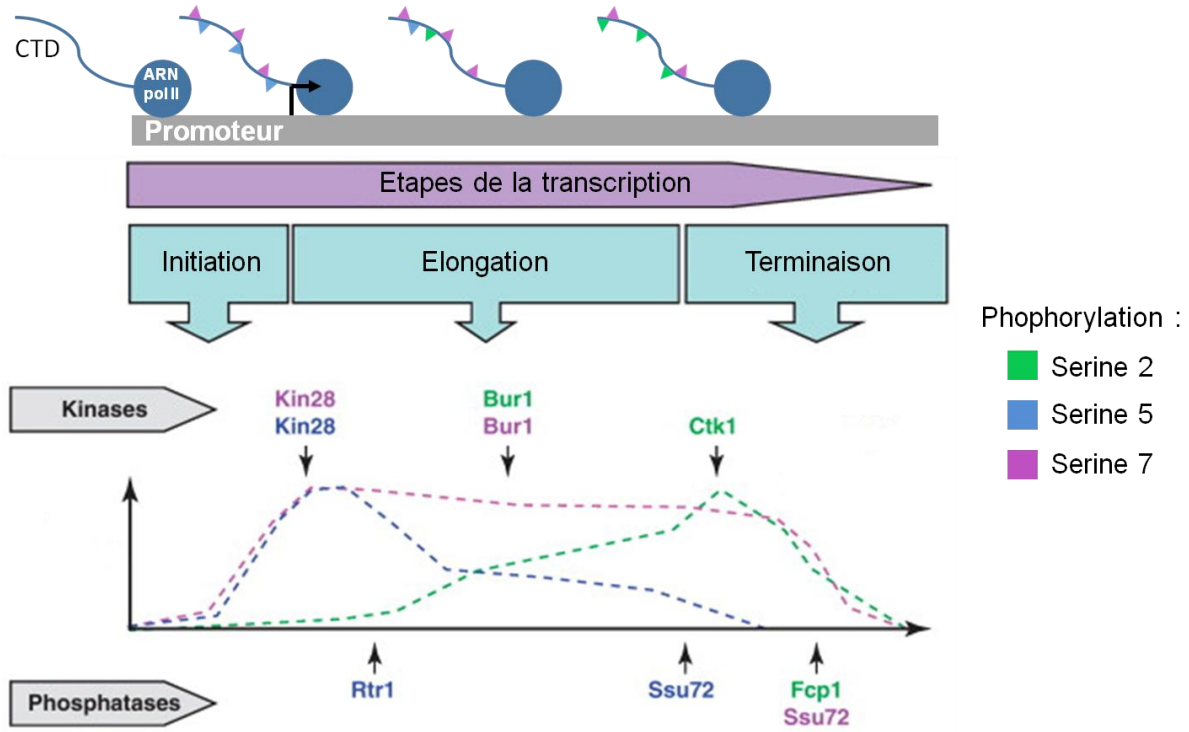

Figure 1.2. Etat de phosphorylation du CTD de l'ARN pol II au cours de la transcription d'un gène chez *S. cerevisiae*.
Le niveau moyen de phosphorylation des sérines 2, 5 et 7 du domaine C-terminal de Rpb1p par rapport à la quantité totale d'ARN pol II est représenté au cours des différentes étapes de la transcription. Les noms des kinases et des phosphatases responsables de ce profil de phosphorylation sont notés. Figure adaptée de Egloff et al., 2012.

I.2.2. L'initiation de la transcription

La transcription par l'ARN pol II est contrôlée par des facteurs généraux de transcription (GTF) ainsi que par une multitude d'activateurs et de répresseurs. L'ARN pol II et les GTF forment le complexe de pré-initiation (PIC) et se lient aux promoteurs en amont des sites d'initiation de la transcription et participent à la modulation de la fréquence d'initiation, communément appelée la « force » de la transcription.

I.2.2.1. Reconnaissance du promoteur

L'ARN pol II est recrutée sur l'ADN au niveau des promoteurs et initie la transcription à travers un processus qui implique les GTF : TFIIA, TFIIB, TFIID, TFIIE, TFIIF et TFIIH (TFII signifie Facteur de Transcription de l'ARN pol II, figure 1.3.).

Le premier élément du promoteur à avoir été décrit chez les eucaryotes est le site de fixation de la protéine de liaison à la boîte TATA (TBP), une sous-unité de TFIID. Cette séquence fortement conservée est située entre 25 et 35 paires de base (pb) en amont du site d'initiation de la transcription (TSS) chez la plupart des eucaryotes. Chez *S. cerevisiae* le consensus correspondant à la boite TATA, TATAWAWR (W correspond à A/T et R à A/G), est plus éloigné du TSS, à une distance de 40 à 120pb. Ce consensus tout d'abord pensé comme étant un élément universel n'est en fait présent que dans environ 20% des promoteurs de la levure (revue : Dikstein, 2011). Cependant une étude récente montre que la vaste majorité des promoteurs de levure contiennent tout de même des séquences proches de la boîte TATA. Celles-ci diffèrent du consensus habituel par 2 bases ou moins et TBP est recrutée au niveau de ces séquences (Rhee and Pugh, 2012) bien qu'avec une moindre affinité. TBP est également requise pour la transcription par les ARN pol I et III (Cormack and Struhl, 1992).
Un deuxième motif appelé initiateur est fréquemment présent dans les promoteurs eucaryotes. Chez *S. cerevisiae* sa séquence consensus riche en adénosine est A(A-rich)$_5$NY<u>A</u>(A/T)NN(A-rich)$_6$, le A souligné correspond au site d'initiation de la transcription (Zhang and Dietrich, 2005).

Chez les métazoaires d'autres éléments au cœur du promoteur (Downstream Promoter Element, B Recognition Element) ont été décrits mais ils ne sont pas retrouvés chez la levure.

I.2.2.2. Formation du PIC

Les facteurs généraux de transcription et l'ARN pol II
Le modèle conventionnel de recrutement de l'ARN pol II au travers de la formation du PIC suit une série d'événements interdépendants (Sikorski and Buratowski, 2009). Les régions promotrices dont la structure chromatinienne est décondensée sont reconnues par TBP associée à TFIID, un complexe d'environ 12 sous-unités dont certaines sont impliquées dans la sélection et la liaison aux promoteurs. TFIIA stabilise ensuite la liaison TBP-ADN (Buratowski et al., 1989), interagit avec de nombreux activateurs et déplace ou bloque des inhibiteurs de la transcription. TFIID possède une surface d'interaction avec TFIIB qui est impliqué dans la sélection du site d'initiation de la transcription et stabilise lui aussi la fixation de TBP sur l'ADN.

L'ARN pol II associée à TFIIF se lie ensuite au PIC en cours de formation. Le recrutement de TFIIE suit celui de l'ARN pol II, il semble impliqué dans l'ouverture de l'ADN double brin. Vient enfin TFIIH dont les sous-unités portent au moins trois activités enzymatiques, ATPase ADN-dépendante, hélicase ATP-dépendante et CTD kinase. L'activité hélicase désapparie l'ADN et permet le démarrage de la transcription tandis que l'activité kinase, portée par Kin28 (Cdk7 chez l'humain) phosphoryle les sérine 5 et 7 du CTD de l'ARN pol II (Lee and Young, 2000).

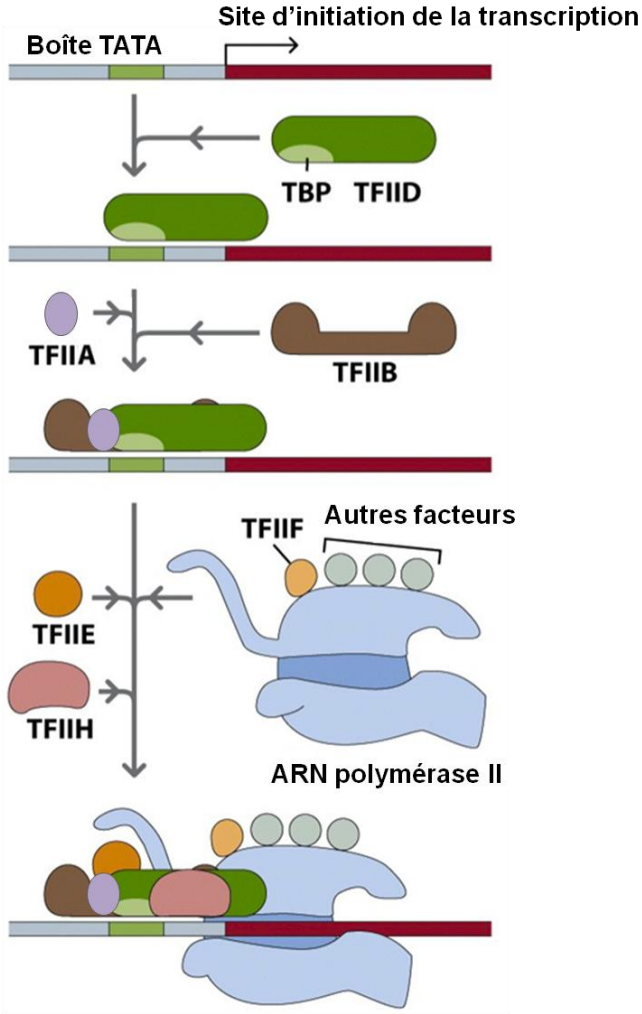

Figure 1.3. Formation du complexe de pré-initiation.
La boîte TATA est reconnue par une sous-unité de TFIID, TBP. TFIIA et TFIIB se fixent ensuite sur TFIID et permettent le recrutement de l'ARN pol II accompagnée de nombreux facteurs dont TFIIE, TFIIH et TFIIF.
Figure adaptée de Human Molecular Genetics, quatrième édition par Strachan et Read, 2010.

Chez la levure, dans 10% des PIC, TBP est associée au complexe SAGA (pour Spt-Ada-Gcn5-Acetyl transferase) plutôt qu'à TFIID (Huisinga and Pugh, 2004). Les deux

complexes SAGA et TFIID partagent de nombreuses sous-unités et possèdent à la fois une affinité pour TBP et une activité histone acétyltransférase ce qui souligne un lien entre initiation de la transcription et modification de la chromatine (Grant et al., 1997; Mizzen et al., 1996). Les gènes contenant une boîte TATA conforme au consensus correspondent à des gènes de réponse aux stress hautement régulés et sont préférentiellement associés au complexe SAGA plutôt qu'à TFIID (Basehoar et al., 2004).

Le Médiateur

Un autre complexe appelé Médiateur favorise l'initiation de nombreux transcrits. C'est une plateforme co-activatrice permettant de transmettre les signaux d'activation de l'expression des gènes à l'ARN pol II qui est également fréquemment associée au PIC. Ce complexe multi-protéique est notamment ciblé par des facteurs de transcription impliqués dans le programme d'expression de gènes en réponse à des signaux développementaux et environnementaux (revue : Poss et al., 2013).

Les facteurs de transcription

Pour réguler spécifiquement l'expression des gènes la cellule emploie des facteurs de transcription. A l'inverse des facteurs « généraux » de transcription, les facteurs de transcription ne ciblent qu'une quantité limitée de gènes. Ils se lient sur l'ADN en reconnaissant des séquences régulatrices spécifiques parfois éloignées de plusieurs centaines de paires de bases du promoteur. Ces facteurs peuvent activer ou réprimer la transcription de leurs gènes cibles.

Les facteurs de transcription activateurs de la transcription sont composés d'au moins deux domaines : un domaine de fixation à l'ADN et un domaine activateur. Les domaines activateurs utilisent de nombreux modes d'action pour favoriser la transcription. Ils peuvent notamment attirer, positionner ou modifier les facteurs généraux de transcription, le Médiateur ou l'ARN pol II ; soit par interaction directe avec ces éléments, soit indirectement, en modifiant la structure de la chromatine au niveau du promoteur.

Les facteurs de transcription répresseurs de la transcription sont capables de limiter la transcription par des méthodes variées telles que la compétition pour la fixation sur une séquence régulatrice reconnue par un activateur, le masquage de la surface

d'activation d'un activateur, l'interaction directe avec des facteurs généraux de transcription ou du remodelage de la structure chromatinienne du promoteur.

I.2.2.3. Structure chromatinienne du promoteur

Les génomes eucaryotes sont compactés sous forme de chromatine dont l'unité de base est le nucléosome. Les nucléosomes sont constitués de 147pb d'ADN enroulées autour d'un complexe protéique composé de deux copies de chaque histone H2A, H2B, H3 et H4. Les nucléosomes sont placés de manière ordonnée sur l'ADN, environ tous les 200 nucléotides, et ils sont reliés les uns aux autres par les courtes séquences d'ADN situées entre eux ainsi que par l'histone H1 qui assure un rôle de connecteur (revue : Jansen and Verstrepen, 2011). Comparé à l'ADN nu, la chromatine est moins accessible aux facteurs régulateurs et la structure de la chromatine influence de nombreux processus cellulaires tels que la transcription, la réplication, la réparation de l'ADN, la recombinaison ou la ségrégation des chromosomes. Cette compaction protège également l'ADN contre les dommages et la formation de structures aberrantes.

La cartographie du positionnement des nucléosomes sur l'ensemble du génome de *S. cerevisiae* a révélé que les nucléosomes sont très précisément placés et, dans des conditions physiologiques similaires, ils sont retrouvés à la même position dans quasiment toutes les cellules (Lee et al., 2007; Yuan et al., 2005). La plupart des gènes de levure (95%) sont caractérisés par une déplétion de nucléosomes aux extrémités 5' et 3' des gènes (Jiang and Pugh, 2009; Kaplan et al., 2010), ces régions sont appelées NDR (nucleosome depleted regions) ou NFR (nucleosome free regions). La NDR en 5' s'étend sur environ 150pb au niveau des régions promotrices et la transcription démarre généralement au niveau des 10 premières paires de base de l'ADN associé au nucléosome +1 (Albert et al., 2007) (figure 1.4.).

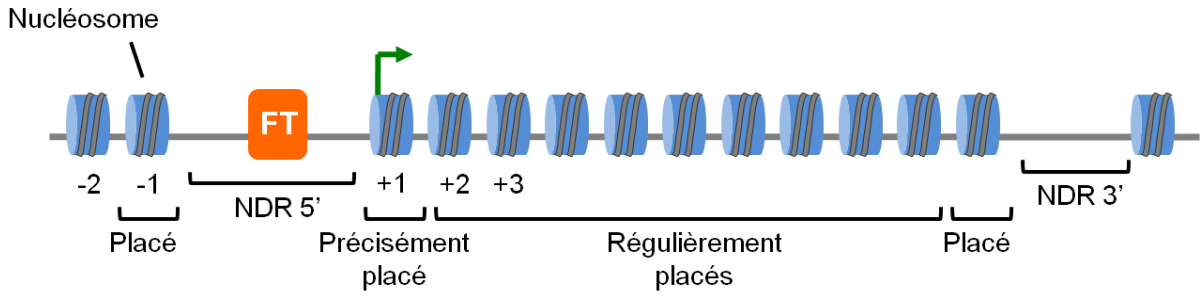

Figure 1.4. Organisation typique des nucléosomes sur un gène de levure.
En 5', une région déplétée en nucléosome (NDR) d'environ 150 pb est délimitée par deux nucléosomes, -1 et +1, précisément placés. La NDR 5' contient le promoteur et de nombreuses séquences régulatrices dont des sites de fixation de facteur de transcription (FT). Le nucléosome +1 est situé au niveau du site d'initiation de la transcription (flèche verte) et le positionnement des nucléosomes suivants est de moins en moins précis en s'éloignant du promoteur. A l'extrémité 3', un nucléosome est précisément placé en amont de la NDR 3'. Figure adaptée de Jansen and Verstrepen, 2011.

Le positionnement des nucléosomes au niveau des promoteurs est déterminé à la fois par la séquence de l'ADN elle même et par le recrutement de facteurs capables de modifier l'organisation des nucléosomes.

Il a été montré que les séquences d'ADN qui présentent un fort taux de A et T sont corrélées avec une faible présence de nucléosomes (Tillo and Hughes, 2009). Les répétitions poly(dA:dT), retrouvées dans les NDR « repoussent » les nucléosomes (revue : Segal and Widom, 2009). Il a été proposé que ces répétitions forment des structures rigides qui ne sont pas adaptées à l'enroulement autour des octamères d'histones (Iyer and Struhl, 1995).

Le positionnement des nucléosomes est aussi influencé par des complexes de remodelage de la chromatine qui utilisent l'énergie de l'hydrolyse de l'ATP pour déplacer, éjecter ou restructurer les nucléosomes (revue : Clapier and Cairns, 2009). Il existe quatre grandes familles de complexes de remodelage de la chromatine, conservées de la levure à l'homme : les familles SWI/SNF (switching defective/sucrose nonfermenting), ISWI (imitation switch), CHD (chromodomain, helicase, DNA binding) et INO80 (inositol requiring 80). Les sous-unités catalytiques de ces quatre familles partagent un domaine ATPase conservé et des domaines flanquants spécifiques. Un certain nombre d'études montrent que ces complexes jouent un rôle majeur dans la structure des promoteurs. Il a notamment été montré que le complexe RSC de la famille SWI/SNF participe à l'exclusion des nucléosomes

sur des centaines de promoteurs (Badis et al., 2008; Hartley and Madhani, 2009). Par ailleurs la délétion de ISW2 membre de la famille ISWI provoque une perturbation de la structure dans plus de 1000 régions génomiques dont environ 12% des promoteurs (Whitehouse and Tsukiyama, 2006; Whitehouse et al., 2007).

Le positionnement des nucléosomes est aussi influencé par des facteurs de transcription. Par exemple, les facteurs Reb1, Rap1 et Abf1 promeuvent la formation des NDR autour de leurs sites de liaison à l'ADN, à la fois par leurs propriétés intrinsèques capables de déstabiliser les nucléosomes (Yarragudi et al., 2004) et par le recrutement de complexes de remodelage de la chromatine (Gavin et al., 2002; Hartley and Madhani, 2009; Idrissi et al., 2001; Reid et al., 2000). Leur déplétion provoque une réduction des NDR au niveau de leur gènes cibles (Fourel et al., 2002; Hartley and Madhani, 2009; Yu et al., 2003).
D'autre part, les sites de liaison de certains facteurs de transcription peuvent être masqués par les nucléosomes, ce qui limite l'initiation de la transcription (Jansen and Verstrepen, 2011). Conformément à ce modèle, de nombreux sites fixés par des facteurs de transcription se trouvent dans des NDR où leur accès à l'ADN est favorisé (Goh et al., 2010).

Ces différents mécanismes d'exclusion des nucléosomes coopèrent puisque les motifs poly(dA:dT) enrichis dans les NDR peuvent faciliter l'accès des facteurs de transcription à l'ADN qui, pour certains, recrutent des complexes de remodelage de la chromatine participant eux aussi à l'exclusion des nucléosomes. Cela permet de créer des régions d'ADN décompacté compatibles à l'assemblage d'un PIC et à l'initiation de la transcription.

I.2.3. L'élongation de la transcription

La formation du PIC est suivie par le départ de l'ARN pol II du promoteur et par une phase d'élongation processive.

I.2.3.1. Départ du promoteur

Une fois que l'ARN pol II a accès au promoteur elle doit faire face à de multiples challenges avant d'atteindre un stade d'élongation processive. Les GTF associés au

PIC encombrent l'ARN pol II, une portion de TFIIB occupe le tunnel de sortie de l'ARN et TFIIE et TFIIF lient la polymérase de part et d'autre de la région dans laquelle doit se positionner le brin d'ADN matrice. Ces interactions doivent être éliminées lors du dégagement du promoteur. Par ailleurs, l'ARN pol II ne peut pas entamer la transcription tant que le double brin d'ADN situé au niveau du promoteur n'est pas ouvert par les activités hélicases de Ssl2 et Rad3 (Xpb et Xpd chez l'Homme), deux sous-unités de TFIIH.

La façon dont le site d'initiation de la transcription (TSS) est choisi n'est pas connue. Chez la levure, le TSS peut être situé de 40 à 120 nucléotides de la boîte TATA, le modèle consensuel propose que l'ARN pol II scanne l'ADN à partir du promoteur jusqu'à identifier le TSS à l'aide de TFIIB et TFIIH (Yang and Ponticelli, 2012; Khaperskyy et al., 2008). Une fois la transcription initiée, l'hybride ARN-ADN s'allonge tout d'abord de 6 à 9 nucléotides, puis le transcrit nouvellement synthétisé entre dans le tunnel de sortie de l'ARN pol II et sort de la polymérase après avoir atteint une taille d'environ 18 nucléotides (revue : Luse, 2013). Des études effectuées sur l'ARN pol II humaine montre que la bulle de dénaturation s'agrandit dans un premier temps jusqu'à 17-18 bases puis se referme pour ne laisser qu'une bulle de dénaturation d'environ 10 bases non appariées typique du complexe d'élongation (Holstege et al., 1997; Pal et al., 2005). La réduction de la taille de la bulle de dénaturation est considérée comme la fin du dégagement du promoteur. L'initiation de la transcription est marquée par de nombreux événements de transcription avortés, d'avancés et de reculs de la polymérase mais une fois que le transcrit a dépassé environ 30 nucléotides, l'ARN pol II a perdu ses contacts avec les facteurs d'initiation et entame une élongation processive (revue : Luse, 2013).

Le départ de l'ARN pol II est aussi favorisé par la phosphorylation du CTD de la polymérase au niveau du promoteur, cette modification facilite la dissociation de l'ARN pol II du PIC. Dans le PIC, le Médiateur interagit avec l'ARN pol II non phosphorylée et son intégration au PIC stimule fortement l'activité de Kin28, la CTD kinase de TFIIH. Kin28 phosphoryle alors le CTD et cette phosphorylation perturbe la fixation du Médiateur (Max et al., 2007) et participe à la libération de l'ARN pol II tandis que le Médiateur reste fixé au promoteur.

I.2.3.2. Elongation de la transcription à travers la chromatine

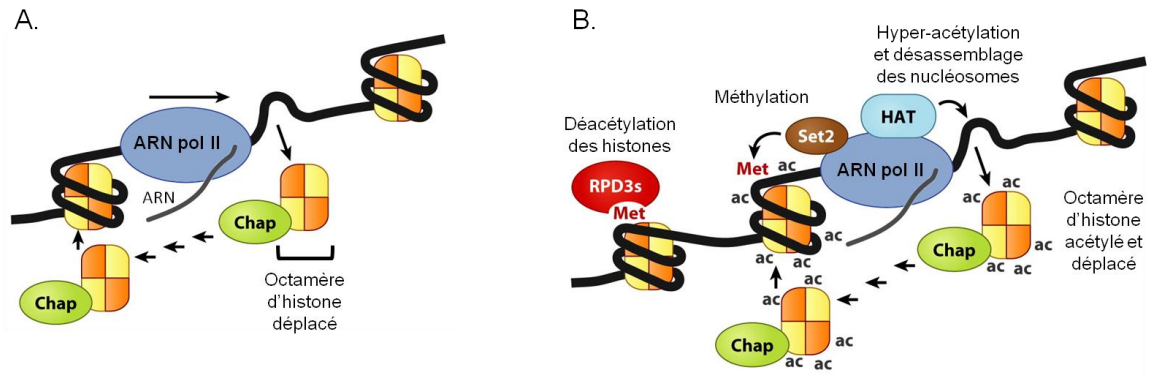

Figure 1.5. Elongation de la transcription à travers les nucléosomes.

A. La transcription à travers les nucléosomes peut être favorisée par l'action de chaperonnes (Chap) d'histone qui déplacent les histones au passage de l'ARN pol II.
B. Les modifications co-transcriptionnelles des histones interviennent dans la transcription à travers les nucléosomes. Des histones acétyltransférases (HAT) acétylent les histones au passage de l'ARN pol II, ce qui stimule le désassemblage des nucléosomes et l'action des chaperonnes d'histone. Après le replacement des nucléosomes, les histones hyper-acétylés sont méthylés par Set2. Cette méthylation est spécifiquement reconnue par le complexe Rpd3S qui déacétyle les histones.

Abréviations : ac, acétylation, Met, méthylation.
Figure adaptée de Selth et al., 2010/

La machinerie de transcription doit progresser à travers les gènes dont l'ADN est organisé sous forme de chromatine dont l'unité de base est le nucléosome. Les nucléosomes sont des obstacles à l'avancée de l'ARN pol II, ils doivent donc être retirés pour permettre le passage de la polymérase et être replacés une fois le complexe d'élongation passé afin de maintenir la structure compacte et protectrice de l'ADN. Il existe donc une série de facteurs permettant le déplacement dynamique des nucléosomes au cours de l'élongation de la transcription (revue : Selth et al., 2010, figure 1.5.).

Les histones formant les nucléosomes peuvent porter des modifications telles que des acétylations, méthylations, ubiquitination. Ces modifications ont notamment un rôle important dans l'élongation grâce à deux mécanismes : d'une part ces modifications peuvent altérer les contacts entre nucléosomes et décompacter la chromatine, d'autre part elles forment des surfaces de contact avec des facteurs de remodelage et de modification (Selth et al., 2010). De nombreuses enzymes de modification d'histone et de remodelage de la chromatine sont impliquées dans le déplacement des nucléosomes au cours de l'élongation. Les facteurs et les

modifications participant à la transcription au travers de la chromatine sont succinctement décrits dans la section suivante.

Facteurs de remodelage de la chromatine recrutés lors de l'élongation

Des résultats obtenus à la fois *in vitro* et *in vivo* chez la levure montrent qu'il existe des complexes de remodelage de la chromatine capables de décompacter l'ADN des régions transcrites et de favoriser ainsi l'élongation. Un complexe membre de la famille SWI/SNF, le complexe RSC, remodèle la chromatine au niveau de séquences codantes et stimule l'élongation de la transcription (Carey et al., 2006; Mas et al., 2009).

De plus, le complexe FACT (Facilitating chromatin transcription), une chaperonne d'histone, favorise l'élongation de la transcription à travers la chromatine. Il a été montré que ce complexe, formé des protéines Spt16 et Pob3, déstabilise la structure du nucléosome menant au détachement d'un dimère d'histone H2A-H2B au passage de l'ARN pol II, facilitant ainsi l'élongation (Belotserkovskaya et al., 2003; Reinberg and Sims, 2006).

Une fois le complexe d'élongation passé, les nucléosomes sont reformés. Les octamères d'histones acétylés détachés au passage de l'ARN pol II sont pris en charge par les facteurs de remodelages Chd1 et Isw1 et replacés sur l'ADN une fois la machinerie de transcription passée (figure 1.5.B.). Il est important de replacer ces octamères d'histones acétylés puisqu'ils sont ensuite ciblés par les facteurs de compaction des nucléosomes Set2 et Rpd3S, décrits plus bas, permettant la recompaction de l'ADN (Simic et al., 2003; Smolle et al., 2012).

La maintenance de la structure de la chromatine et la reformation des nucléosomes après le passage de l'ARN pol II implique également la chaperonne Spt6 chez la levure. Son inactivation provoque une perte de nucléosomes sur les régions transcrites, ce qui rend accessible de nouveaux sites d'initiation aberrants à l'intérieur de ces régions transcrites (Kaplan et al., 2003). Cela montre que la chromatine n'est pas seulement un obstacle à la progression de l'ARN pol II mais qu'elle participe aussi au contrôle de l'initiation de la transcription.

Modification des histones lors de l'élongation de la transcription

L'élimination ou le déplacement des nucléosomes sont associés à la modification des histones au cours de l'élongation. Comme cela a été précisé précédemment, toute une série d'enzymes catalysent des modifications covalentes des histones telles que des acétylations, méthylations et ubiquitinations permettant une déstabilisation des nucléosomes ou créant de nouvelles surfaces d'interaction avec des facteurs d'élongation. Ces modifications jouent un rôle majeur dans la régulation de l'expression des gènes.

L'acétylation des histones

L'acétylation des histones est catalysée par des enzymes appelées histones acétyltransférases (HAT), elles permettent l'ajout d'un groupement acétyl sur des lysines dans la région N-terminale des histones. Cette modification neutralise la charge positive des histones et diminue leur affinité pour l'ADN (Hong et al., 1993) ce qui altère la conformation des nucléosomes (Norton et al., 1989) et favorise l'accès des facteurs de transcription aux promoteurs (Lee et al., 1993) et la progression de l'ARN pol II (Selth et al., 2010). Les régions promotrices actives sont caractérisées par une hyper acétylation des histones (Pokholok et al., 2005) tandis que les régions activement transcrites arborent un mécanisme d'acétylation-déacétylation au passage de la polymérase ce qui permet d'assurer la décompaction-recompaction des séquences transcrites. Les histones des régions transcriptionnellement inactives sont, quant à elles, hypoacétylées.
L'acétylation des histones est un phénomène réversible, elle est éliminée par l'action de nombreuses histones déacétylases (HDAC).

La méthylation des histones

Contrairement aux acétylations, les méthylations n'altèrent pas directement la structure des nucléosomes mais entraînent plutôt le recrutement de nouveaux facteurs. Les histones peuvent porter plusieurs groupements méthyl, ils sont principalement ajoutés sur des lysines et arginines spécifiques des histones H3 et H4. Ces modifications sont ajoutées par des histones méthyltransférases.

Le CTD de l'ARN pol II en cours d'élongation porte des sérines 2 phosphorylées, ce qui est spécifiquement reconnu par l'histone méthyltransférase Set2. Elle catalyse la méthylation de la lysine 36 l'histone H3 provoquant une accumulation de cette marque sur les régions fortement transcrites (Krogan et al., 2003; Li et al., 2003; Schaft et al., 2003; Xiao et al., 2003). Ces méthylations sont à leur tour reconnues par le complexe HDAC Rpd3S (Carrozza et al., 2005; Keogh et al., 2005). Par ce mécanisme, la méthylation cotranscriptionnelle des histones mène à leur déacétylation et permet la compaction des nucléosomes après le passage de l'ARN pol II.

Dans la région 5' des unités de transcription, la phosphorylation de la sérine 5 du CTD par le GTF TFIIH est reconnue par le complexe COMPASS (complex proteins associated with Set1), un complexe capable d'ajouter jusqu'à 3 groupements méthyl sur les lysines 4 des histones H3. De manière intéressante, le profil de méthylation par le complexe COMPASS décroit du promoteur jusque dans la région transcrite. Les histones H3 proches du TSS portent une tri-méthylation, la région juste en aval porte, elle, une di-méthylation tandis que le reste du gène possède une mono-méthylation (Liu et al., 2005; Pokholok et al., 2005). La tri-méthylation des histones semble favoriser l'acétylation des nucléosomes et leur exclusion des régions promotrices tandis que la di-méthylation entraîne le recrutement du complexe Set3 qui déacétyle les histones dans la région 5' des gènes et participe ainsi au blocage de la décompaction de la chromatine provenant du promoteur (Kim and Buratowski, 2009).

Compte tenu du temps de demi-vie de la méthylation des histones, qui est comparable au temps de demi-vie des histones elles-même, il a longtemps été considéré que cette modification était irréversible (Shi et al., 2004). Le seul mécanisme connu de suppression des histones méthylées était leur remplacement. Depuis 2004 et la découverte de la première histone déméthylase, LSD1 (lysine-specific histone demethylase 1) (Shi et al., 2004), plus d'une vingtaine d'histone déméthylases ont été identifiées (Shi and Tsukada, 2013).

L'ubiquitination des histones

L'ajout d'ubiquitine, un polypeptide de 76 acides aminés, aux protéines mène à diverses conséquences. Celle-ci est ajoutée sous forme d'un monomère sur lequel peut être synthétisée une chaîne linéaire d'ubiquitine. Il existe différentes formes de chaînes, certaines d'entre elles mènent à la dégradation des protéines par le protéasome 26S alors que d'autres formes de chaînes, ainsi que la mono-ubiquitination, régulent l'activité des protéines (revues : Heride et al., 2014; Sadowski et al., 2012).

Dans le cas des histones, la mono-ubiquitination de H2B favoriserait l'élongation à travers la chromatine. L'ubiquination des histones est organisée en trois étapes impliquant successivement une enzyme E1 activatrice de l'ubiquitine, une enzyme E2 de conjugaison de l'ubiquitine et enfin une enzyme E3, l'ubiquitine ligase. Chez *S. cerevisiae*, les histones H2B sont mono-ubiquitinées par une série d'enzymes E3 (Wright et al., 2012). Elles sont notamment mono-ubiquitinées par l'action combinées des enzymes Rad6 (E2) et Bre1 (E3) (Robzyk et al., 2000; Wood et al., 2003).
Cette marque favoriserait, par un mécanisme encore mal compris, la méthylation des lysine 4 et 79 de l'histone H3, une modification favorisant l'élongation décrite dans la section précédente (Briggs et al., 2002; Ng et al., 2002; Sun and Allis, 2002).
De plus, l'ubiquitination de l'histone H2B favorise le passage de l'ARN pol II à travers les nucléosomes *in vitro* en stimulant la déstabilisation des nucléosomes par l'histone chaperonne FACT (Pavri et al., 2006). Cette observation est soutenue par une étude *in vivo* chez la levure montrant que la mutation du site d'ubiquitination de H2B provoque une modification de la distribution des nucléosomes et de l'ARN pol II sur les gènes (Tanny et al., 2007; Shukla and Bhaumik, 2007).

L'ubiquitination des histones est réversible, la mono-ubiquitination de l'histone H2B peut être supprimée par l'action de deux enzymes distinctes, Ubp8 et Ubp10 (Daniel et al., 2004; Emre et al., 2005; Henry et al., 2003).

I.2.3.3. Contrôle de l'arrêt transcriptionnel

Les mécanismes favorisant la transcription à travers la chromatine décrits précédemment ne sont pas les seuls à aider l'ARN pol II à progresser le long de l'ADN. En effet la polymérase peut s'arrêter temporairement et parfois reculer ce qui mène à un état appelé « pause transcriptionnelle ». Si la pause de l'ARN pol II persiste, elle évolue progressivement vers le statut « d'arrêt transcriptionnel » (Gu and Reines, 1995a). Soit l'état de pause se résout de lui-même, l'ARN pol II étant capable de repartir, soit la polymérase est incapable de reprendre l'élongation et passe à un état d'« arrêt » transcriptionnel. Une ARN pol II arrêtée nécessite l'intervention de facteurs additionnels pour reprendre une élongation processive ou pour être éliminée (Kwak and Lis, 2013). L'arrêt de l'ARN pol II peut être provoqué par de nombreux événements tels que des dommages à l'ADN, des séquences difficiles à transcrire, des contraintes topologiques, une déplétion en nucléotide, un obstacle physique sur l'ADN ou, comme évoqué précédemment, la structure de la chromatine (revue : Wilson et al., 2013).
Une cartographie à haute résolution des ARN pol II en cours d'élongation, établie en utilisant la technique NET-Seq (native elongation transcript sequencing) développée par Churchman et Weissman montre notamment une hausse de la pause transcriptionnelle devant chacun des quatre premiers nucléosomes situés en aval des promoteurs (Churchman and Weissman, 2011).

Facteurs impliqués dans la résolution de l'arrêt transcriptionnel

L'arrêt de la transcription est caractérisé par une « marche arrière » de l'ARN pol II le long de l'ADN matrice, généralement de 2 à 14pb (Gu and Reines, 1995b). Durant cette marche arrière l'appariement ARN-ADN est maintenu mais l'extrémité 3' du transcrit sort du site actif de la polymérase. Les complexes d'élongation en pause transcriptionnelle et arrêtés sont ciblés par des facteurs de transcription, dont TFIIS et TFIIF qui agissent de concert pour résoudre la pause ou l'arrêt transcriptionnel et redémarrer l'élongation une fois que l'obstacle est éliminé ou surmonté. TFIIS aide notamment l'ARN pol II à couper l'ARN lorsqu'elle a reculé afin de replacer l'extrémité 3' de l'ARN dans le centre catalytique de la polymérase (revue : Svejstrup, 2007).

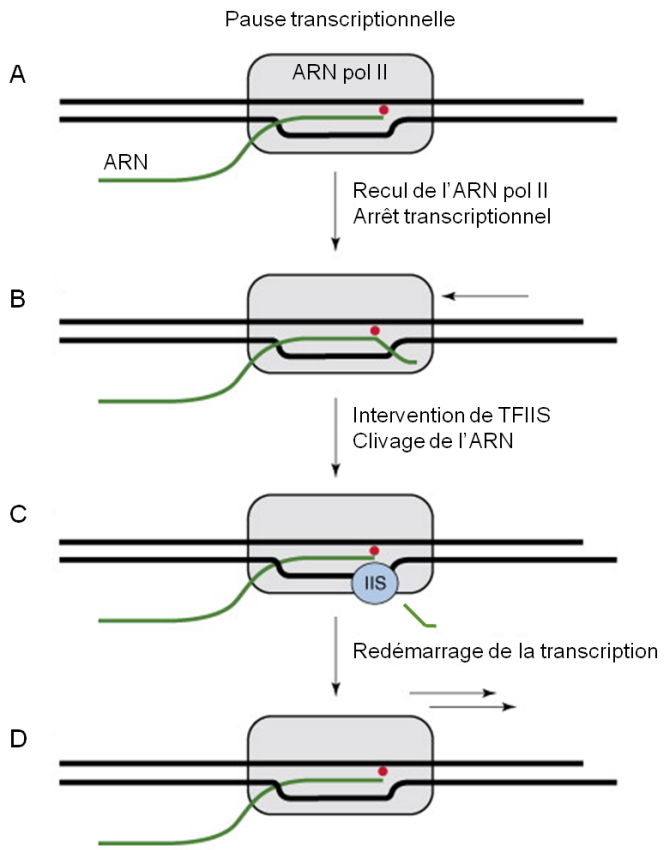

Figure 1.6. Résolution de l'arrêt transcriptionnel par TFIIS.
Lors d'une pause de la transcription (A), l'ARN pol II peut reculer et passer dans un état d'arrêt transcriptionnel dans lequel l'extrémité 3' de l'ARN est déplacée du site catalytique (point rouge) de l'ARN pol II (B). TFIIS (IIS) stimule le clivage endonucléolytique de l'ARN par l'ARN pol II, ce qui permet de repositionner l'extrémité 3' du transcrit dans le site catalytique de l'ARN pol II (C). De cette manière, la transcription peut redémarrer permettant éventuellement à l'ARN pol II de surmonter l'obstacle initial (D).
Figure adaptée de Svejstrup, 2007.

TFIIS (Dst1 chez *S. cerevisiae*) est un facteur général d'élongation s'associant aux ARN pol II arrêtées connu pour favoriser le clivage endonucléolytique des ARN associés aux ARN pol II reculées, ce qui permet un redémarrage de la transcription (Izban and Luse, 1992a; Reines, 1992, figure 1.6.). TFIIS aide également l'ARN pol II à transcrire à travers les sites de pause sur l'ADN en favorisant la marche arrière des polymérases pausées (Schweikhard et al., 2014).

Dans le complexe d'élongation arrêté et reculé, 8 nucléotides de l'ARN sorti du site actif interagissent avec une région conservée chez les eucaryotes dans « l'entonnoir » de la polymérase provoquant des changements conformationnels de l'ARN pol II, stabilisant le complexe polymérase-ADN-ARN et bloquant l'élongation. Pour résoudre ce blocage, TFIIS entre dans l'entonnoir de la polymérase, rectifie les réarrangements conformationnels et stimule le clivage endonucléolytique du transcrit

par l'ARN pol II permettant le réalignement de son extrémité 3' avec le site actif de la polymérase (Cheung and Cramer, 2011). Cette activité de clivage semble essentielle pour la survie cellulaire puisque l'utilisation d'un mutant de TFIIS bloquant le clivage par l'ARN pol II est létal (Sigurdsson et al., 2010).

TFIIF est un facteur général de la transcription associé au PIC. Il a également été montré que ce facteur est capable de se lier au complexe d'élongation et qu'il stimule *in vitro* le taux de transcription, notamment en résolvant la pause transcriptionnelle (Izban and Luse, 1992b; Renner et al., 2001; Zawel et al., 1995). Des études structurales menées plus récemment suggèrent que TFIIF se lie à la surface de l'ARN pol II, au niveau du tunnel d'entrée de l'ADN (Chen et al., 2010; Eichner et al., 2010), et stimule l'ouverture de la bulle de dénaturation de l'ADN (Gong et al., 2005; Zhang et al., 2005a). De cette manière, TFIIF pourrait aider les polymérases arrêtées à redémarrer.

Dans certain cas TFIIS et TFIIF ne sont pas suffisant pour redémarrer la transcription une ARN pol II arrêtée, la polymérase est alors éliminée par un mécanisme de poly-ubiquitination menant à la dégradation de la grande sous-unité de l'ARN pol II par le protéasome.

Ubiquitination et dégradation de l'ARN pol II

Le mécanisme d'élimination des ARN pol II arrêtées a d'abord été identifié dans le cas de polymérases bloquées par des dommages à l'ADN (Bregman et al., 1996; Ratner et al., 1998).

Lorsqu'une ARN pol II est bloquée par une lésion sur l'ADN elle peut avoir deux effets antagonistes sur la réparation de la lésion. D'une part elle favorise le recrutement de facteurs de réparation de l'ADN par le mécanisme du « transcription-coupled nucléotide excision repair » (TC-NER). Mais d'autre part, lorsque ce système de réparation n'est pas suffisant pour réparer la lésion, l'ARN pol II bloque l'accès à l'ADN des autres machineries de réparation requises. Pour libérer l'accès à l'ADN, les ARN pol II arrêtées sont ubiquitinées et désassemblées, ce qui favorise la réparation des lésions de l'ADN (Wilson et al., 2013a).

Par la suite, il a été montré que l'ubiquitination et la dégradation de l'ARN pol II est une réponse générale aux complexes d'élongation arrêtés. Par exemple une délétion de TFIIS entraîne une augmentation de l'ubiquitination de l'ARN pol II (Somesh et al., 2005). En effet, en absence de TFIIS la cellule perd un mécanisme de résolution des arrêts transcriptionnels et compense cette perte par l'utilisation d'un second mécanisme.

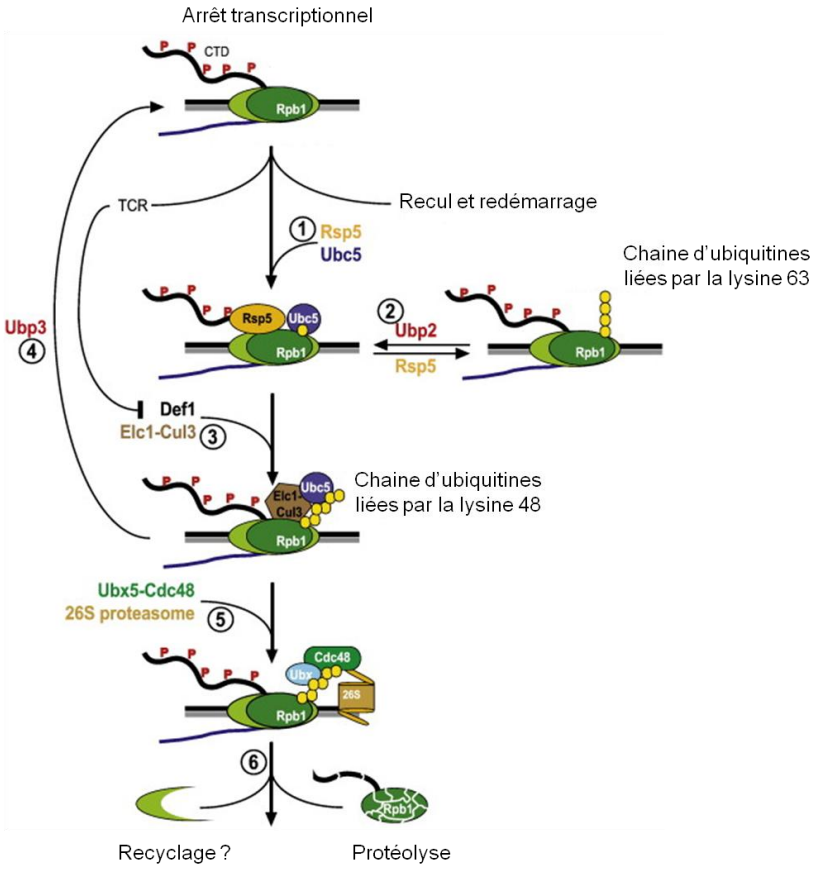

Figure 1.7. Modèle pour l'ubiquitination et la dégradation des ARN pol II arrêtées.
Les ARN pol II arrêtées peuvent être éliminées par un mécanisme d'ubiquitination et de dégradation la grande sous-unité, Rpb1. Rpb1 est tout d'abord mono-ubiquitinée par Rsp5 et Ubc5 (1). Cela peut donner lieu à la formation par Rsp5 d'une chaine d'ubiquitines reliées par la lysine 63 qui est dégradée pas l'enzyme Ubp2 (2). Elc1, Cul3 et Def1 reconnaissent la forme mono-ubiquitinée de Rpb1 et synthétisent une chaine d'ubiquitines reliées par la lysine 48 (3). Pour empêcher la dégradation de Rpb1 cette chaine peut être éliminée par Ubp3 (4). Si la chaine est maintenue, le complexe Cdc48-Ubx5 et le protéasome 26S sont recrutés au niveau de l'ARN pol II (5). Cela mène au désassemblage du complexe d'élongation et à la dégradation de Rpb1 par le protéasome alors que les autres sous-unités sont libérées (6).
Figure adaptée de Wilson et al. 2013

L'ubiquitination de l'ARN pol II suit plusieurs étapes consécutives : une première ubiquitine ligase (E3) mono-ubiquitine Rbp1 puis une deuxième intervient pour allonger la chaîne d'ubiquitine. La première mono-ubiquitination de Rpb1 est assurée

par l'enzyme de conjugaison de l'ubiquitine (E2) Ubc5 et l'E3 ligase Rsp5 (NEDD4 chez les mammifères) (Somesh et al., 2007, figure 1.7.).
Le mécanisme de reconnaissance des ARN pol II arrêtées par le complexe Rsp5-Ubc5 est inconnu, cependant il existe quelques indices permettant de comprendre la façon dont ce dimère se lie aux complexes d'élongation. Un certain nombre d'observations *in vitro* et *in vivo* montrent une forte préférence de ce complexe pour l'ARN pol II en cours d'élongation. D'une part, Rsp5 cible l'ARN pol II en cours d'élongation en interagissant directement avec le CTD de Rpb1 dont les sérines 5 ne sont pas phosphorylées, état caractéristique de l'élongation (Somesh et al., 2005). D'autre part, la structure de l'ARN pol II en cours d'élongation est spécifiquement reconnue par Ubc5. Ubc5 est la première E2 identifiée comme jouant un rôle dans la reconnaissance spécifique d'un substrat d'ubiquitination, ce rôle étant normalement endossé les ubiquitines ligases (E3) (Svejstrup, 2007).

Rsp5 catalyse l'ubiquitination de Rpb1. Rsp5 synthétise une chaîne d'ubiquitine mais la structure de celle-ci est inadaptée au désassemblage de l'ARN pol II. Pour permettre le désassemblage de l'ARN pol II, la chaine d'ubiquitine formée par Rsp5 est dégradée par Ubp2 jusqu'à laisser un simple monomère d'ubiquitine (Harreman et al., 2009). Ce monomère d'ubiquitine est alors reconnu par la protéine Def1 qui recrute un complexe capable de former une chaîne d'ubiquitine à partir de cette mono-ubiquitination (Wilson et al., 2013b). Ce complexe comprend les élongines Elc1 et Ela1 et l'ubiquitine ligase Cul3, il synthétise une chaîne d'ubiquitine qui mène à la dégradation de Rbp1 par le protéasome (revue : Wilson et al., 2013).

Une fois poly-ubiquitinée par Cul3, l'ARN pol II est désassemblée par Cdc48. Cdc48, est capable de se fixer à la fois aux chaînes d'ubiquitine ainsi qu'à Rpb1 et à Cul3 à l'aide de l'adaptateur Ubx5 et d'extraire Rpb1 de l'ARN pol II pour permettre sa dégradation (revue : Dantuma and Hoppe, 2012). Dans le cas des ARN pol II arrêtées, Cdc48 extrait spécifiquement Rpb1 poly-ubiquitinée du complexe d'élongation et l'adresse au protéasome 26S (Verma et al., 2011). De plus, il a été montré que le protéasome 26S est recruté directement au niveau de l'ADN, sur les régions transcrites, à la fois constitutivement et après induction de dommages à l'ADN (Auld et al., 2006; Gillette et al., 2004; Krogan et al., 2004), probablement pour y supprimer les ARN pol II arrêtées.

Rpb1 semble être la seule sous-unité de l'ARN pol II à être dégradée suite à un arrêt transcriptionnel (Malik et al., 2008).

L'ubiquitination de Rpb1 est réversible. Upb3 est capable de dégrader les chaînes d'ubiquitines ajoutées par Cul3, empêchant la dégradation de Rpb1 (Kvint et al., 2008). Ainsi lorsqu'*UPB3* est délété, Rpb1 est dégradée plus rapidement. Ubp3 semble limiter le taux de dégradation de Rpb1, peut être pour laisser le temps d'agir aux autres systèmes de résolution de l'arrêt transcriptionnel capables de redémarrer la transcription. En effet, ce processus de dégradation des ARN pol II n'est pas économique, il peut être vu comme un mécanisme de « dernier recours » utilisé pour assurer la continuité de la transcription des gènes et la réparation de l'ADN, sacrifiant le recyclage des ARN pol II.

I.2.4. La terminaison de la transcription

La terminaison de la transcription par l'ARN pol II suit deux principales voies alternatives. Pour chacune d'elle, la terminaison est provoquée par le recrutement de complexes de terminaison qui se fixent à la fois à des séquences spécifiques sur le transcrit naissant et à l'ARN pol II.
Les deux complexes de terminaison alternativement responsables de la terminaison sont le complexe de clivage et de polyadénylation (CPF-CF) et le complexe Nrd1-Nab3-Sen1 (NNS). Le CPF-CF est impliqué dans la terminaison des ARNm et le complexe NNS participe à la terminaison de nombreux transcrits non-codants tels que les sn(o)ARN et certains produits de la transcription envahissante présentés ultérieurement. Dans la section suivante je décris la composition et le fonctionnement de ces deux complexes.

I.2.4.1. La terminaison par le complexe CPF-CF

Le CPF-CF reconnait un ensemble de séquences spécifiques sur l'ARN naissant puis il clive cet ARN et ajoute une queue poly(A) à l'extrémité 3' ainsi produite. Cette queue poly(A) est constituée d'environ 70 adénosines (250 chez les mammifères) et elle est importante pour la stabilité, l'export et la traduction des ARN (revues : Mischo and Proudfoot, 2013; Porrua and Libri, 2015)

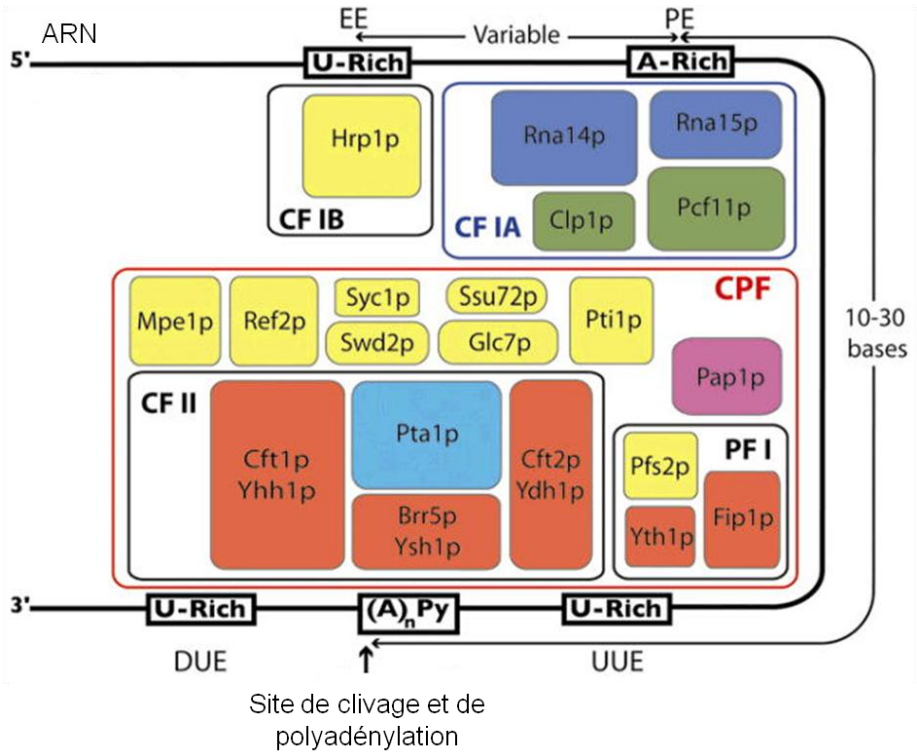

Figure 1.8. Représentation schématique du complexe CPF-CF associé à un ARN naissant.
Les différentes sous-unités du CPF-CF et les motifs constituant le signal de terminaison sur l'ARN sont brièvement décrits dans le texte ci-dessous.
Abréviations : EE, élément d'efficacité; PE, élément de positionnement; UUE, élément riche en U en amont; DUE élément riche en U en aval.
Figure adaptée de Mandel et al., 2008.

Les signaux de terminaison sur l'ARN naissant

Chez *S. cerevisiae*, les signaux de recrutement du complexe CPF-CF sont composés de cinq éléments situés dans la région du site de clivage et de polyadénylation. Ces séquences sont appelées « élément d'efficacité », « élément de positionnement », « élément riche en U en amont », « élément riche en U en aval » et « site de clivage et de polyadénylation » (Guo and Sherman, 1996; Graber et al., 1999; Mischo and Proudfoot, 2013). L'ensemble de ces séquences sont liées par des sous-unités du CPF-CF qui participent au positionnement du complexe sur l'ARN.

L'élément d'efficacité est situé de 35 à 70 nt en amont du site de clivage. Ce site stimule le clivage et il correspond au consensus TA(C/T)(A/G)TA spécifiquement lié par la sous-unité Hrp1 du CPF-CF (Minvielle-Sebastia et al., 1998).

L'élément de positionnement est la séquence la plus conservée au sein des eucaryotes (Graber et al., 1999), placé de 10 à 30 nt en amont du site de clivage, il signale la position de ce site. Cette séquence riche en adénosine (AA(A/T)AAA) est liée par Rna15, composant du CPF-CF (Gross and Moore, 2001b).

Le site de clivage est entouré de séquences riches en uracile (Graber et al., 1999) qui sont liées par les sous-unités Yth1, Cft1 et Cft2 (Barabino et al., 2000; Dichtl et al., 2002a; Dichtl and Keller, 2001). Le recrutement de ces facteurs sur ces séquences est important pour la sélection du site de clivage et pour stimuler le clivage de l'ARN *in vitro* (Dichtl and Keller, 2001).

Le site de clivage et de polyadénylation correspond à une séquence très courte, il est dans la majorité des cas constitué d'une pyrimidine suivie d'une ou plusieurs adénosines (Russo et al., 1993). Ce site est très probablement ciblé par Ysh1, l'endonucléase putative du CPF-CF (Jenny et al., 1996; Chanfreau et al., 1996; Mandel et al., 2006).

Ces cinq éléments permettent le recrutement de l'ensemble du CPF-CF constitué de plus de 20 sous-unités décrites ci-dessous (figure 1.8).

Les facteurs de terminaison du CPF-CF

L'ensemble des facteurs impliqués dans ce mécanisme de terminaison ont été purifiés et largement étudiés. En 1992, le fractionnement d'extraits de levure a mené à la séparation de quatre facteurs responsables de la formation des extrémités 3' d'ARNm *in vitro*. Ils ont été nommés PFI (polyadenylation factor I), CFI, CFII (cleavage factor I et II) et Pap1 la poly(A) polymérase (Chen and Moore, 1992). Ce fractionnement a permis de montrer *in vitro* que le clivage de l'ARN est réalisé par la combinaison des fractions CFI et CFII et que la polyadénylation est effectuée par CFI, PFI et Pap1. Par la suite, il a été démontré que le clivage et la polyadénylation sont mécanistiquement liés *in vivo*. Ces différents facteurs s'assemblent et forment un unique complexe CPF-CF (cleavage and polyadenylation factor / cleavage factor). Le CPF-CF se fixe aux signaux de polyadénylation dès leur sortie de l'ARN pol II et interagit avec le CTD de l'ARN pol II phosphorylé sur la sérine 2. Cette interaction est

importante mais pas essentielle pour la terminaison puisque la délétion du CTD atténue seulement l'efficacité du clivage des ARNm et la longueur des queues poly(A) (Licatalosi et al., 2002).

Le CPF

Le complexe CPF, constitué de 15 protéines impliquées à la fois dans le clivage et la polyadénylation des transcrits, peut être subdivisé en trois sous-complexes. Deux sont essentiels : PFI et CFII, quant au sous-complexe APT (associated with Pta1), il est facultatif. Le CPF est l'analogue fonctionnel du CPSF (cleavage and polyadenylation specificity factor) des mammifères.

PFI

PFI recrute la poly(A) polymérase Pap1 au sein du CPF permettant ainsi la polyadénylation des ARN ciblés par cette voie. Il est composé des protéines Fip1, Pfs2 et Yth1.

Fip1 est la sous-unité qui recrute Pap1, en effet Pap1 et Fip1 interagissent directement (Preker et al., 1995). De plus, Fip1 restreint l'élongation des queues poly(A) par Pap1 en limitant l'accès de Pap1 à l'ARN puisque la surface d'interaction de Pap1 et de Fip1 recouvre le domaine de fixation à l'ARN de Pap1 (Zhelkovsky et al., 1998; Helmling et al., 2001). Fip1 est fixée au reste du sous complexe PFI par son interaction avec Yth1 (Barabino et al., 2000).

Pfs2 est au cœur de l'assemblage des complexes CPF et CF via des interactions directes avec Fip1 dans le PFI, Rna14 dans le CFI et Ysh1 dans le CFII. Pfs2 participe à la coordination de l'assemblage du complexe CPF-CF notamment grâce à ses 7 répétitions WD, des surfaces d'interaction très efficaces au cœur de nombreux complexes cellulaires (Ohnacker et al., 2000; Stirnimann et al., 2010)

Yth1 se fixe aux séquences riches en U situées autour du site de polyadénylation de l'ARN et interagit avec Fip1 et Ysh1 (Barabino et al., 1997; Barabino et al., 2000). Yth1 établit ses interactions avec l'ARN et ses partenaires protéiques grâce à 5 doigts de zinc Cys-Cys-Cys-His (CCCH) et participe ainsi à l'assemblage et au positionnement spécifique du CPF-CF.

CFII

Le sous complexe CFII est directement impliqué dans le clivage de l'ARN néo-synthétisé (Zhao et al., 1997). Il est composé de 4 protéines : Ysh1, Cft1, Cft2 et Pta1.

Ysh1/Brr5 serait l'endonucléase putative responsable du clivage en 3' des transcrits chez *S. cerevisiae*. Bien que la preuve directe d'une activité nucléase portée par Ysh1 soit encore manquante, elle possède un domaine β-CASP, un motif commun parmi les hydrolases métal-dépendantes, impliqué dans le positionnement de deux ions zinc catalytiques. De plus c'est l'homologue de la protéine CPSF-73, l'endonucléase responsable du clivage 3' des ARN chez les mammifères (Jenny et al., 1996; Chanfreau et al., 1996; Mandel et al., 2006).

Cft1 et Cft2 par leurs multiples interactions participent à la formation d'une interface entre le CPF-CF, l'ARN pol II et le site de polyadénylation de l'ARN naissant.
Cft1/Yhh1 est la plus grande sous-unité du complexe CPF-CF, elle joue un rôle central dans la reconnaissance du site de clivage en liant l'ARN au niveau de la séquence riche en uracile proche de ce site (Dichtl et al., 2002a). Par ailleurs Cft1 interagit *in vitro* avec le CTD de l'ARN pol II phosphorylé avec un extrait nucléaire de cellule HeLa (Dichtl et al., 2002b) et elle se lie également avec un grand nombre de sous-unités du complexe CPF-CF.

Cft2, tout comme Cft1, joue un rôle important dans le positionnement du CPF-CF sur l'ARN en se fixant aux séquences riches en uracile situées autour du site de clivage (Dichtl and Keller, 2001; Zhao et al., 1997), ainsi qu'au CTD de l'ARN pol II et à plusieurs sous-unités du CPF-CF (Kyburz et al., 2003).

Pta1 est un élément structural du complexe CPF qui interagit avec les sous-unités Ssu72, Pti1 et Ysh1 (Ghazy et al., 2009). Pta1 participe également au sous-complexe APT décrit ci-dessous (Nedea et al., 2003).

APT

Environ la moitié des complexes CPF de la levure possède un sous-complexe supplémentaire, APT, qui joue un rôle dans l'efficacité de la terminaison. En son absence on observe une augmentation de la translecture des signaux de terminaison (Nedea et al., 2003). Ce complexe est composé de 6 protéines associées à Pta1 : Ref2, Pti1, Swd2, Glc7, Ssu72 et Syc1.

Ref2 est un régulateur négatif de l'élongation des queues poly(A) (Mangus et al., 2004). Cette sous-unité est aussi impliquée dans la maturation des extrémités 3' des snoARN (Dheur et al., 2003; Nedea et al., 2003). C'est une des deux seules sous-unités non essentielles du CPF-CF avec Syc1.

Pti1 semble favoriser la reconnaissance des sites de terminaison puisqu'en son absence on observe une altération du choix du site de clivage et de polyadénylation (Skaar and Greenleaf, 2002). Pti1 interagit directement avec Pta1 et Rna14 au sein du CPF-CF et il a également été montré que Pti1, tout comme Ref2, est requis pour la formation des extrémités 3' de snoARN (Dheur et al., 2003).

Swd2 a un rôle structural important dans le CPF-CF en reliant les sous complexes APT, CPF et CF par des interactions avec Pta1, Cft1 et Ysh1 du CFII, Pfs2 du PFI, et avec les membres de l'APT Glc7 et Pti1 (Dichtl et al., 2004). Comme pour Pfs2, Swd2 établit ses nombreuses interactions grâce à 7 répétitions WD. Cette protéine est requise pour la formation des extrémités 3' de certains ARNm et snoARN (Dichtl et al., 2004).

Glc7 fait partie d'au moins deux complexes, d'une part c'est la sous-unité catalytique de la phosphatase de protéine de type 1 et d'autre part c'est une sous-unité du CPF-CF. Dans le complexe CPF-CF, Glc7 déphosphoryle Pta1, ce qui semble favoriser le recrutement de Fip1 qui interagit avec la poly(A) polymérase Pap1 (He and Moore, 2005). La déplétion de Glc7 mène à une accumulation de la forme phosphorylée de Pta1 et à un raccourcissement de la queue poly(A). Ce raccourcissement pourrait être provoqué par un défaut de recrutement de Fip1 et Pap1 au sein du CPF-CF. Glc7 est recrutée dans le complexe APT par Ref2 (Nedea et al., 2003) et elle est aussi recrutée par le complexe de terminaison NNS décrit plus bas, dans lequel elle

déphosphoryle Sen1 (Nedea et al., 2008) mais le rôle précis de ces déphosphorylations dans la terminaison reste à établir.

Ssu72 est une phosphatase qui joue un rôle à la fois dans l'initiation et la terminaison de la transcription. Ssu72 intervient dans la terminaison des ARNm et des snoARN (Ganem et al., 2003; Steinmetz and Brow, 2003) où elle est essentielle pour la formation des extrémités 3' *in vitro* indépendamment de son activité phosphatase (He et al., 2003; Dichtl et al., 2002b).
Par ailleurs, lors de l'initiation, Ssu72 interagit avec TFIIB et Rpb2 dans le PIC et participe à la sélection du site d'initiation de la transcription (Sun and Hampsey, 1996; Pappas and Hampsey, 2000; Dichtl et al., 2002b). Son activité phosphatase permet la déphosphorylation des sérines 5 et 7 du CTD de l'ARN pol II (figure 1.2.) (Krishnamurthy et al., 2004; Zhang et al., 2012).

Syc1 (Similar to Ysh1 C-terminal), semble avoir un rôle de régulateur négatif de la terminaison par le CPF-CF. En effet, sa délétion entraîne une amélioration de la croissance de nombreux mutants du CPF-CF et *in vitro*, son absence réduit le défaut de terminaison de mutants Ysh1-1 (Zhelkovsky et al., 2006). Comme son nom l'indique, Syc1 possède une grande similarité avec Ysh1, l'endonucléase putative du CPF-CF puisque 199 acides aminés de sa séquence ont 35% d'identité et 52% de similarité avec la région C-terminale d'Ysh1 (Nedea et al., 2003). Syc1, par sa similarité avec la région C-ter d'Ysh1, pourrait concurrencer Ysh1 et interagir avec certains de ses partenaires tout en ne portant pas d'activité endonucléase. De cette façon Syc1 pourrait ralentir ou diminuer l'efficacité de la terminaison par le CPF-CF limitant l'utilisation de terminateurs aberrants (Zhelkovsky et al., 2006).

Le CFI

Le facteur de clivage I a été identifié comme un facteur nécessaire pour les réactions de clivage et de polyadénylation (Chen and Moore, 1992), il est composé des deux sous-facteurs CFIA et CFIB. Ces deux sous-facteurs se fixent aux motifs de terminaison en amont du site de clivage de l'ARN et participent ainsi au recrutement et au positionnement de l'ensemble du CPF-CF (figure 1.8.). CFIA est composé des 4 protéines Rna14, Rna15, Pcf11 et Clp1. Sa reconstitution à partir de protéines recombinantes montre que la stœchiométrie de ses 4 sous-unités est respectivement

2-2-1-1 (Gordon et al., 2011). CFIB correspond à une seule protéine, Hrp1/Nab4 (Kessler et al., 1996; Gross and Moore, 2001a).

CFIA

Rna14 et Rna15 ont été identifiées comme des protéines nécessaires pour la production des queues poly(A) et pour la stabilité des ARN (Minvielle-Sebastia et al., 1991). De nombreuses études ont montré que ces deux protéines sont fortement associées et forment un hétéro-tétramère (Kessler et al., 1996; Gross and Moore, 2001a; Noble, 2004; Gordon et al., 2011).

Rna14 semble être un élément important dans la structure du CFIA, elle contient 10 répétitions HAT (half a TRP), des structures uniquement retrouvées dans les protéines au cœur de complexes de traitement des ARN. Ces répétitions sont probablement impliquées dans les interactions protéine-protéine nécessaires à la formation du CPF-CF (Preker and Keller, 1998).

Rna15 se fixe spécifiquement à l'élément de positionnement sur l'ARN naissant grâce à un motif de reconnaissance de l'ARN dans sa région N-terminale (Minvielle-Sebastia et al., 1991). Sa fixation à l'ARN a été montrée par pontage aux UV (Gross and Moore, 2001b) et par chromatographie d'affinité (Birse et al., 1998). Cependant, l'interaction spécifique de Rna15 pour l'élément de positionnement dépend de la présence de Rna14 et Hrp1 (Gross and Moore, 2001b).

Pcf11 est une des protéines au cœur de la structure du complexe CPF-CF, elle interagit avec de nombreuses sous-unités du CFIA et du CPF (Amrani et al., 1997; He et al., 2003; Ghazy et al., 2009; Ghazy et al., 2012). De plus, un domaine d'interaction avec le CTD a été identifié dans sa région N-terminale. Il permet l'interaction directe de Pcf11 avec le CTD phosphorylé sur la sérine 2 caractéristique d'une élongation avancée (figure 1.2.) (Barillà et al., 2001; Licatalosi et al., 2002; Sadowski et al., 2003). De cette manière, la phosphorylation dynamique du CTD au cours de l'élongation régule le recrutement des facteurs de terminaison.

Clp1 interagit également avec de nombreuses sous-unités du CFIA et du CPF. Elle lie notamment Pcf11 avec qui elle forme un pont entre le CFIA et le CPF (Kyburz et al., 2003; Ghazy et al., 2009; Ghazy et al., 2012).

CFIB

Le sous-facteur CFIB est constitué uniquement de la protéine Hrp1 (Kessler et al., 1996) (Kessler et al., 1997). Hrp1 n'est pas nécessaire à la coupure de l'ARN naissant mais participe à la sélection du site de clivage en réprimant l'utilisation de sites de clivage alternatifs (Minvielle-Sebastia et al., 1998). Hrp1 semble remplir ce rôle en se liant à l'élément d'efficacité en amont du site de polyadénylation des ARN. En effet, Hrp1 possède dans sa région centrale deux motifs de reconnaissance de l'ARN arrangés en tandem (Henry et al., 1996), capables de se fixer spécifiquement aux répétitions UA de l'élément d'efficacité (Chen and Hyman, 1998; Pérez-Cañadillas, 2006; Valentini et al., 1999).

La poly(A) polymérase Pap1

Pap1 est l'enzyme qui catalyse la polyadénylation des ARN clivés par le CPF-CF, c'est le premier facteur du CPF-CF a avoir été purifié et caractérisé chez la levure (Lingner et al., 1991). Son activité enzymatique est portée par la région N-terminale qui est très structurée et hautement conservée. En effet, Pap1 et son homologue chez les mammifères partagent 47% d'identité parmi leurs 400 premiers acides aminés (Martin and Keller, 1996). Cette région permet aussi l'interaction de Pap1 avec les membres du complexe CPF-CF Cft1 et Pta1 (Ezeokonkwo et al., 2012). La région C-terminale, moins conservée, contient un domaine permettant à la fois de lier l'ARN ou d'interagir avec la protéine Fip1, facteur dont le rôle est de contrôler le recrutement spécifique et l'activité de Pap1 sur les transcrits ciblés par le CPF-CF. Pap1 possède un second domaine de liaison à l'ARN qui reconnaît l'extrémité 3' d'amorces ARN *in vitro* (Zhelkovsky et al., 1998). Ce domaine pourrait donc participer au positionnement de la région catalytique de Pap1 sur l'extrémité 3' des transcrits.

L'activité de Pap1, étroitement contrôlée par le CPF-CF, mène à la production d'extension 3' d'environ 70 nt chez la levure. Chez *S. cerevisiae* les queues poly(A) sont liées par Pab1 et Nab2, deux protéines essentielles qui sont impliquées dans le contrôle de l'activité de Pap1, dans l'export des ARNm et dans l'initiation de la transcription.

Les protéines de fixation aux extensions poly(A) Pab1 et Nab2

Les queues poly(A) sont liées dans le noyau et dans le cytoplasme par Pab1 (poly(A) binding protein 1), une protéine très conservée qui joue un rôle majeur dans le contrôle de la taille des queues poly(A) et dans l'export des ARNm (Sachs et al., 1986).

Pab1 contient 4 motifs de reconnaissance de l'ARN dans sa partie N-terminale lui permettant de se lier à des répétitions d'au moins 12 adénosines. Une même queue poly(A) peut être liée par plusieurs protéines espacées d'environ 25 nucléotides (Sachs et al., 1987), soit en moyenne 3 protéines par ARN chez la levure. Dans le noyau, Pab1 interagit avec Rna15, sous-unité du CPF-CF, et est impliquée dans le contrôle de la taille des extensions poly(A) par deux mécanismes : d'une part elle limite l'accès de Pap1 à l'ARN et d'autre part elle recrute et stimule la déadénylation par le complexe poly(A) ribonucléase Pan2-Pan3 (Amrani et al., 1997; Kessler et al., 1997; Sachs and Deardorff, 1992; Siddiqui et al., 2007; Wolf et al., 2014; Zhelkovsky et al., 1998). Pab1 est donc un exemple intéressant de protéine capable à la fois de protéger l'extrémité 3' des ARN contre l'activité d'exonucléases non spécifiques et de favoriser l'action d'exonucléases spécifiquement impliquées dans la maturation des ARN.

La fonction nucléaire de Pab1 semble aussi être liée à l'export nucléaire puisque sa déplétion provoque une rétention des ARNm à leur site de transcription (Dunn et al., 2005) . De plus, il a été montré que Pab1 interagit directement avec Xpo1, le récepteur des signaux d'export nucléaire (Hammell et al., 2002) et indirectement avec des nucléoporines (Allen et al., 2001).

Dans le cytoplasme, l'association de Pab1 aux ARNm stimule la traduction en interagissant avec une sous-unité du complexe d'initiation de la traduction, eIF4G (Tarun and Sachs, 1996 ; Tarun et al., 1997).

Dans le noyau, les queues poly(A) sont également fixées par Nab2 (Anderson et al., 1993). Cette protéine contient 7 domaines doigt de zinc en tandem qui se fixent spécifiquement aux queues poly(A) de l'ARN (Kelly et al., 2007). Jusqu'ici, deux fonctions majeures de Nab2 ont été identifiées. Tout comme Pab1, Nab2 joue un rôle dans le contrôle de la taille des queues poly(A) (Viphakone et al., 2008), la déplétion de Nab2 entraînant leur allongement (Hector et al., 2002; Kelly et al., 2010). De plus,

Nab2 est impliquée dans l'export des ARNm vers le cytoplasme : la déplétion de Nab2 provoque aussi une accumulation d'ARN polyadénylés dans le noyau (Hector et al., 2002; Green et al., 2002). Nab2 associe les ARNm aux pores nucléaires par l'interaction de son extrémité N-terminale, qui porte un domaine PWI-like, avec des protéines associées aux complexes de pore nucléaire dont Gle1, nécessaire pour l'export des ARNm (Suntharalingam et al., 2004). Nab2 reste attachée aux ARN lors de leur passage à travers les pores nucléaires puis elle est séparée de l'ARN à l'arrivée dans le cytoplasme par l'ARN hélicase Dbp5 (Tran et al., 2007).

Deux mécanismes alternatifs de terminaison

Le mécanisme de désassemblage du complexe d'élongation n'est pas complètement défini. Il existe deux modèles de couplage du détachement de l'ARN pol II et de la production de l'extrémité 3' des ARN, ils sont appelés modèle allostérique (ou anti-terminateur) et modèle torpedo.

Le modèle allostérique suggère que des modifications conformationnelles de l'ARN pol II ou la perte de facteurs « anti-terminateurs » serait provoquée par la reconnaissance du site de polyadénylation, ce qui mènerait à un décrochage progressif de la polymérase (Logan et al., 1987) (Greenblatt et al., 1993).
Le modèle torpedo propose, quant à lui, que le clivage endonucléolytique de l'ARN au site de polyadénylation permettrait le recrutement d'un facteur capable de désassembler du complexe d'élongation. Le produit du clivage encore associé à la polymérase en cours d'élongation serait alors dégradé par une exonucléase 5'-3', l'enzyme Rat1, et lorsque la nucléase atteint l'ARN pol II elle provoquerait le désassemblage du complexe ARN pol II-ARN-ADN (Connelly and Manley, 1988). A l'inverse du modèle allostérique, le modèle torpedo requiert un clivage endonucléolytique de l'ARN pour engendrer le décrochage de l'ARN pol II.

Des résultats soutiennent et contredisent chacun des deux modèles. Le modèle allostérique est soutenu par des résultats qui suggèrent que le recrutement ou la libération de facteurs de terminaison influencerait le détachement de l'ARN pol II (Licatalosi et al., 2002; Kim et al., 2004a; Calvo and Manley, 2005). Il a notamment été montré que le domaine d'interaction avec le CTD de Pcf11, composant du CFIA requis pour la formation des extrémités 3', est capable de provoquer le

désassemblage du complexe d'élongation *in vitro* (Zhang et al., 2005b). De manière intéressante, le domaine d'interaction de Pcf11 avec le CTD n'est pas nécessaire au clivage de l'ARN par le CPF-CF (Sadowski et al., 2003). Le rôle de cette interaction entre Pcf11 et l'ARN pol II reste donc énigmatique.
Le modèle torpedo est soutenu par la découverte que l'exonucléase ARN 5'-3' Rat1 est localisée à la fin de nombreux gènes et qu'elle est requise pour la terminaison (Kim et al., 2004b). Dans un mutant *rat1*, le clivage et la polyadénylation des pré-ARNm ne sont pas affectés mais un défaut de désassemblage des complexes d'élongation est observé. De plus, la déplétion de Rat1 stabilise le produit de clivage associé à la polymérase suggérant que Rat1 est l'exonucléase qui dégrade l'ARN en aval du site de clivage et de polyadénylation.

Cependant Rat1 n'est pas suffisant pour provoquer la libération de la polymérase et étonnement cette exonucléase favoriserait le recrutement de facteurs de terminaison, dont Pcf11 (Luo et al., 2006). Ces observations ont permis de proposer un modèle unifié allostérique/torpedo, présenté dans la figure 1.9., dans lequel Rat1 est un composant intégré à la machinerie de clivage et de polyadénylation.

La terminaison par le CPF-CF permet la production d'ARN polyadénylés, capables d'être exportés et traduits, c'est le mode de terminaison des ARNm. Il existe une seconde voie de terminaison responsable de la terminaison de la transcription de plusieurs classes de transcrits non-codants (Steinmetz and Brow, 1996; Wyers et al., 2005 ; Arigo et al., 2006a).

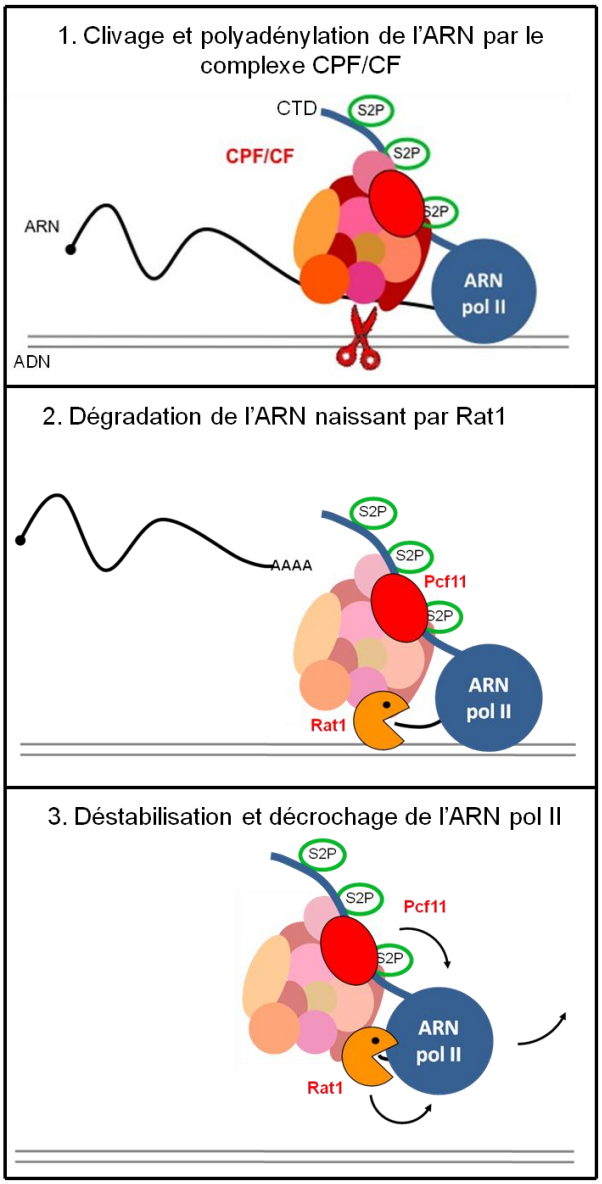

Figure 1.9. Modèle hybride de coordination de la formation des extrémités 3' et de la terminaison des cibles du complexe CPF-CF.

Ce modèle incorpore des caractéristiques des modèles allostérique et torpedo.

1. Le complexe CPF-CF est recruté à la fois sur le CTD de l'ARN pol II et sur l'ARN naissant, *via* des interactions de 6 de ses sous-unités avec des motifs spécifiques sur l'ARN et par l'interaction de 3 sous-unités avec le CTD phosphorylé sur les sérines 2. Le complexe CPF-CF clive l'ARN naissant et polyadényle l'extrémité 3'.
2. L'extrémité 5' générée par le clivage est ciblée par l'exonucléase 5'-3' Rat1 qui dégrade l'ARN jusqu'à rejoindre l'ARN pol II.
3. Les interactions combinées de Rat1 et Pcf11 avec l'ARN pol II déstabilisent la polymérase et provoque son décrochage de l'ADN.

Figure adaptée de Luo et. al, 2006.

I.2.4.2. La terminaison par le complexe Nrd1-Nab3-Sen1 :

Le NNS se lie de manière spécifique aux ARN naissants et à l'ARN pol II. Les sous-unités Nrd1 et Nab3 forment un dimère et reconnaissent des motifs courts sur l'ARN naissant. Un seul de ces motifs n'est pas suffisant pour provoquer la terminaison, ils doivent être regroupés pour former un signal de terminaison efficace. Ces motifs sont retrouvés dans les sn(o)ARN et dans de nombreux transcrits envahissants (revue : Porrua and Libri, 2013b).

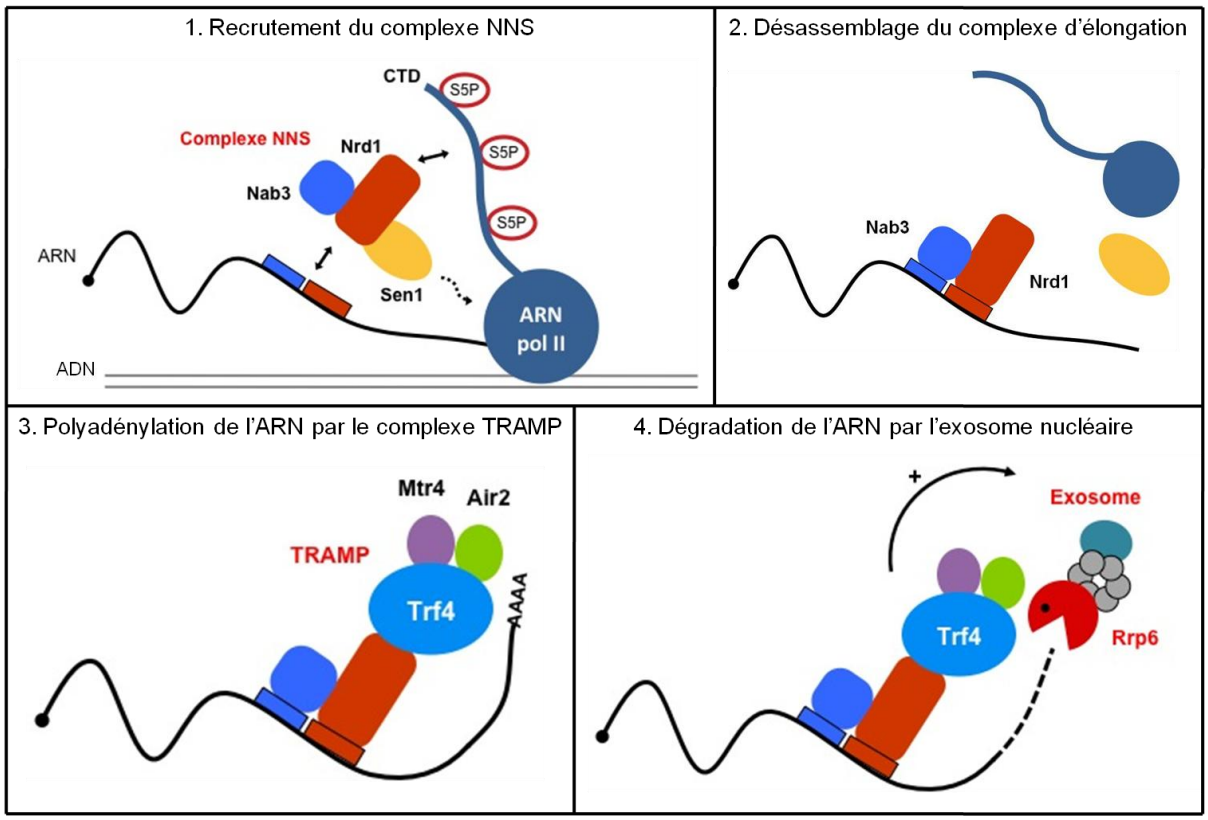

Figure 1.10. Modèle de la coordination entre la terminaison de la transcription et la dégradation des ARN ciblés par le complexe NNS.

1. Le complexe NNS interagit, via Nrd1, avec le CTD de l'ARN pol II dont les sérines 5 sont phosphorylées (S5P) et avec l'ARN naissant au niveau de sites spécifiques de liaison de Nrd1 et Nab3 (boîtes colorées).
2. Dissociation du complexe d'élongation par Sen1.
3. Le complexe TRAMP (Trf4/5 Air1/2 Mtr4 Polyadenylation complex) est recruté par Nrd1 au niveau de l'ARN via une interaction entre le CID de Nrd1 et Tfr4. Cette interaction remplace l'interaction CID-CTD. Le complexe TRAMP polyadényle ensuite l'extrémité 3' de l'ARN terminé.
4. Le complexe TRAMP contribue au recrutement de l'exosome nucléaire *via* une interaction entre Trf4 et Rrp6, favorisant ainsi la dégradation de l'ARN.

Nrd1 lie aussi le CTD phosphorylé sur la sérine 5 de l'ARN pol II par son domaine d'interaction avec le CTD (CID, CTD interacting domain). La phosphorylation de la sérine 5 du CTD etant une marque caractéristique de l'élongation précoce (Steinmetz and Brow, 1998; Vasiljeva et al., 2008) (figure 1.2.), cela permet le recrutement précoce du dimère Nrd1-Nab3 lors de la transcription. De ce fait les transcrits terminés par la voie NNS ont généralement une taille inferieure à 1000nt (Gudipati et al., 2008).
La fixation de Nrd1 et Nab3 permet de recruter l'hélicase Sen1 au niveau du complexe d'élongation, qui est dissocié par Sen1 menant à la libération du transcrit et de l'ARN pol II et provoquant ainsi la terminaison. Ce mode de terminaison n'est pas très précis, il se produit de manière progressive en aval des motifs de terminaison. Les transcrits terminés par cette voie ont donc des tailles hétérogènes.
Les transcrits terminés sont ensuite ciblés par deux facteurs de dégradation : le complexe TRAMP et l'exosome recrutés par le complexe NNS (figure 1.10.). Ces deux facteurs seront décrits ultérieurement.

Le complexe Nrd1-Nab3-Sen1

Le complexe NNS est composé de 3 protéines essentielles : Nrd1, Nab3 et Sen1. Nrd1 (19600 molécules/cellule) et Nab3 (5830 molécules/cellule) sont deux protéines abondantes (Ghaemmaghami et al., 2003) qui se lient coopérativement à l'ARN (Hobor et al., 2011). Compte tenu de la faible abondance de Sen1, 125 molécules/cellule (Ghaemmaghami et al., 2003), l'association entre Sen1 et les dimères Nrd1-Nab3 est probablement transitoire.

Nrd1 (nuclear pre-mRNA down-regulation) est une protéine de 68kDa. Capable d'interagir avec de nombreux facteurs, elle est au cœur du mécanisme de terminaison par le NNS. Elle coordonne les étapes de ciblage de l'ARN naissant, de sa libération, de sa polyadénylation et du recrutement de l'exosome responsable de la dégradation du transcrit.
Un motif de reconnaissance de l'ARN situé dans sa région centrale lui permet de se fixer à l'ARN de manière spécifique (Carroll et al., 2004; Steinmetz and Brow, 1996).
Nrd1 interagit avec les deux autres sous-unités du NNS, Sen1 et Nab3 (Conrad et al., 2000; Vasiljeva and Buratowski, 2006) et il interagit directement avec Trf4, la poly(A) polymérase du complexe TRAMP décrit plus bas (Tudek et al., 2014).

De manière intéressante, l'expression du gène *NRD1* est soumise à une autorégulation. En effet la région 5' de sa séquence codante contient 14 sites de fixation de Nrd1 et Nab3. Ces sites permettent de réguler la quantité de protéine Nrd1 dans la cellule. Lorsque Nrd1 est en excès, la terminaison précoce de la transcription du gène *NRD1* au niveau des motifs de recrutement du NNS est particulièrement efficace, ce qui fait diminuer la quantité d'ARNm fonctionnels au profit de transcrits courts terminés par le NNS. Il a été montré que la mutation des sites de fixation de Nrd1 et Nab3 situés dans la séquence codante de *NRD1* entraine une augmentation de la densité d'ARN pol II dans la région située en aval, ce qui montre un défaut de terminaison (Arigo et al., 2006b). Dans ce cas l'ARN est terminé au niveau du terminateur CPF-CF situé en aval des sites Nrd1 et Nab3, permettant la synthèse de l'ARNm de NRD1.

Nab3 (nuclear polyadenylated RNA-binding), protéine de 90kDa, lie l'ARN grâce à un motif de reconnaissance des ARN dans sa partie centrale (Wilson et al., 1994). Nab3 interagit aussi avec Nrd1 et Sen1 (Chinchilla et al., 2012; Conrad et al., 2000).
La région C-terminale de Nab3 est capable de multimériser *in vitro* (Loya et al., 2012; Loya et al., 2013). Bien qu'à l'heure actuelle la multimérisation de la protéine Nab3 entière ne soit pas montrée, on peut imaginer qu'elle pourrait permettre aux multiples hétérodimères Nrd1-Nab3 liés à l'ARN d'interagir entre eux, ce qui pourrait favoriser leur fixation.

Sen1 (splicing endonuclease), 252kDa, est une hélicase ARN/ADN dont la mutation provoque de nombreux phénotypes tels que des défauts de terminaison de la transcription, de réparation de l'ADN, de maturation des ARN ou d'assemblage des ribonucléoprotéines. Comme décrit précédemment, Sen1 fait partie du complexe NNS et partage des interactions directes avec les deux autres membres du complexe.
L'interaction entre Sen1 et l'ARN pol II est nécessaire pour le mécanisme de terminaison. En effet, le domaine N-terminal est nécessaire pour une terminaison efficace (Ursic et al., 2004) ce qui peut être expliqué par sa capacité à interagir spécifiquement avec le CTD de l'ARN pol II (Chinchilla et al., 2012).
Le domaine ATPase essentiel de Sen1 est lui aussi nécessaire pour la terminaison puisque sa mutation engendre des défauts de terminaison des snoARN, des CUT et

de certains ARNm (Steinmetz et al., 2006). Il a également été récemment montré dans un système *in vitro* que Sen1 seule est capable de reconnaître les transcrits naissants et de dissocier le complexe d'élongation de façon dépendante de l'ATP (Porrua and Libri, 2013b).

Les signaux de terminaison sur l'ARN naissant

Le complexe NNS contient deux protéines de liaison à l'ARN, Nrd1 et Nab3. Ces deux protéines forment un dimère mais chacune de ces deux protéines lient un motif spécifique.

L'analyse des motifs de terminaison de snoARN a permis d'identifier deux motifs fréquents en 3' de ces transcrits liés par Nrd1 et de Nab3 (Carroll et al., 2004; Steinmetz et al., 2001). La composition de ces motifs a ensuite été affinée par des expériences de pontage aux UV de Nrd1 et Nab3 sur leurs cibles naturelles (Creamer et al., 2011; Wlotzka et al., 2011). Ces travaux ont montré que le site de fixation optimal de Nab3 est UCUUG, et le site de fixation de Nrd1 est UGUA(A/G).

Récemment, une expérience de sélection *in vivo* de terminateurs à partir de séquences aléatoires synthétiques menée au laboratoire a permis de mettre en lumière un autre motif impliqué dans la terminaison par le complexe NNS (Porrua et al., 2012). Dans cette expérience, les sites de fixation de Nrd1 et Nab3 déjà connus ont été enrichis dans les terminateurs sélectionnés et ils étaient fréquemment accompagnés de séquences riches en A et U longues d'environ 7 nucléotides jamais identifiées jusqu'alors. Ces séquences riches en A-U sont nécessaires pour la terminaison de la transcription par la voie NNS au niveau d'un terminateur synthétique issu de cette expérience et testé *in vivo*. Il a également été montré par anisotropie de fluorescence que ce motif A-U favorise la fixation de Nrd1 et Nab3 à l'ARN.

Nrd1 et Nab3 forment un hétéro-dimère qui se lie de manière coopérative aux sites de fixation sur l'ARN (Hobor et al., 2011). De plus, les terminateurs naturels ciblés par le complexe NNS possèdent souvent de nombreux sites de liaison de Nrd1 et Nab3. Les nombreux dimères fixés aux terminateurs pourraient interagir les uns avec les autres et se fixer coopérativement sur l'ARN, ce qui favoriserait l'efficacité de la

terminaison (Carroll et al., 2007). On peut donc imaginer que les sites de fixation de Nrd1 et Nab3 sont donc organisés les uns par rapport aux autres afin de favoriser le recrutement du complexe NNS. Cependant cette notion étant mal connue, je me suis penché sur l'étude de la composition et de la répartition de sites de fixation de Nrd1 et Nab3 au cours de ma thèse.

Une des caractéristiques de la terminaison par le complexe NNS est la dégradation des transcrits produits. Elle dépend du complexe TRAMP et de l'exosome décrits dans le chapitre suivant.

Le complexe TRAMP

Le complexe (Trf4/5-Air1/2-Mtr4) est impliqué dans la maturation, la dégradation et le contrôle de qualité des ARN. Ce complexe polyadényle une grande variété d'ARN et favorise leur dégradation ou leur maturation par l'exosome. Le complexe TRAMP cible les précurseurs et les intermédiaires anormaux de maturation des sn(o)ARN et des ARNr (LaCava et al., 2005) et les CUT (Wyers et al., 2005). Il est composé de trois protéines : une poly(A) polymérase : Trf4 ou Trf5, une protéine de liaison à l'ARN : Air1 ou Air2, et une ARN hélicase : Mtr4.
Lors de la terminaison par le NNS, le complexe TRAMP est recruté au niveau des ARN néo-synthétisés par Nrd1 via une interaction directe entre le domaine CID de Nrd1 et le domaine NIM (Nrd1 interaction motif) de Trf4 (figure 1.10.) (Tudek et al., 2014).

Deux formes du complexe existent *in vivo*, TRAMP4 (Trf4-Air2-Mtr4) et TRAMP5 (Trf5-Air1-Mtr4). TRAMP4 est le complexe majoritaire, probablement parce que Trf4 a un taux d'expression supérieur (Trf4 : 7550 molécules/cellule, Trf5 : 2240 molécules/cellules) (Ghaemmaghami et al., 2003). Ces deux complexes agissent sur des substrats à la fois de manière redondante et spécifique en fonction des transcrits. Leur spécificité de substrat est contrôlée par les protéines Air1 et Air2. En effet, le séquençage des ARN issus de souches dans lesquelles Air1 ou Air2 sont alternativement délétés montre l'accumulation de transcrits différents dans chacun des deux mutants (Schmidt et al., 2012).

Trf4/Pab2 et Trf5 sont deux poly(A)-polymérases codées par des gènes paralogues. A l'inverse d'autres poly(A)-polymérases telles que Pap1, Trf4 et Trf5 ne possèdent pas de domaine de fixation à l'ARN. Les deux protéines Air1 et Air2 servent très probablement d'intermédiaires à Trf4 et Trf5 pour se fixer à leurs ARN cibles puisqu'il a été montré que Trf4 est incapable de polyadényler son substrat *in vitro* en absence d'Air1 ou Air2 (Vanácová et al., 2005; Wyers et al., 2005).

Air1 et Air2 sont aussi codées par deux gènes paralogues, elles partagent une séquence d'acides aminés très conservées contenant 5 motifs Cys-Cys-His-Cys « zink knuckles » qui permettent des interactions protéine-protéine ou protéine-ARN. Des études biochimiques et cristallographiques ont montré que le cinquième motif d'Air2 est impliqué dans l'interaction avec Trf4 et que les deuxième, troisième et quatrième motifs interagissent avec l'ARN (Hamill et al., 2010; Holub et al., 2012).

Mtr4 est une ARN hélicase 3'-5' dépendante de l'ATP. Identifiée comme cofacteur de l'exosome dans la maturation de l'extrémité 3' de l'ARNr 5,8S (de la Cruz et al., 1998), Mtr4 joue un rôle majeur dans la maturation et la dégradation nucléaire de nombreux ARN (Hoof et al., 2000). Elle est recrutée au sein du complexe TRAMP par les deux régions N-terminales de l'hétéro-dimère Trf4-Air2 (Falk et al., 2014) et limite l'extension des queues poly(A) par Trf4/5 (Jia et al., 2011). Ce contrôle pourrait empêcher la synthèse de queues poly(A) suffisamment longue pour être fixés par Pab1. En effet, Pab1, qui nécessite au moins 12 A successifs pour se lier, protège les extrémités 3' contre les exonucléases, ce qui est antagoniste à la fonction du TRAMP.
Il a été montré que Mtr4 se fixe préférentiellement aux extrémités 3' poly(A) des ARN (Bernstein et al., 2008). De cette façon elle pourrait débobiner l'ARN de 3' vers 5' et produire une extrémité 3' d'ARN non structurée capable d'entrer dans le canal du cœur de l'exosome (structure décrite dans le paragraphe suivant) et de favoriser leur dégradation. En effet, l'exosome a une activité exonucléase faible *in vitro* sur des substrat très structurés tels que des ARNt et leur dégradation efficace nécessite à la fois Mtr4 et une polyadénylation par Trf4 (Vanácová et al., 2005).

L'Exosome

L'exosome est un complexe protéique portant 3 activités ribonucléases conservées chez les eucaryotes, il participe à la dégradation, à la maturation et au contrôle qualité de toutes les familles d'ARN chez *S. cerevisiae* (revue : Chlebowski et al., 2013 et Porrua and Libri, 2013b). L'exosome est composé de 10 sous-unités dans sa version cytoplasmique et 11 dans le noyau.

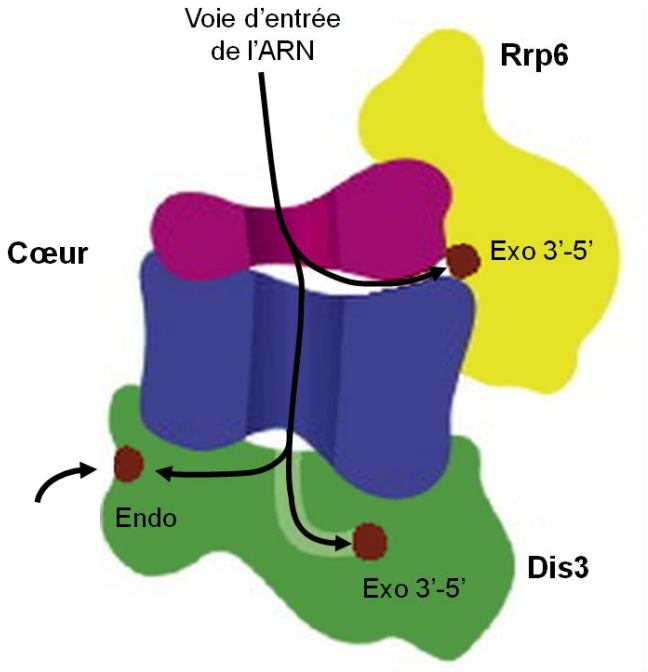

Figure 1.11. Représentation schématique de l'exosome nucléaire.
Les voies menant l'ARN aux sites catalytiques de l'exosome sont indiquées par des flèches. Pour accéder aux sites exonucléases (exo) de Rrp6 et Dis3, l'ARN passe par le tunnel central du cœur de l'exosome. Pour accéder au domaine endonucléase (endo) de Dis3, l'ARN pourrait soit passer par le tunnel central, soit y parvenir directement à partir du solvant.
Figure adaptée de Chlebowski et al., 2013.

Le cœur de l'exosome commun aux deux formes de l'exosome est composé de 9 protéines de relativement petite taille (de 20 à 50kDa). La structure du cœur de l'exosome est bien établie grâce aux données cristallographiques obtenues chez l'Homme (Liu et al., 2006). Ces 9 sous-unités sont ordonnées en 2 anneaux superposés. L'anneau inférieur est composé des 6 sous-unités Rrp41/Ski6, Rrp42, Rrp43, Rrp45, Rrp46 et Mtr3. L'anneau supérieur de ce cœur est, quant à lui, composé des 3 protéines de liaison à l'ARN : Rrp4, Rrp40 et Csl4. Ces protéines n'interagissent pas entre elles mais sont chacune directement attachées à des paires

de sous-unités de l'hexamère inférieur. L'étroit passage central qui traverse les deux anneaux n'est suffisamment large que pour laisser passer un ARN simple brin.

La fonction de ce cœur n'est pas complètement claire et pourrait servir à fixer et à débobiner l'ARN pour faciliter sa dégradation par les sous-unités catalytiques (figure 1.11). Il a notamment été montré que le cœur de l'exosome humain reconnaît spécifiquement les ARN riches en U et en AU (Anderson et al., 2006).

L'activité enzymatique de l'exosome est portée par les sous-unités Dis3 et Rrp6 qui s'associent avec le cœur sous deux formes. La forme cytoplasmique de l'exosome contient le cœur et Dis3, soit 10 sous-unités alors que la forme nucléaire de 11 sous-unités est composée du cœur, de Dis3 et de Rrp6. Ces deux sous-unités catalytiques sont de grands polypeptides multi-domaines. Dis3 et Rrp6 ont à la fois des rôles redondants et spécifiques dans la dégradation des différentes classes de transcrits (Gudipati et al., 2012). Ils ont une fonction redondante dans la dégradation des produits de la transcription envahissante mais la fonction de Rrp6 dans la maturation des sn(o)ARN semble être prédominante par rapport à Dis3. Dis3, quant à lui, semble avoir un rôle plus important dans la dégradation de précurseurs d'ARNt et de pré-ARNm contenant des introns.

Dis3/Rrp44 possède des activités endoribonucléase et exoribonucléase 3'-5' spécifiques d'ARN simple brin. L'activité exonucléase est très processive, elle attaque les extrémités 3' d'ARN simples brins et une fois fixée, Dis3 dégrade entièrement ses substrats ne laissant que des nucléosides monophosphates et un produit long de 3 à 5 nucléotides (Lorentzen et al., 2008). Par ailleurs, Dis3 est aussi capable de défaire les structures intra- et intermoléculaires des ARN tant qu'il existe une région non structurée d'au moins 4 ou 5 nucléotides en 3' (Lorentzen et al., 2008).
Dis3 est un homologue des RNAses II et R d'*E. coli* (Lorentzen et al., 2008) ce qui en fait un membre de la superfamille RNR (RNase II/RNase R-like). A l'instar des membres de cette famille, Dis3 possède dans sa partie C-terminale deux domaines CSD (cold-shock domain), un domaine RNB (domaine catalytique de la ribonucléase II) et un domaine S1. Les domaines CSD et S1 se fixent de manière non spécifique à l'ARN et le domaine RNB porte une activité exoribonucléase 3'-5' processive (Zuo

and Deutscher, 2001). Ce domaine actif est situé au cœur de Dis3, au font d'un canal étroit uniquement accessible par un ARN simple brin (Lorentzen et al., 2008).
Dis3 possède également deux autres domaines à son extrémité N-terminale : un domaine CR3, important pour l'association de Dis3 au cœur de l'exosome (Schaeffer et al., 2012) et un domaine actif PIN retrouvé dans de nombreuses protéines qui disposent d'activités exo- ou endoribonucléase (Arcus et al., 2004). Ce domaine PIN a deux fonctions, il a une activité endoribonucléase (Lebreton et al., 2008) et il permet la fixation de Dis3 à l'anneau inférieur du cœur de l'exosome par une interaction directe avec Rrp41 (Bonneau et al., 2009). Il est pensé que la fonction endonucléolytique du domaine PIN pourrait servir à créer des points d'entrée pour l'activité exonucléase de Dis3 (Gudipati et al., 2012).

Rrp6 est la sous-unité catalytique spécifique de l'exosome nucléaire, elle a une activité exoribonucléase 3'-5' (Allmang et al., 1999; Briggs et al., 1998).
Rrp6 appartient à la superfamille des nucléases DEDD qui contient de nombreuses exonucléases ou des domaines de correction d'ADN polymérases (Midtgaard et al., 2006). Tous ces membres possèdent un domaine DEDD constitué d'un site actif de 4 acides aminés conservés (3 aspartates et un glutamate, à l'origine du nom du domaine) repartis dans 3 motifs différents hautement conservés.
Rrp6 possède également des domaines HRDC (helicase and RNase D C-terminal) et PMC2NT dont les fonctions précises demeurent inconnues. Cependant, le domaine HRDC est essentiel pour la maturation de l'ARNr 5,8S et de certains snoARN (Phillips and Butler, 2003) et pourrait servir à positionner les substrats complexes, tels que ceux cités précédemment, dans la bonne orientation dans le site actif de Rrp6 (Midtgaard et al., 2006). Le domaine PMC2NT apporte lui une interface pour la fixation de Rrp47, un cofacteur qui aiderait Rrp6 à se fixer à ses substrats ARN (Stead et al., 2007). Des données de microscopie cryo-électronique et cristallographiques indiquent que Rrp6 se fixerait au niveau de l'anneau supérieur du cœur de l'exosome (Cristodero et al., 2008; Schuch et al., 2014; Wasmuth et al., 2014).

Comme décrit précédemment, l'ampleur de la dégradation par l'exosome est différente selon les transcrits. Le complexe NNS recrute l'exosome sur les précurseurs des sn(o)ARN et sur les CUT. Les précurseurs des sn(o)ARN sont

raccourcis dans le cadre de leur maturation alors que les CUT sont complètement dégradés. A l'heure actuelle, le mécanisme limitant la dégradation des sn(o)ARN n'a pas été démontré. Cependant, il est connu que ces transcrits forment des complexes avec de nombreuses protéines et certaines d'entre elles se lient à ces ARN au cours de leur transcription, elles pourraient bloquer l'avancée de l'exosome et limiter ainsi son activité exonucléase de 3' vers 5' (Grzechnik and Kufel, 2008; Vasiljeva and Buratowski, 2006).

I.2.4.3. Mécanismes de terminaison de la transcription par les ARN polymérases I et III

Chacune des trois ARN polymérases possède ses propres mécanismes d'initiation, d'élongation et de terminaison. Dans la section suivante, l'état des connaissances à propos des mécanismes de terminaison des ARN pol I et III sont rapidement décrits.

Terminaison de la transcription par l'ARN pol I

La levure contient entre 150 et 200 de copies d'ADN ribosomique concentrées sur le chromosome 12. Ces répétitions sont transcrites par l'ARN pol I qui synthétise principalement un transcrit pré-ARNr 35S dont la maturation mène à la production des trois ARNr 5,8S, 18S et 25S (revue : Németh et al., 2013).

La terminaison de la transcription par l'ARN pol I nécessite la fixation d'un facteur de terminaison sur une séquence spécifique de l'ADN. Ce facteur, Nsi1, bloque la progression de l'ARN pol I (Merkl et al., 2014; Reiter et al., 2012). Ce blocage semble être une étape nécessaire pour le recrutement des autres facteurs impliqués dans ce mécanisme de terminaison (Merkl et al., 2014; Németh et al., 2013). Un modèle de terminaison actuel propose que le clivage co-transcriptionnel d'une tige boucle sur le pré-ARNr par l'endonucléase Rnt1 créerait une extrémité 5' servant de point d'entrée à l'exonucléase Rat1 qui dégraderait l'ARN de 5' vers 3' (Hage et al., 2008). La 5' polynucléotide kinase Grc3 aiderait Rat1 en maintenant la phosphorylation de l'extrémité 5' de l'ARN nécessaire à l'action de cette exonucléase (Braglia et al., 2010). Il a également été montré que l'hélicase Sen1 est retrouvée au site de terminaison de l'ARNr et qu'elle serait impliquée dans la dégradation de l'ARN ciblé par Rat1 (Kawauchi et al., 2008). L'ARN pol I serait déstabilisée par Rat1

ainsi que par l'helicase Sen1 et par la polynucléotide kinase Grc3 ce qui provoquerait sa dissociation de l'ADN (Németh et al., 2013).

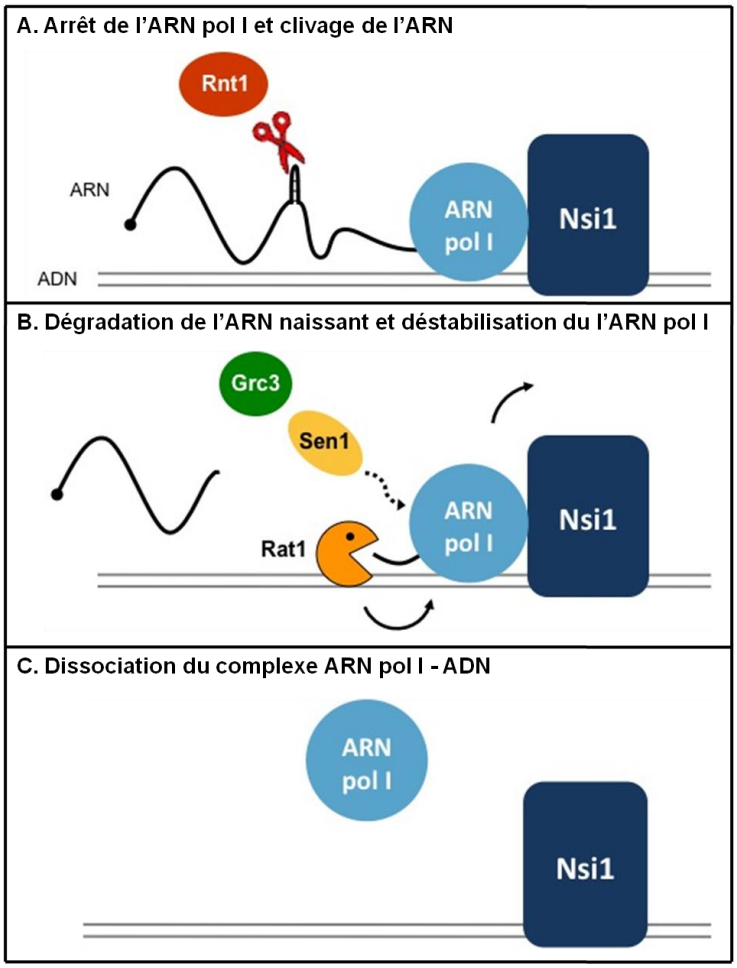

Figure 1.12. Modèle de mécanisme de terminaison de la transcription par l'ARN pol I
A. La terminaison de la transcription par l'ARN pol I reposerait sur un blocage de la transcription provoqué par la fixation de Nsi1 sur l'ADN accompagné par le clivage de l'ARN naissant. Ce clivage serait réalisé par l'endonucléase ARN Rnt1 au niveau d'une tige boucle.
B et C. L'extrémité 5' du transcrit associé à l'ARN pol I serait dégradé par l'exonucléase 5'-3' Rat1 jusqu'à atteindre la polymérase. L'ARN pol I serait déstabilisée par Rat1 ainsi que par l'helicase Sen1 et par la polynucléotide kinase Grc3 ce qui provoquerait sa dissociation de l'ADN.
Figure adaptée de Németh et al., 2013.

La terminaison de la transcription par l'ARN pol III

L'ARN pol III transcrit une diversité de petits transcrits non-codants tels que les ARNt, l'ARNr 5S et le snARN U6. De manière remarquable, le signal de la terminaison de ces transcrits est extrêmement simple, il est constitué d'au moins 5

thymidines consécutives sur le brin « codant » chez *S. cerevisiae* (Allison and Hall, 1985; Braglia et al., 2005; revue : Orioli et al., 2012).

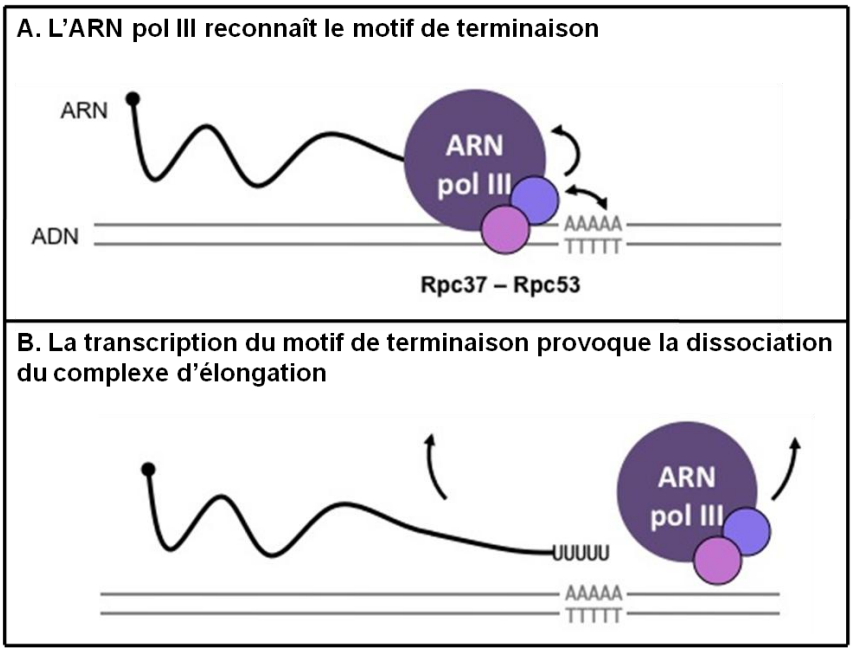

Figure 1.13. Modèle de terminaison de la transcription par l'ARN pol III.

A. L'ARN pol III, par l'intermédiaire de ses sous-unités Rpc37 et Rpc53, reconnaît le motif de terminaison constitué d'au moins cinq A consécutifs ce qui mènerait à un ralentissement de l'élongation par l'ARN pol III.
B. La transcription de ce motif qui forme un duplexe instable avec l'ADN matrice associée au ralentissement de l'élongation mènerait à la dissociation du complexe d'élongation.

Figure adaptée de Richard and Manley, 2009.

Malgré la simplicité du signal de terminaison, le mécanisme moléculaire permettant la terminaison est encore mal connu. Il semblerait que l'ARN pol III soit capable de terminer la transcription par elle-même, probablement sans l'intervention de facteurs supplémentaires. Des expériences *in vitro* de terminaison de la transcription par l'ARN pol III ont montré que les sous-unités Rpc37 et Rpc53 de cette polymérase sont impliquées dans la reconnaissance du signal de terminaison (Landrieux et al., 2006). Ces deux sous-unités forment un duplex positionné dans la région d'entrée de l'ADN d'où elles pourraient « scanner » l'ADN entrant (Fernández-Tornero et al., 2007). La reconnaissance de la séquence poly(T) par le dimère Rpc37-Rpc53 pourrait provoquer un changement conformationnel de la polymérase menant à une pause de la transcription. Cette pause faciliterait la libération de l'ARN naissant, dont l'appariement avec l'ADN (dA:rU) est particulièrement instable, et le décrochement de l'ARN pol III (Landrieux et al., 2006; Martin and Tinoco, 1980; Orioli et al., 2012).

I.3. La transcription envahissante

I.3.1. Découverte de la transcription envahissante

Le développement de nouveaux outils permettant des études transcriptomiques à grande échelle a permis de révéler une complexité inattendue du transcriptome eucaryote. Une grande portion des ARN synthétisés ne correspondent pas à des unités de transcription fonctionnelles clairement définies telles que des ARNm, ARNr ou ARNt et des milliers de nouvelles unités de transcription ont été découvertes. Ces nouveaux transcrits sont souvent rapidement dégradés et demeurent invisibles tant que les machineries de dégradation des ARN sont actives (revue Jensen et al., 2013; Tisseur et al., 2011) et ces observations ont donné naissance aux concepts de transcription cachée et de transcription envahissante.

Chez *S. cerevisiae*, ces événements de transcription non conventionnels ont tout d'abord été observés lorsque des facteurs de remodelage de la chromatine ont été mutés. En effet, il a été montré que la mutation de Spt6, une chaperonne nécessaire à la maintenance de la structure de la chromatine (voir chapitre I.2.3.2.), provoque des événements d'initiation de la transcription aberrants dans les régions intragéniques (Kaplan et al., 2003).

Peu de temps après, des analyses transcriptomiques à l'échelle du génome par puce à ADN ont montré l'existence de très nombreux transcrits instables (Davis and Ares, 2006; Houalla et al., 2006; Wyers et al., 2005). Ces observations sont à l'origine du concept de la transcription cachée. En absence de Rrp6, une des deux sous-unités catalytiques de l'exosome nucléaire, de nombreux ARN sont stabilisés et deviennent observables au niveau de régions inter et intragéniques, ainsi que chevauchant en sens ou en anti-sens des phases ouvertes de lecture. Ces nouveaux transcrits ont été appelés CUT pour cryptic unstable transcripts. Les CUT sont synthétisés par l'ARN pol II et leur instabilité est liée à leur mode de terminaison faisant intervenir le complexe NNS qui provoque leur dégradation par l'exosome nucléaire (Thiebaut et al., 2006). Compte tenu de leur mode de terminaison, spécifique des ARN pol II en début d'élongation, les CUT sont relativement courts, d'une taille moyenne située entre 200 et 500 nt et dotés d'extrémités 3' hétérogènes.

La caractérisation de l'ensemble des CUT a mené à la découverte d'autres transcrits envahissants moins sensibles à Rrp6. Ces ARN ont donc été appelés SUT pour stable unannotated transcripts (Neil et al., 2009; Xu et al., 2009). Ces transcrits, plus longs que les CUT, d'une taille médiane de 760pb, ne sont que partiellement sensibles à la dégradation par l'exosome nucléaire. Ils sont principalement affectés par les voies de dégradation cytoplasmique telles que le Nonsense Mediated decay (NMD), chargé de supprimer les transcrits aberrants, et par la dégradation 5'-3' par l'exoribonucléase Xrn1 (van Dijk et al., 2011; Marquardt et al., 2011).Le choix du mécanisme de terminaison pourrait expliquer la différence entre CUT et SUT. Alors que les CUT sont terminés par la voie NNS, la transcription des SUT semble être terminée par le CPF-CF, ce qui permettrait leur export (Marquardt et al., 2011).

L'inactivation de l'exoribonucléase cytoplasmique Xrn1 a permis l'identification d'une troisième classe d'ARN instables synthétisés par l'ARN pol II et dégradé par cette exoribonucléase, ils ont été nommés XUT (Xrn1-sensitive unstable transcripts) (van Dijk et al., 2011). Cela montre que les transcrits envahissants peuvent être dégradés par différents systèmes de contrôle qualité.

Les CUT, SUT et XUT forment trois classes de transcrits définis par leurs modes de terminaison et de dégradation. Cependant la frontière entre ces trois catégories n'est pas toujours claire. Par exemple, ces transcrits partagent parfois des sites d'initiation de la transcription communs et une part d'entre eux peut être terminée par des signaux de terminaison du NNS, ils correspondent alors à des CUT et sont dégradés par l'exosome nucléaire. Une autre part de ces transcrits peut passer à travers ces premiers signaux de terminaison et être terminés en aval par un terminateur CPF-CF. Ces ARN non-codants terminés par le CPF-CF peuvent être stabilisés et exportés vers le noyaux où ils sont dégradés par le NMD ou Xrn1, ce qui peut les définir comme des SUT ou des XUT (Marquardt et al., 2011). Chacun des transcrits formant ces trois catégories d'ARN non-codants sont donc ciblés par une des machineries de dégradation qui peut être remplacée par une autre lorsqu'elle est contournée ou inactivée.

La neutralisation d'une autre enzyme de dégradation également a permis de découvrir une classe supplémentaire de transcrits instables : la délétion de Rat1, l'exoribonucléase 5'-3' nucléaire, a mené à l'identification chez *S. cerevisiae* de transcrits instables provenant des régions télomeriques appelés « telomeric repeat-containing RNA » (TERRA) (Luke et al., 2008). Ces ARN étaient déjà connus chez les vertébrés où ils participent au maintien de la structure des télomères (Azzalin et al., 2007).

Des études similaires portant sur la caractérisation complète du transcriptome ont été effectuées chez l'Homme et elles ont également permis la découverte de nouvelles classes d'ARN telles que les PASR (promoter-associated small RNA), les TASR (terminator-associated short RNA) (Kapranov et al., 2007) et les PROMPT (promoter upstream transcript) (Preker et al., 2008). Des petits ARN, proches des PASR, ont ensuite été découverts par séquençage d'ARN et nommés TSSa-RNA (transcription start site associated RNA) (Seila et al., 2008) et ti-RNA (tiny RNA) (Taft et al., 2009). Parmi ces ARN, les PROMPT sont proches des CUT puisqu'ils sont dégradés par l'exosome nucléaire (Flynn et al., 2011; Ntini et al., 2013; Preker et al., 2008) cependant leur mode de terminaison est probablement différent puisqu'aucun homologue de Nrd1 ou Nab3 n'est connu chez l'Homme.

I.3.2. Origines de la transcription envahissante

Les transcrits envahissants ont tout d'abord été définis par leur mode d'élimination mais les mécanismes permettant leur synthèse commencent à être compris. Les CUT, les SUT et les XUT sont produits par l'ARN pol II et proviennent en majorité des régions déplétées en nucléosome (NDR) en 5' et en 3' des gènes, l'ADN décondensé de ces régions étant plus accessible pour les ARN pol II (figure 1.14.).

Les CUT et les SUT sont par exemple fréquemment issus de transcription divergente à partir de promoteurs de gènes (Neil et al., 2009). Au niveau de nombreux promoteurs, deux complexes de pré-initiation semblent se former. L'un d'eux permet la production d'un ARN messager dans un sens et l'autre d'un transcrit non-codant dans l'autre direction. L'étude de la répartition des facteurs généraux de la transcription sur l'ensemble du génome a confirmé que dans le cas de la transcription divergente de CUT et d'ARNm, initiés à partir d'une même NDR, les

deux transcrits emploient deux PIC indépendants (Murray et al., 2012; Rhee and Pugh, 2012). Dans certains cas, ces deux PIC semblent être en compétition pour le recrutement des facteurs d'initiation. En effet la mutation de la boîte TATA nécessaire pour la transcription d'un ARNm peut avoir deux effets opposés sur la synthèse de l'ARNm et du CUT divergent. La mutation de la boîte TATA diminue la production de l'ARNm et elle augmente aussi la transcription du CUT divergent (Neil et al., 2009). Cependant, dans d'autres cas, les ARNm et leurs transcrits divergents associés semblent être co-régulés, cela signifie qu'ils seraient contrôlés et partageraient les même activateurs transcriptionnels (Xu et al., 2009). Etant donné que les CUT sont extrêmement répandus et que la plupart d'entre eux correspondent à des transcrits divergents, il a été proposé que les promoteurs de *S. cerevisiae* soient fondamentalement bidirectionnels (Neil et al., 2009). Le choix de la voie de terminaison permet alors la synthèse d'un ARNm stable par le complexe de terminaison CPF-CF dans un sens et la dégradation du produit de la transcription divergente par le recrutement du complexe NNS.
Il existe également des CUT transcrits dans la même direction qu'un ARNm, à partir du même promoteur. Ce type d'événement peut être utilisé pour des fonctions de régulation par atténuation de la transcription de l'ARNm notamment dans le cas d'au moins deux gènes participant à la synthèse de nucléotides (Kuehner and Brow, 2008; Thiebaut et al., 2008). Ce type de mécanisme est décrit dans le chapitre I.3.3.1.

Le génome de *S. cerevisiae* étant très dense, la région NDR située en 3' d'un gène correspond dans environ la moitié des cas à la NDR située en 5' du gène en aval. Les CUT, SUT et XUT transcrits à partir de telles régions sont donc à la fois issus d'un promoteur divergent et synthétisés en anti-sens d'un gène. Il existe également des transcrits non-codants synthétisés en anti-sens de gènes, à partir de la NDR en 3', indépendamment de la présence de promoteurs divergents. En effet, de courtes régions correspondant aux 3' de gènes, insérées au milieu d'ORF sont capables d'à la fois initier un transcrit anti-sens et de terminer la transcription sens (Murray et al., 2012). Dans le cas des CUT anti-sens, leur transcription se termine généralement après quelques centaines de nucléotides et n'interfère pas avec la transcription de l'ARNm sur l'autre brin. A l'inverse, les transcrits qui échappent à la terminaison précoce, tels que les SUT ou les XUT, peuvent atteindre le promoteur de l'ARNm

synthétisé en anti-sens et interférer avec la transcription de celui-ci (Castelnuovo et al., 2013).

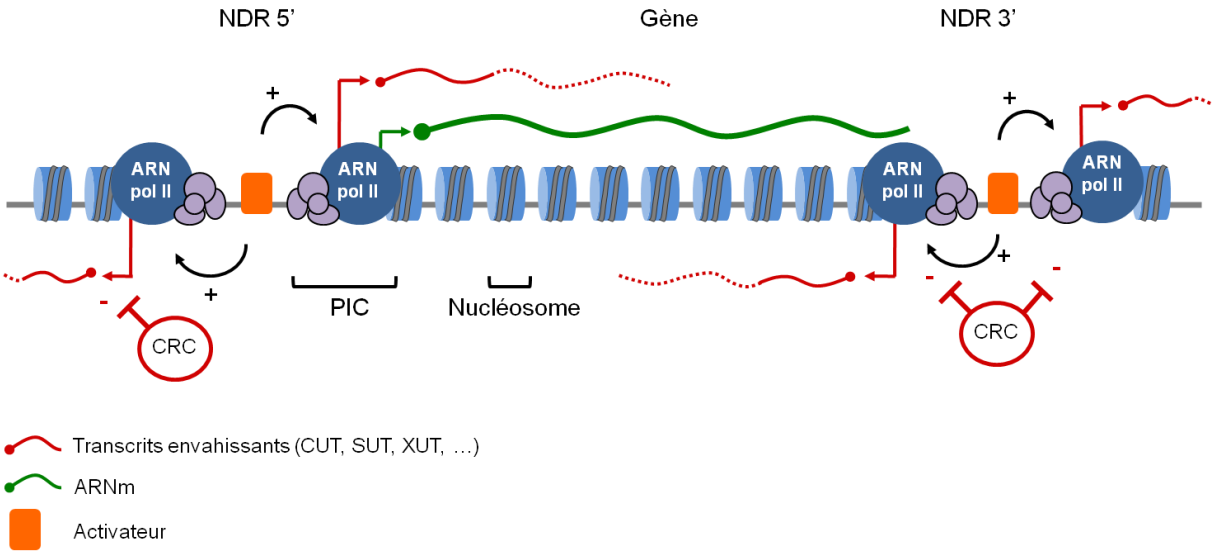

Figure 1.14. Origine de la transcription envahissante.
Représentation schématique d'une unité de transcription comprenant un gène et des ARN non-codants transcrits par l'ARN pol II. L'ARNm est représenté en vert. Les régions déplétées en nucléosome (NDR) en 5' et 3' sont susceptibles d'être le site d'initiation de transcrits envahissants, colorés en rouge. La transcription envahissante anti-sens en amont du gène est initiée à partir d'un complexe de pré-initiation (PIC) indépendant dont la formation peut être favorisée par la présence d'activateur de la transcription du gène. Un promoteur peut diriger la transcription à partir de différents sites d'initiation dans le même sens (voir *URA2*), certain d'entre eux pouvant mener à la production de transcrits instables et d'autres permettront la synthèse d'un ARNm.
La taille des NDR est limitée par les complexes de remodelage de la chromatine (CRC) qui forment la structure répressive de la chromatine ce qui a pour conséquence de limiter les événements d'initiation aberrants.
Figure adaptée de Colin et al., 2011 et Jensen et al., 2013.

Tout comme pour les ARNm, la transcription des CUT, SUT et XUT varie en fonction des conditions physiologiques.Par exemple, la méiose provoque une grande modification du profil de transcription des ARN non-codants avec l'induction des « meiotic unannotated transcripts » (MUT) qui sont sensibles à Rrp6 et peuvent donc être considérés comme des CUT transcrits de façon transitoire (Lardenois et al., 2011). Bien que ces variations du niveau de transcription soient probablement indirectement provoquées par les changements d'expression de nombreux gènes, elles pourraient aussi indiquer un rôle important de ces ARN non-codants dans le contrôle de l'expression génique.

I.3.3. Fonctions de la transcription envahissante

Les analyses transcriptomiques à grande échelle ont donc permis de découvrir de nouvelles classes de transcrits. Ces découvertes ont très rapidement mené à de nouvelles interrogations quant à l'effet de ces ARN non-codants. Sont-ils simplement le fruit d'événements de transcription futiles ou peuvent-ils porter une quelconque fonction biologique ? (revue : Wu et al., 2012).

I.3.3.1. Régulation de la transcription par des ARN non-codants

Comme cela a été rapidement évoqué dans les paragraphes précédents, la transcription d'au moins une partie de ces nouveaux ARN joue un rôle dans la régulation de la transcription. Il semblerait donc qu'au moins une partie de ces événements de transcription envahissante ait une fonction biologique.

Le premier exemple de transcription non-codante responsable de la régulation d'un gène par interférence transcriptionnelle chez *S. cerevisiae* a été décrit pour le gène *SER3* impliqué dans la synthèse de la sérine et de la glycine (Martens et al., 2004). Il a été montré que le transcrit *SRG1* initié en amont de *SER3* lorsque la sérine est abondante et que sa transcription réprime *SER3*. La transcription de *SRG1* transforme la NDR du promoteur de *SER3* en région d'ADN condensé sous forme de nucléosome impropre à l'initiation. En effet, l'élongation de la transcription de *SRG1* provoque un repositionnement des nucléosomes sur le promoteur de *SER3* de façon dépendante du facteur d'élongation Spt6, une chaperonne d'histone responsable de la reformation des nucléosomes après le passage de l'ARN pol II (voir chapitre I.2.3.2). Les nucléosomes ainsi positionnés bloquent l'accès du promoteur de *SER3* aux facteurs de transcription (Hainer et al., 2011). En absence de sérine la transcription de *SRG1* est réprimée, ce qui permet l'expression de *SER3* (Martens et al., 2004; figure 1.15.A.). *SRG1* produit plusieurs transcrits, certains étant stables alors que d'autres sont ciblés par l'exosome nucléaire en fonction de l'utilisation de la voie de terminaison CPF-CF ou NNS (Davis and Ares, 2006; Thiebaut et al., 2006; Thompson and Parker, 2007).

Le gène *URA2* est un exemple de régulation par choix alternatif du site d'initiation de la transcription de l'ARNm ou d'un CUT transcrit dans la même orientation. Le

promoteur du gène *URA2* permet la production d'un CUT et d'un ARNm dans la même direction. La transcription peut démarrer à partir de deux groupes de sites d'initiation. Les sites en amont, les plus proches du promoteur, mènent à la production d'un CUT et les sites situés plusieurs dizaines de nucléotides en aval sont utilisés pour synthétiser l'ARNm (figure 1.15.B.).
En présence d'uracile, l'expression d'*URA2* n'est pas requise et seul le CUT est transcrit; quand l'expression d'*URA2* est nécessaire, la carence en uracile provoque l'utilisation du site d'initiation de la transcription de l'ARNm, de cette façon le transcrit évite les signaux de terminaison et de dégradation et devient un ARNm fonctionnel (Thiebaut et al., 2008). Le CUT est transcrit quelque soit la quantité d'uracile et sa transcription est requise pour la régulation de l'expression d'*URA2*. Ce type d'organisation a également été caractérisé pour d'autres gènes impliqués dans la synthèse de nucléotides tels que *IMD2*, *URA8* et *ADE12* (revue : Colin et al., 2011).

La transcription d'un CUT passant à travers une région promotrice peut aussi réprimer la transcription en établissant des modifications de la chromatine inhibitrices. L'étude du locus *GAL1-GAL10* a montré qu'un CUT initié en amont de *GAL1* en présence de glucose (lorsque les gènes *GAL1* et *GAL10* ne sont pas exprimés) atténue l'induction de *GAL1* (figure 1.15.A.). En effet, la transcription du CUT provoque la compaction de l'ADN par la déacétylation co-transcriptionnelle des histones, provoquant ainsi l'atténuation de la transcription de *GAL1* (Houseley et al., 2008; Pinskaya et al., 2009).

Le gène *PHO84* code un transporteur de phosphate, il possède un CUT initié en anti-sens à partir de sa région 3' pouvant réprimer l'activité basale du promoteur du gène. En absence de *RRP6* une forme allongée du CUT est accumulée. D'une manière encore mal comprise, lorsque *RRP6* est délété, le CUT ne semble pas être terminée au niveau de signaux de terminaison précoces par le complexe NNS, sa transcription est étendue et traverse le promoteur de *PHO84*. Cette transcription modifie la structure locale de la chromatine, la faisant passer de région promotrice en région transcriptionnelle active inadaptée au recrutement de l'ARN pol II. Comme dans le cas du locus *GAL1-10*, le mécanisme de répression fait notamment intervenir les histones méthyltransférases Set1 et Set2 ainsi que la HDAC Rpd3 (Camblong et al., 2007; Castelnuovo et al., 2013) (figure 1.15.C.).

Le gène *PHO5* présente un rare exemple de régulation positive par la transcription d'un CUT anti-sens (figure 1.15.D). Cet ARN anti-sens est initié dans la région 3' de *PHO5* et sa transcription se termine au niveau de la région promotrice de ce même gène. Dans ce cas, la transcription à travers le promoteur ne provoque pas d'interférence transcriptionnelle mais semble plutôt favoriser une éviction des histones par un mécanisme qui reste à élucider (Uhler et al., 2007). Cette modification de la structure du promoteur faciliterait le recrutement des facteurs de transcription nécessaires à l'expression de *PHO5*.

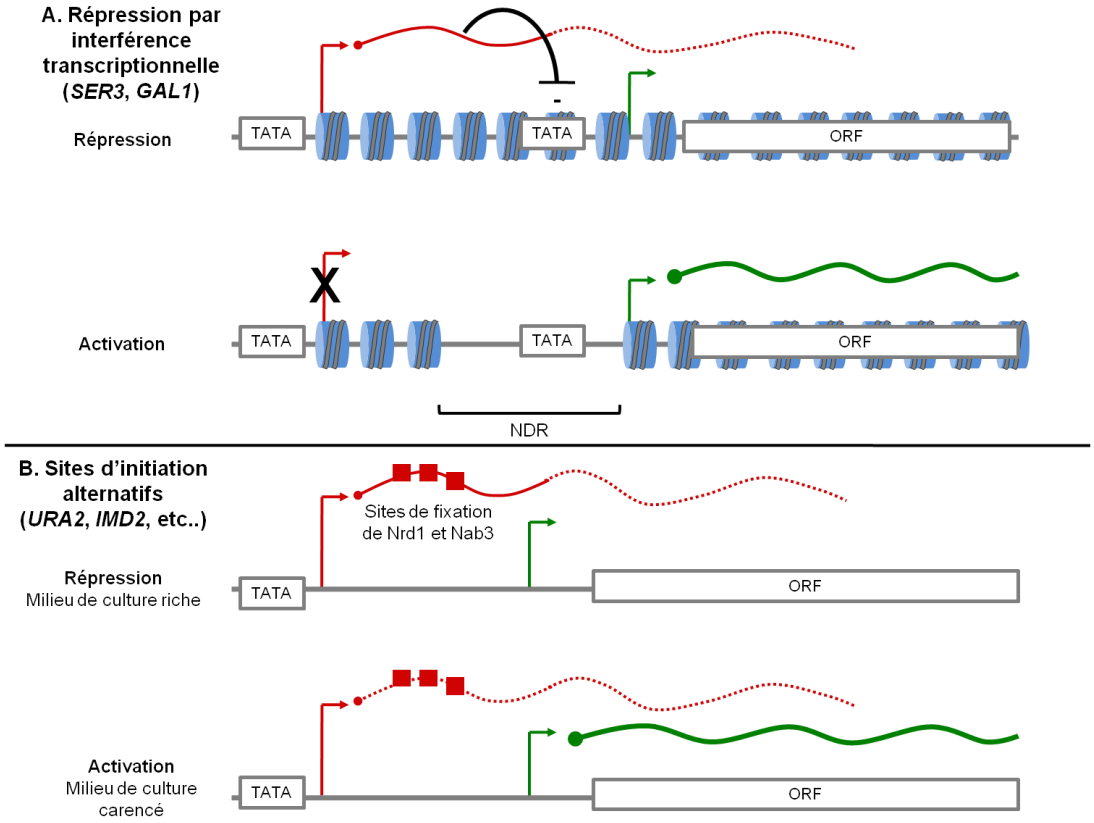

Figure 1.15. Exemples de mécanismes de régulation de l'expression de gènes par la transcription envahissante.

A. La transcription d'ARN instables (en rouge) en amont des gènes *SER3* et *GAL1* réprime la transcription des ARNm (en vert). Cette répression est basée sur la formation co-transcriptionnelle de chromatine répressive au niveau du promoteur du gène. Cette répression est levée lorsque la transcription de l'ARN instable est inhibée (croix noire).
B. Certains gènes impliqués dans la synthèse de nucléotides, tels qu'*URA2* et *IMD2*, ont un promoteur capable d'initier la transcription à partir de deux sites d'initiation. Le site d'initiation le plus proche du promoteur permet la synthèse d'un ARN instable contenant des sites de fixation de Nrd1 et Nab3 (carrés rouges). Le second site d'initiation n'est utilisé que lorsque l'expression du gène est nécessaire (en carence de nucléotide) et mène à la transcription de l'ARNm.

Figure adaptée de Colin et al., 2011 et Wu et al., 2012.

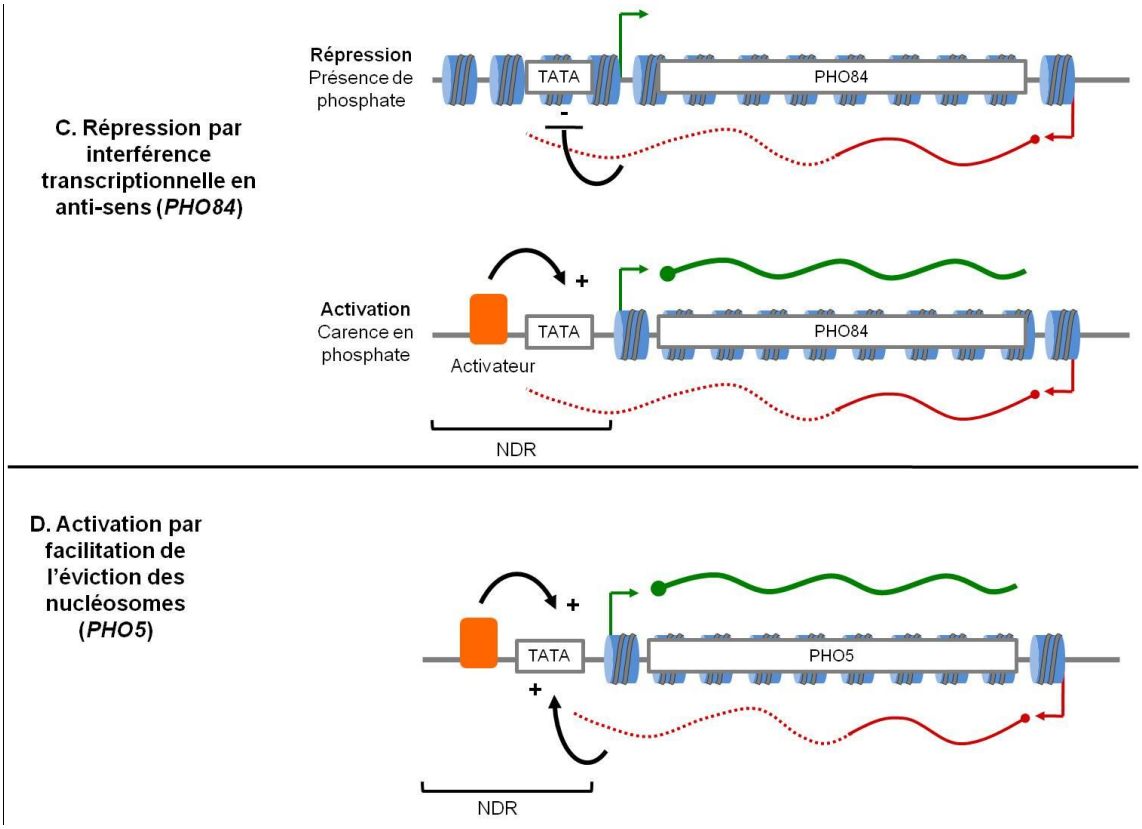

Figure 1.15. Exemples de mécanismes de régulation de l'expression de gènes par la transcription envahissante.

C. Dans le cas du gène *PHO84* la transcription d'un ARN non-codant en anti-sens du gène provoque la formation de chromatine répressive sur le promoteur de *PHO84*.
D. L'ARN non-codant transcrit en anti-sens du gène *PHO5* favorise l'activation de la transcription de l'ARNm en facilitant l'éviction des histones sur le promoteur de *PHO5* permettant le recrutement d'activateurs de la transcription (rectangle orange).

Figure adaptée de Colin et al., 2011 et Wu et al., 2012.

Dans une grande partie des cas de régulation par la transcription d'ARN non-codants instables chez *S. cerevisiae*, le mécanisme de régulation implique le passage des complexes d'élongation de ces transcrits à travers les promoteurs des gènes entrainant le repositionnement des nucléosomes au niveau du promoteur. L'ARN non-codant ne semble souvent pas fonctionnel par lui-même, ce sont plutôt les facteurs de remodelage de la chromatine associés aux complexes d'élongation qui transforment les régions promotrices déplétées en nucléosome en région d'élongation de la transcription dans lesquelles l'ADN est plus compact. Un mécanisme de compaction qui semble avoir été sélectionné pour contrôler et éviter l'initiation de la transcription au sein des ORF.

PHO5 est un contre exemple de ce mécanisme de répression. Il pourrait être intéressant d'étudier si, dans ce cas, la terminaison de la transcription de l'ARN anti-

sens provoquerait la formation d'une NDR juste au niveau du promoteur de *PHO5* permettant de cette façon le recrutement de l'ARN pol II.
Il existe aussi des exemples de gènes réprimés par des ARN instables transcrits en anti-sens et dégradés dans le cytoplasme pour lesquels il a été proposé que l'ARN soit un acteur direct de la régulation. Ces ARN favoriseraient, en trans, le recrutement de répresseurs de la transcription ou de facteurs de remodelage de la chromatine sur le promoteur du gène en anti-sens (van Dijk et al., 2011). C'est notamment ce qui a été proposé dans le cas de la répression de la transcription du retro-transposon *TY1*, qui possède un XUT anti-sens. Ce XUT semble favoriser en trans l'action de l'histone methyltransférase Set1 au niveau du promoteur de *TY1,* ce qui inhibe la transcription du retro-transposon (Berretta et al., 2008).

I.3.3.2. Les transcrits envahissants : une source de nouveaux gènes ?

Une fonction régulatrice associée à la transcription envahissante n'a été montré que pour une minorité de CUT, de SUT ou de XUT. Par ailleurs, une partie de ces transcrits sont situés dans des régions intergéniques et ne sont pas associé à des séquences codantes et compte tenu de leur dégradation rapide ils n'ont probablement pas de fonction directe *per se*. Cependant, on peut imaginer que la transcription d'ARN non fonctionnels puisse servir de « pépinière de gènes ». La mutation de nucléotides au niveau de ces régions transcrites pourrait transformer une de ces unités de transcription en un nouveau gène. Dans certains cas, ces nouveaux gènes pourraient être utiles pour donner un avantage compétitif à leurs hôtes notamment dans l'adaptation à des stress environnementaux ou améliorer l'accès à une source de nutriment. Ce système pourrait être un outil très utile d'un point de vue de l'adaptation et de l'évolution.

De ce point de vue, l'exemple du gène *FYV5/MDF1,* uniquement présent chez *S. cerevisiae,* est particulièrement intéressant. Fyv5 a pour fonction de lier la protéine Matα2, un des déterminants du signe sexuel de *S. cerevisiae*. En se liant à Matα2, Fyv5 diminuerait l'efficacité de la conjugaison et conserverait les cellules à l'état haploïde. L'expression de *FYV5* pourrait ainsi accélérer la vitesse de croissance de *S. cerevisiae* en favorisant la croissance végétative des cellules haploïdes.

L'apparition de ce gène a pu donner un avantage compétitif aux levures pour la colonisation de nouveaux milieux et aurait été conservé.
Des analyses bioinformatiques et expérimentales suggèrent que le gène *FYV5* soit apparu chez *S. cerevisiae*. Il semblerait que ce gène soit issu d'une séquence non-codante ancestralement transcrite en anti-sens du gène *ADF1* (Li et al., 2010). Tandis qu'*ADF1* est un gène conservé chez quasiment tous les hémiascomycètes dont le génome est connu, *FYV5*, quant à lui, n'est présent que chez *S. cerevisiae*. Il n'est pas retrouvé chez les autres levures du genre *Saccharomyces,* seules des ORF tronquées homologues de FYV5 et non traduites sont retrouvées chez certaines espèces (Li et al., 2010). *FYV5* est donc très probablement un gène récent sur le plan évolutif.

La transcription envahissante et la synthèse de transcrit non fonctionnel pourrait donc être une source de nouveaux gènes en plus des principaux mécanismes d'acquisition de gènes dérivés de gènes préexistants qui peuvent être modifiés suites à des duplications, divergences, retro-transpositions, modifications d'exon ou transferts horizontaux (Long et al., 2003).

I.3.4. Mécanismes de contrôle de la transcription envahissante

Bien que la transcription envahissante puisse avoir des fonctions, elle doit être limitée car, comme le montrent les exemples de régulation négative par la transcription envahissante, elle est potentiellement délétère pour l'expression des gènes. De plus, une trop grande quantité d'ARN non fonctionnels peut être toxique s'ils sont traduits, s'ils séquestrent des facteurs de liaison à l'ARN limitants ou s'ils interagissent avec des séquences d'ADN complémentaires. La levure possède donc une batterie de systèmes lui permettant de maîtriser la transcription envahissante par le contrôle de l'initiation de la transcription et par la dégradation des transcrits non fonctionnels (revue : Jensen et al., 2013).

I.3.4.1. Contrôle de l'initiation de la transcription

La quantité d'information nécessaire pour le recrutement de l'ARN pol II sur l'ADN est faible. Lorsque l'ADN est décompacté et exposé il est sujet à de nombreux événements d'initiation de la transcription. Dans tous les cas de décompaction de

l'ADN, qu'elle soit naturelle aux deux extrémités des gènes ou provoquée artificiellement par l'inactivation de facteurs de remodelage de la chromatine, on observe des événements d'initiation de transcription envahissante.

S. cerevisiae emploie donc toute une série de facteurs capables de compacter l'ADN en positionnant des nucléosomes « protecteurs » en dehors des extrémités 5' et 3' des unités de transcriptions, ces facteurs sont décrits dans le chapitre I.2.3.2. Parmi ces facteurs, le complexe HDAC Rpd3S et la chaperonne d'histone Spt6 sont nécessaires pour la maintenance de la structure de la chromatine et la reformation des nucléosomes après le passage de l'ARN pol II. Il a été montré que leur inactivation provoque l'apparition de sites d'initiation de la transcription aberrants (Churchman and Weissman, 2011; Kaplan et al., 2003).

La levure contrôle également la dimension des régions déplétées en nucléosome. Il existe deux principales classes de NDR, les NDR 5' et les NDR 3', situées de part et d'autre des phases ouvertes de lecture. Ces NDR peuvent être chevauchantes dans le cas de gènes en tandem proches. Elles font généralement de 80 à 300pb et sont délimitées par deux nucléosomes précisément placés. Les NDR 5' correspondent aux promoteurs de nombreux gènes et sont enrichies en site de fixation de facteur de transcription (Yuan et al., 2005) et elles démarquent le site d'initiation de la transcription de la région transcrite en aval (Albert et al., 2007; Lee et al., 2007). Les NDR 3' sont, comme leur nom l'indique, placées à l'extrémité 3' des gènes et sont enrichies en sites de terminaison de la transcription (Mavrich et al., 2008).
De nombreux facteurs tels que des facteurs de transcription, des complexes de remodelage de la chromatine (Badis et al., 2008) ou la structure intrinsèque de l'ADN (Iyer and Struhl, 1995) favorisent la formation ou l'extension des NDR. L'activité de ces facteurs faciliterait l'initiation de la transcription en offrant un meilleur accès à l'ADN aux facteurs de transcription.
S. cerevisiae est également capable de contrebalancer l'effet de ces facteurs et de limiter l'extension des NDR par l'activité du complexe de remodelage de la chromatine Isw2 (Whitehouse and Tsukiyama, 2006; Whitehouse et al., 2007). Isw2 est retrouvée associé à des centaines de positions différentes dans le génome. Ce complexe fonctionne à la limite des régions promotrices et positionne les nucléosomes à l'interface entres les régions géniques et intergéniques en

« poussant » les nucléosomes vers les NDR limitant ainsi leur dimension. De manière intéressante, la délétion d'*ISW2* provoque la transcription de CUT divergents à partir des NDR des promoteurs. La limitation de l'extension des NDR par Isw2 permettrait de limiter l'initiation de la transcription de transcrits envahissants (Yadon et al., 2010).

I.3.4.2. Contrôle des transcrits envahissants par le couplage de la terminaison et de la dégradation des ARN

Un grand nombre de transcrits envahissants échappent au contrôle de l'initiation de la transcription. Ils atteignent alors un deuxième niveau de contrôle lors de la terminaison. Parmi ces transcrits, les CUT possèdent des motifs de fixation de Nrd1 et Nab3 ce qui provoque leur terminaison par le complexe NNS et leur dégradation par l'exosome nucléaire (voir chapitre I.2.4.2.). De cette manière ces ARN sont supprimés par un couplage entre leur terminaison et la dégradation. Les autres classes de transcrits envahissants ne sont pas ciblés par le complexe NNS et sont dégradés de manière post-transcriptionnelle.

I.3.4.3. Contrôle post-transcriptionnel des transcrits envahissants

Un grand nombre de transcrits envahissants ne sont pas supprimés par le couplage entre la terminaison et la dégradation. La transcription des XUT et des SUT est terminée par le CPF-CF, le mécanisme limitant leur accumulation est donc différent de celui des CUT. Ces ARN sont exportés vers le cytoplasme et sont pris en charge par les machineries de dégradation cytoplasmique.
Les SUT sont partiellement sensible à l'exosome nucléaire mais sont principalement dégradés par le NMD (Marquardt et al., 2011). Comme décrit précédemment les XUT sont, eux, dégradés par Xrn1 (van Dijk et al., 2011).

I.3.5. Un nouveau mécanisme de contrôle de la transcription envahissante

Au cours d'une expérience menée au laboratoire dans le but d'identifier de nouveaux éléments de séquence impliqués dans la terminaison de la transcription par l'ARN pol II, de nouveaux motifs de terminaison ont été identifiés. Cette expérience était basée sur une sélection *in vivo* de terminateurs à partir de séquences aléatoires. Les résultats obtenus ont été particulièrement intéressants, ils ont notamment permis

d'identifier de nouvelles séquences provoquant la terminaison. Parmi celles-ci, le motif de liaison du facteur de transcription Reb1 était particulièrement enrichi (Colin et al., 2014).

I.3.5.1. Le facteur de transcription Reb1

Reb1 est une protéine nucléaire abondante (7500 molécules/cellule (Ghaemmaghami et al., 2003)). C'est un facteur de transcription capable de se lier à des motifs spécifiques sur de très nombreux promoteurs *via* deux motifs myb-like (Ju et al., 1990). Le nombre de sites de liaison liés par Reb1 dépasse le millier (Rhee and Pugh, 2011). Un certain nombre d'études ont montré que le recrutement de Reb1, comme c'est le cas pour d'autres membres de la famille des facteurs généraux de régulation Rap1, Abf1 et Tbf1, est capable de limiter la propagation de la chromatine et agit comme un « insulateur » délimitant les régions génomiques (Fourel et al., 2002; Yu et al., 2003). Reb1 est un facteur impliqué dans la régulation de la transcription par les ARN polymérases I et II en favorisant la formation de NDR au niveau des promoteurs (Hartley and Madhani, 2009; Raisner et al., 2005). Au moins deux mécanismes de formation de NDR par l'intermédiaire de Reb1 ont été proposés. D'une part Reb1 pourrait favoriser la déplétion des nucléosomes *via* le recrutement du complexe de remodelage de la chromatine RSC, en interagissant physiquement avec Rsc2, Rsc3 et Npl6 (Gavin et al., 2002; Hartley and Madhani, 2009), des sous-unités du complexe RSC (Cairns et al., 1996). D'autre part Reb1 pourrait favoriser la formation de NDR en formant une barrière bloquant l'expansion de la chromatine dans les NDR puisque les sites de liaison de Reb1 sont très fréquemment situés au niveau du nucléosome -1, tout juste en amont des NDR (Koerber et al., 2009).
Par ailleurs Reb1 se fixe également aux régions subtélomériques et semble jouer un rôle dans la limitation de l'élongation des télomères (Berthiau et al., 2006).

Reb1 a tout d'abord été identifié par erreur comme le facteur de terminaison de la transcription des ARNr (Morrow et al., 1989; Lang and Reeder, 1993; Lang et al., 1994). *In vitro* Reb1 est bien capable de terminer la transcription des ARNr par l'ARN pol I mais *in vivo* c'est Nsi1, paralogue de Reb1, qui est le véritable responsable de la terminaison des ARNr par un mécanisme appelé « road-block » (chapitre I.2.4.3.) (Merkl et al., 2014; Reiter et al., 2012).

La terminaison par Reb1

Comme l'analyse de la sélection de terminateurs l'avait suggéré, le site de fixation de Reb1 est capable de provoquer la terminaison. Le laboratoire a également montré que Reb1 provoque la terminaison de la transcription par l'ARN pol II par un mécanisme road-block similaire à la terminaison de la transcription des ARNr par Nsi1 (figure 1.16.).

Reb1 agit comme une barrière fixée à l'ADN qui arrête l'élongation par l'ARN pol II. Les ARN pol II arrêtées au niveau des sites de fixation de Reb1 suivent la voie d'élimination classique des polymérases arrêtées. Elles sont poly-ubiquitinées et seraient dégradées par le protéasome (voir chapitre I.2.3.3. Contrôle de l'arrêt transcriptionnel) provoquant la libération des ARN qui sont polyadénylés par le complexe TRAMP et dégradés par l'exosome nucléaire (Colin et al., 2014). Ce mécanisme de terminaison tout d'abord caractérisé à partir de séquences synthétiques est également utilisé dans le génome de *S. cerevisiae*.

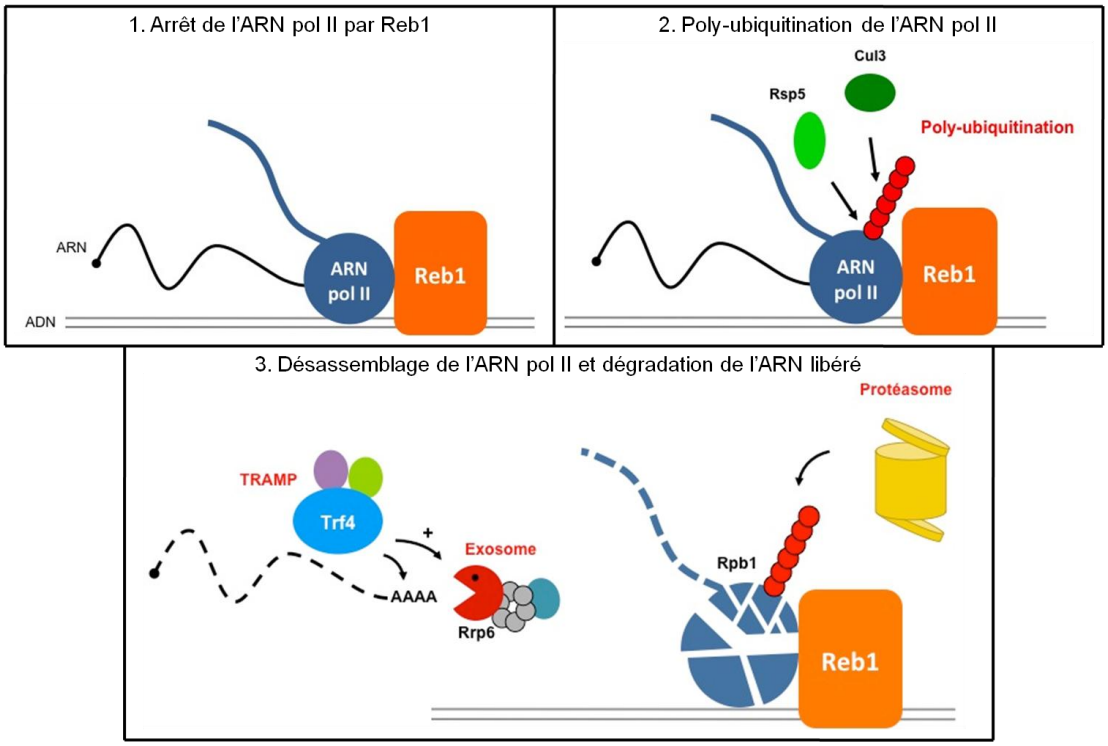

Figure 1.16. Modèle de terminaison de la transcription par Reb1.

1. Reb1, fixé sur l'ADN, bloque la progression de l'ARN pol II ce qui provoque un arrêt de la transcription.
2. L'ARN pol II arrêtée est prise en charge par l'ubiquitine ligase Rsp5 qui mono-ubiquitine Rpb1 puis un complexe ubiquitine ligase contenant Cul3 allongerait la chaîne d'ubiquitine.
3. L'ARN pol II poly-ubiquitinée serait désassemblée et Rpb1 dégradée par le protéasome 26S. L'ARN libéré est polyadénylé par la sous-unité du complexe TRAMP, Trf4. Trf4 favoriserait également la dégradation de l'ARN par l'exosome nucléaire.

Figure adaptée de Colin et al., 2014.

Un mécanisme de terminaison de secours

Compte tenu de la présence fréquente de Reb1 au niveau des promoteurs, cette voie de terminaison correspond à un mécanisme de terminaison de secours capable de bloquer les complexes d'élongation qui auraient échappé à la terminaison par les complexes CPF-CF ou NNS. Reb1 protègerait de cette façon les promoteurs contre les phénomènes d'interférence transcriptionnelle entre des unités de transcription placées en tandem. En effet, il a été montré que dans certains cas la déplétion de Reb1 permet de visualiser l'existence de transcription qui passe à travers le site Reb1 et qui inhibe l'expression de gènes situés en aval (Colin et al., 2014).

La terminaison par Reb1 représente ainsi un mécanisme supplémentaire de contrôle de la transcription envahissante.

I.3.5.2. Rap1, un autre facteur de terminaison ?

L'expérience de sélection *in vivo* de terminateurs à partir de séquences aléatoires a aussi mené à l'identification d'un second motif impliqué dans la terminaison de la transcription. De la même façon que pour Reb1, cette séquence correspond au site de fixation d'un facteur de transcription : Rap1.

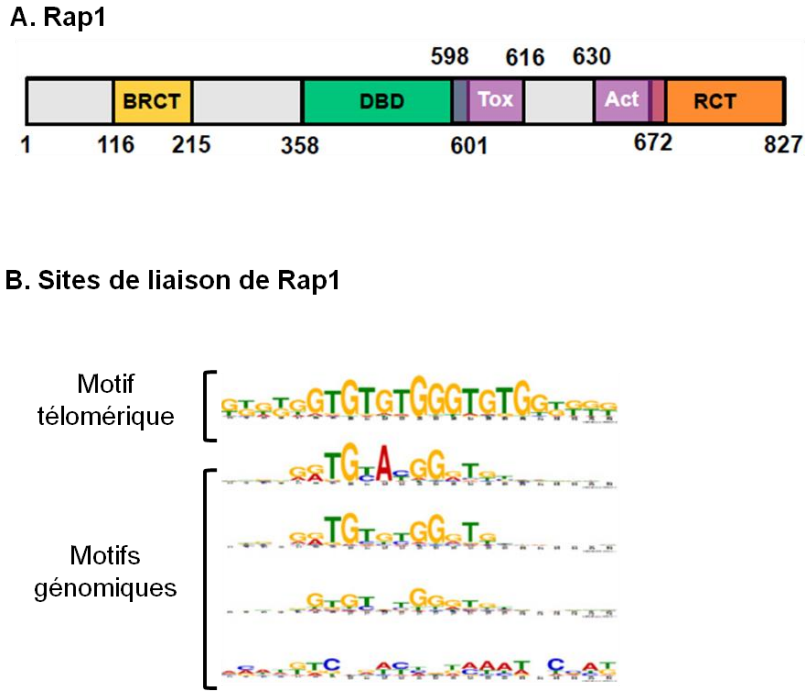

Figure 1.17. La protéine Rap1

A. Représentation schématique des domaines fonctionnels de Rap1. BRCT : domaine C-terminal de BRCA1, DBD : domaine de liaison à l'ADN, Tox : région de toxicité, Act : domaine d'activation, RCT : Domaine C-terminal de Rap1. Figure issue de Feldmann et Galletto, 2014.
B. Représentation des sites de fixation de Rap1 identifiés par ChIP-exo. Le motif télomérique a été obtenu par l'alignement de 73 sites de fixation identifiés sur des séquences télomériques et les trois types de motifs génomiques représentent les séquences obtenues par l'alignement de 503 sites de liaison identifiés sur le reste du génome. Figure adaptée de Rhee et Pugh, 2011.

Rap1 (repressor-activator protein 1) est un régulateur essentiel de la transcription et de la structure des télomères chez *S. cerevisiae* (Shore and Nasmyth, 1987). Rap1 est une protéine abondante, environ 4400 molécules/cellules (Ghaemmaghami et al., 2003), longue de 827 acides aminés qui peut être décomposée en trois principaux domaines : la région N-terminale contenant un domaine BRCT, la région centrale possédant un double domaine de liaison à l'ADN Myb-like et le domaine C-terminal

nécessaire pour le recrutement de facteurs associés à Rap1. En dehors de ces trois domaines, 40% de la chaîne peptidique correspond à des régions non structurées (Matot et al., 2012; figure 1.17.A.).
Les 279 premiers acides aminés contenant le domaine BRCT peuvent être supprimés sans montrer de phénotype particulier bien que cette région semble être requise pour l'interaction avec le facteur de transcription Gcr1 (Mizuno et al., 2004).
La région située entre le domaine de liaison à l'ADN et le domaine C-terminal contient une région « toxique » qui inhibe la croissance cellulaire lorsque Rap1 est surexprimée (Freeman et al., 1995). La région C-terminale est directement impliquée dans les fonctions régulatrices de Rap1. Tout d'abord le domaine situé entre les acides aminés 630 et 675 est nécessaire pour l'activation de la transcription par Rap1 (Hardy et al., 1992a), puis le reste de la région C-terminale est requis pour le contrôle de l'élongation des télomères (Kyrion et al., 1992) par l'interaction avec Rif1 et Rif2 (Hardy et al., 1992b; Wotton and Shore, 1997) et de la compaction de la chromatine (Buck and Shore, 1995; Kyrion et al., 1993) en recrutant directement Sir3 et Sir4 (Moretti et al., 1994).

Rap1, trouvé au niveau de promoteurs, de silencers et des télomères (revue : Piña et al., 2003). Il recrute des facteurs de remodelage de la chromatine responsables de la répression ou de l'activation de la transcription de ses cibles (Ganapathi et al., 2011). Dans un premier temps, il a été estimé que Rap1 ciblerait 5% des gènes (Lieb et al., 2001) mais cette approximation peut être augmentée puisqu'une étude plus récente a identifié près de 500 sites liés par Rap1 en dehors des régions télomériques (Rhee and Pugh, 2011). Comme son nom, l'indique Rap1 peut agir alternativement comme un activateur ou un répresseur de la transcription. Rap1 a été identifié comme un répresseur aux *silent mating type loci HML* et *HMR* (Kurtz and Shore, 1991).
Rap1 assure sa fonction de répresseur de la transcription en favorisant la formation d'hétérochromatine par le recrutement du complexe de déacétylation des histones Sir2-Sir3-Sir4 (Sussel and Shore, 1991; Moretti et al., 1994; Rusché et al., 2002) à la fois au niveau des *silent mating type loci HML* et *HMR* mais aussi au niveau des télomères où il se fixe au niveau des séquences d'ADN répétées télomériques.

Rap1 est également l'activateur de la transcription de nombreux gènes participant à la glycolyse et d'environ 90% des gènes codants des protéines ribosomales

(Lascaris et al., 1999; Lieb et al., 2001). Ces gènes étant très exprimés lors de la croissance des levures, Rap1 participe à l'activation d'environ 37% de tous les évènements d'initiation de la transcription par l'ARN pol II dans des levures en phase exponentielle de croissance (Lieb et al., 2001).

Rap1 est capable d'activer la transcription par des mécanismes variés tels que l'interaction directe avec des sous-unités du PIC, le recrutement de complexe de remodelage de la chromatine et la facilitation de la fixation de facteurs de transcription secondaires.

Rap1 facilite probablement le recrutement de l'ARN pol II au niveau des promoteurs de gènes codant des protéines ribosomales en interagissant directement avec Taf4, Taf5 et Taf12, sous-unités de TFIID, membre du PIC (Garbett et al., 2007). L'activation de ces gènes est aussi dépendante des facteurs de transcription Fhl1 et Ith1 dont le recrutement pourrait être favorisé par Rap1 (Schawalder et al., 2004; Gordân et al., 2009). Par ailleurs, Rap1 est aussi capable de faciliter le recrutement du complexe Gcr1-Gcr2, un activateur transcriptionnel des gènes impliqués dans la glycolyse (Deminoff and Santangelo, 2001; Zeng et al., 1997).

En tant qu'activateur de la transcription, Rap1 semble aussi être capable de recruter ou de faciliter l'activité des complexes de remodelage de la chromatine tel que le complexe SWI/SNF, un complexe capable de désassembler les nucléosomes (Idrissi et al., 2001) et le complexe histone acétylase Esa1 au niveau des gènes de protéines ribosomales (Reid et al., 2000).

Il est intéressant de noter que les sites de fixation de Rap1 sont souvent trouvés près d'autres sites de fixation de facteurs de transcription probablement parce que Rap1, en recrutant des facteurs de remodelage de la chromatine, facilite l'accessibilité de ces sites (Gordân et al., 2009; de Boer and Hughes, 2012). Il a notamment été montré que l'activateur transcriptionnel Gcn4 a besoin de Rap1 pour accéder au promoteur du gène *HIS4* (Yu and Morse, 1999).

Rap1 est potentiellement capable de se lier aux télomères à une densité d'une protéine toutes les 18pb (Gilson et al., 1993; Williams et al., 2010). Au niveau des télomères, Rap1 contrôle à la fois la compaction de la chromatine, par le recrutement du complexe Sir, et limite le nombre de répétitions télomériques (Sussel and Shore, 1991; Kyrion et al., 1993) par le recrutement des protéines Rif1 et Rif2 capables de

« compter » et de déterminer le nombre final de répétitions (Hardy et al., 1992a; Wotton and Shore, 1997; Marcand et al., 1997).

Rap1 cumule donc les fonctions antagonistes d'activateur et de répresseur de la transcription. La sélection du mode de régulation semble être liée à la composition nucléotidique de ses sites de fixation. Les sites de fixation télomèriques et les sites activateurs de la transcription présentent quelques différences au niveau de leurs séquences ce qui pourrait avoir un lien avec leurs effets opposés sur l'activation de la transcription (Rhee and Pugh, 2011; Idrissi et al., 2001; figure 1.17.B.). Par exemple, dans un promoteur, lorsqu'on remplace le site de fixation de Rap1 activateur de la transcription par un site de fixation télomérique, la transcription diminue de deux tiers (Idrissi et al., 2001).
Rap1 pourrait interagir de façon différente sur ces deux types de site. Le domaine de liaison à l'ADN de Rap1 est constitué de deux domaines Myb en tandem qui reconnaissent la séquence ACAYYY (Y : C ou T) (König et al., 1996). Les motifs de fixation de Rap1 sont donc constitués de deux « demi-sites » ACAYYY espacés de trois nucléotides (voir figure 1.17.B., la séquence des motifs est dans l'orientation réverse-complémentaire). Ces deux demi-sites sont très conservés dans les motifs télomériques alors que dans les motifs génomiques le deuxième demi-site est moins conservé que le premier. En liant les motifs télomériques ou génomiques Rap1 pourrait adopter deux structures distinctes et ainsi exposer des surfaces d'interaction spécifiques à des facteurs associés à l'activation, à la répression ou au contrôle de la taille des télomères.

Au cours de ma thèse j'ai étudié et caractérisé une nouvelle fonction de Rap1 dans la terminaison de la transcription. Tout comme le facteur Reb1, Rap1 est capable de bloquer la progression de l'ARN pol II et de provoquer la terminaison de la transcription. Rap1 bloque la transcription et pourrait ainsi avoir une double fonction de remodeleur de la chromatine et de protecteur des NDR contre la transcription envahissante.

II. Travaux de thèse

Mon travail de thèse a concerné deux volets : d'une part j'ai travaillé sur la caractérisation des séquences nécessaires à la terminaison par la voie NNS, et d'autre part j'ai étudié la terminaison « road block » et le rôle du facteur Rap1.

II.1. Etude de séquences nécessaires à la terminaison par la voie NNS

II.1.1. Introduction

Analyse du CUT du locus *IMD2* par mutagénèse aléatoire

Les séquences permettant la liaison de Nrd1 et Nab3 à l'ARN naissant sont relativement bien connues. Cependant la présence de ces sites (ou leur absence) ne permet pas à définir de façon univoque si une région particulière est un terminateur ou pas. En effet, si des sites de fixation des facteurs Nrd1 et Nab3 sont requis pour la terminaison, leur nombre varie beaucoup d'un terminateur à l'autre. Par exemple le CUT *NEL025C* contient 15 sites de fixation de Nrd1 et Nab3 (la plupart contribuant à la terminaison) alors que le CUT du locus *IMD2* n'en contient que 3.
Ces deux terminateurs étant efficaces, on peut formuler différentes hypothèses pour expliquer la grande variation du nombre de sites Nrd1 et Nab3. Par exemple on peut imaginer que des motifs additionnels rendent la fixation de ces deux protéines plus efficaces sur le CUT *IMD2* ou bien qu'ils favorisent la terminaison pour d'autres raisons en compensant l'absence de plusieurs sites Nrd1 et Nab3.

J'ai donc entrepris de déterminer les éléments de séquence du CUT *IMD2* requis pour la terminaison, dans le but de définir les sites de liaison optimaux de Nrd1 et Nab3 et/ou d'identifier de nouvelles séquences impliquées dans la terminaison. Cette expérience ayant été menée en 2010, la composition des motifs de terminaison du NNS n'était pas aussi précise qu'actuellement (voir Table 1, chapitre I.2.4.2.). Les sites de fixation connus étaient, GUA(A/G) pour Nrd1 et UCUU pour Nab3 (Carroll et al., 2004).

L'organisation du locus *IMD2* (Figure 2.1.1) ressemble à celui du locus *URA2* (chapitre I.3.3.1.). *IMD2* code une enzyme de la voie de biosynthèse du GTP et possède un promoteur qui permet l'initiation de la transcription à deux sites distincts (Kuehner and Brow, 2008). Quand la transcription démarre à partir de sites d'initiation proches du promoteur, correspondant à une série de G, la polymérase rencontre des sites de terminaison par le NNS, ce qui conduit à la production d'un CUT ; lorsque l'initiation a lieu au niveau de sites situés plus d'une centaine de paires de base en aval, les signaux de terminaison précoces sont évités, ce qui permet la synthèse de l'ARNm d'*IMD2*. Le choix des sites d'initiation dépend de l'abondance en GTP. Lorsque le GTP n'est pas limitant la transcription démarre au niveau des sites proximaux et seul le CUT est transcrit. En carence de GTP la transcription démarre en aval, ce qui conduit à la synthèse de l'ARNm d'*IMD2*, le CUT n'est plus synthétisé et la transcription est initiée au niveau du TSS de l'ARNm (Kuehner and Brow, 2008, figure 2.1.1.).

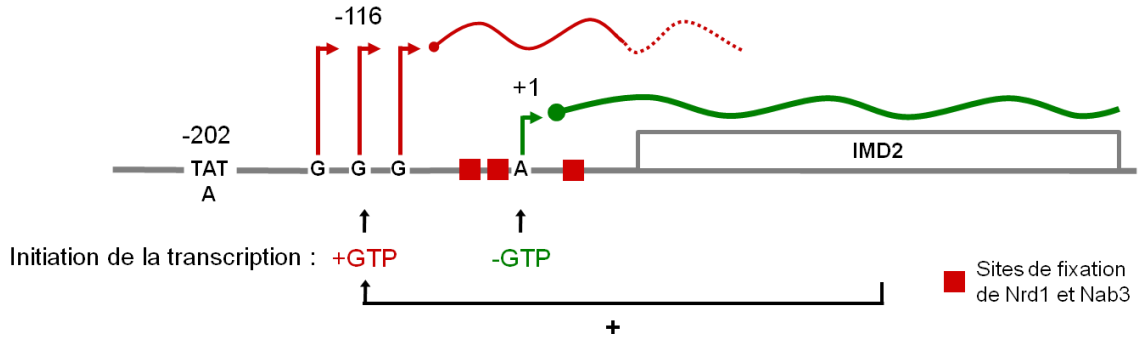

Figure 2.1.1. Modèle de régulation du gène *IMD2*
IMD2 code l'IMP déshydrogénase, une enzyme impliquée dans la biosynthèse du GTP.
Le promoteur d'*IMD2* contrôle la transcription de deux régions d'initiation.
En présence d'une forte concentration de GTP, les premiers sites constitués de G sont utilisés (flèches rouges) et permettent la synthèse d'un CUT (en rouge) terminé par le complexe NNS.
Lors d'une carence en GTP, les sites d'initiation riches en G sont ignorés et la transcription est initiée sur un A plus distal (flèche verte), celui-ci permet la transcription de l'ARNm *IMD2* (en vert). La transcription d'*IMD2* est contrôlée par une boucle de régulation négative : l'expression d'*IMD2* favorise la synthèse de GTP et l'augmentation de la quantité de GTP favorise la transcription du CUT par rapport à l'ARNm.
Figure adaptée de Kuehner et Brow, 2008.

II.1.2. Résultats et discussion

Afin de déterminer les éléments de séquence nécessaires à la terminaison du CUT *IMD2*, j'ai inséré des séquences sauvages ou mutées de ce CUT dans un système rapporteur permettant d'observer des défauts de terminaison (figure 2.1.2.).

II.1.2.1. Le système rapporteur

Ce système constitué de deux unités de transcription en tandem est basé sur l'interférence transcriptionnelle, c'est-à-dire l'inactivation d'un promoteur lorsque ce dernier est « traversé » par des polymérases transcrivant à partir de régions environnantes. Le gène *LacZ,* qui code la β-galactosidase (β-gal) est contrôlé par le promoteur du gène *GAL1.* En amont, Un promoteur répressible par la tétracycline (Tet) contrôle la transcription d'une séquence insérée (figure 2.1.2.).

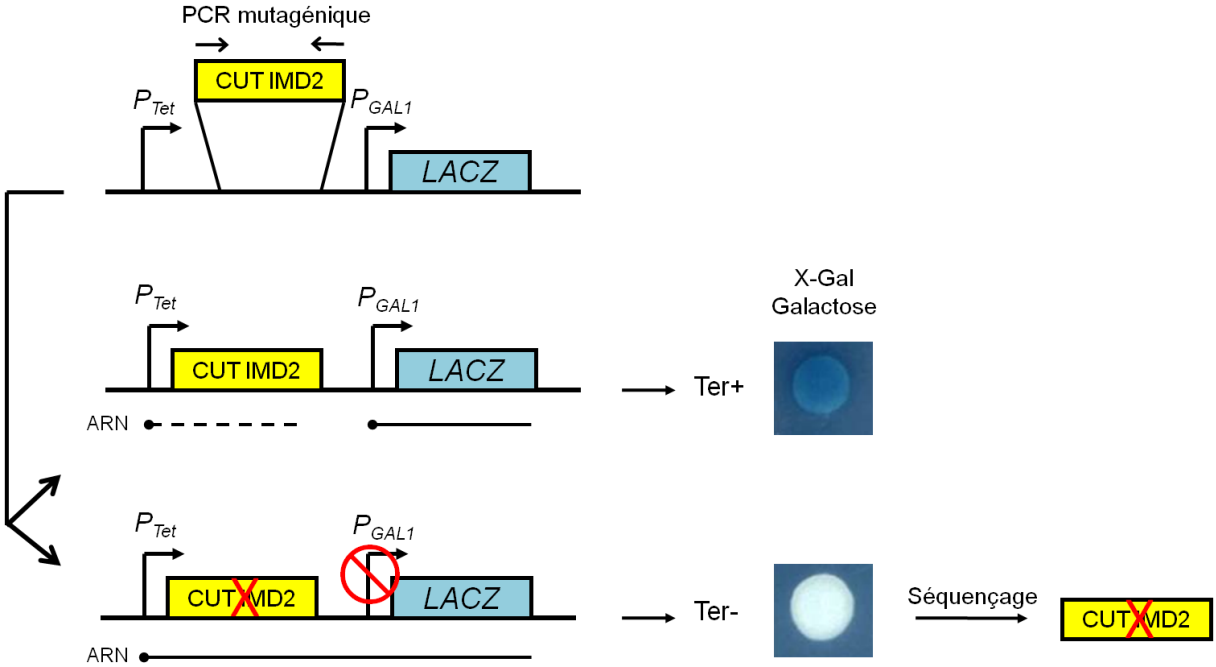

Figure 2.1.2. Procédure expérimentale pour tester l'efficacité de la terminaison par le CUT du locus *IMD2* muté.

Le CUT du locus *IMD2* amplifié par PCR mutagénique est inséré entre un promoteur répressible par la tétracycline (P_{Tet}) et le promoteur du gène *GAL1* (P_{GAL1}). Si le CUT inséré permet la terminaison (Ter+), LACZ est transcrit, la ß-gal est synthétisée et les colonies sont bleues en présence de X-Gal et de galactose. Si les mutations du CUT inséré (croix rouge) altèrent la terminaison (Ter-) la transcription de LACZ est bloquée par interférence transcriptionnelle et les colonies restent blanches en présence de X-Gal. Les clones blancs sont sélectionnés et leurs inserts mutés sont séquencés.

Si la séquence insérée ne permet pas la terminaison, la transcription démarre au niveau du promoteur Tet, traverse l'insert et le promoteur *GAL1*, qui est en conséquence inactivé. Dans ces conditions la β-galactosidase n'est pas produite et la cellule sera blanche dans un milieu contenant du X-Gal.
Si la séquence insérée entre les promoteurs *Tet* et *GAL1* induit la terminaison, la transcription initiée au niveau du promoteur Tet est terminée avant d'atteindre le promoteur *GAL1*. Dans ce cas *LacZ* est transcrit normalement, la β-gal est synthétisée et la cellule sera bleue dans un milieu X-Gal.

Dans le but d'identifier les motifs responsables de la terminaison du CUT *IMD2*, j'ai inséré dans le système rapporteur une série de variantes nucléotidiques du CUT *IMD2 (*320pb) obtenues en mutant le CUT de façon aléatoire par PCR en condition mutagénique. Lorsque le CUT sauvage est inséré, sa séquence assure la terminaison de la transcription et les colonies sont bleues sur milieu gélosé en présence de X-Gal. Lorsque les éléments importants pour la terminaison sont mutés, les cellules sont blanches. J'ai donc sélectionné les colonies blanches en présence de X-Gal, correspondent à l'insertion d'un CUT *IMD2* muté et incapable de terminer la transcription et 20 clones ont été ensuite séquencés.

II.1.2.2. Identification de 5 motifs potentiellement impliqués dans la terminaison du CUT du locus IMD2

Chacune de ces séquences présentait de 3 à 19 mutations, ce qui rendait difficile la discrimination entre les mutations d'intérêt et le bruit de fond de la mutagénèse.
Pour identifier les motifs impliqués dans la terminaison nous sommes partis du principe que ces motifs devraient avoir une fréquence de mutation plus élevée dans l'ensemble des 20 CUT sélectionnés que le reste de la séquence, dont la fréquence de mutation serait le « bruit de fond » de la mutagénèse.
Pour déterminer les régions qui ont une fréquence de mutation élevée j'ai aligné l'ensemble de ces CUT mutés. Cela m'a permis d'identifier 6 « points chauds » de mutation (figure 2.1.3.). Les deux pics de mutation les plus élevés correspondent aux séquences consensus d'un site de fixation de Nrd1 et d'un site Nab3. Ces deux sites correspondant à des motifs de terminaison connus ce résultat valide la pertinence de

l'approche utilisée. Parmi les 4 autres séquences fréquemment mutées on trouve un deuxième site de fixation de Nab3 et deux régions constituées de A et T.

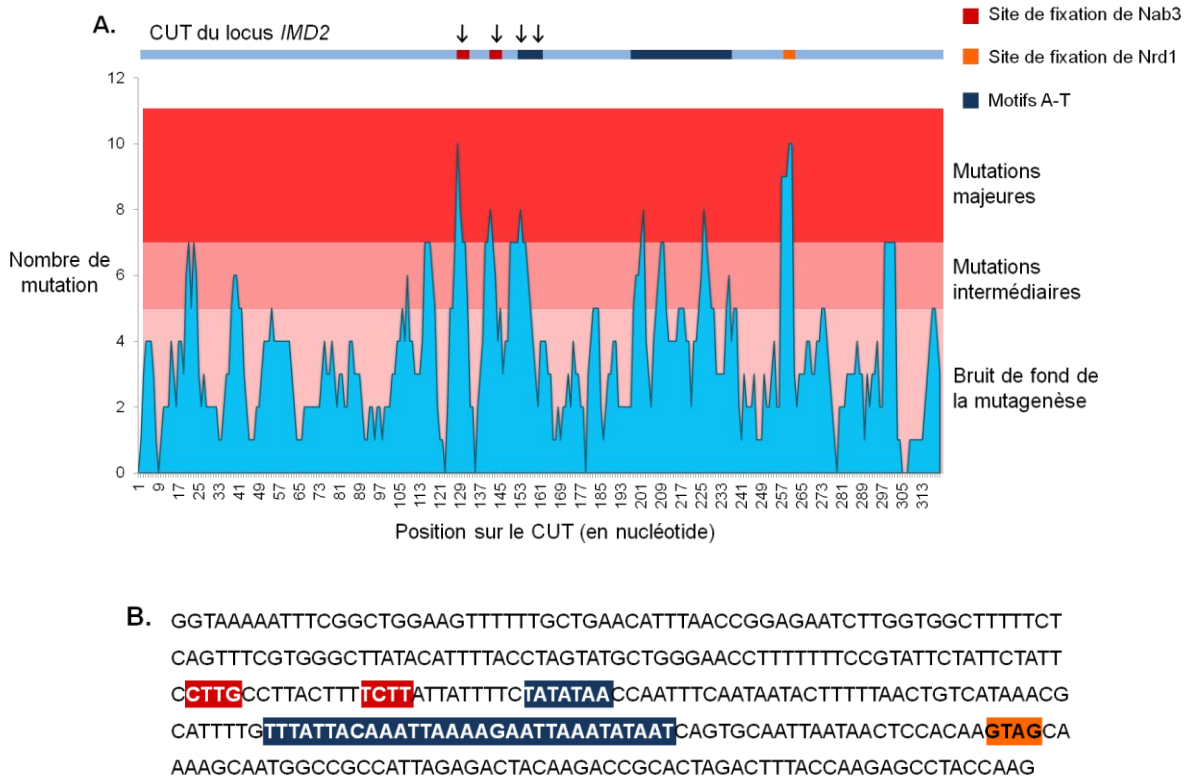

Figure 2.1.3. Résultats de la recherche des motifs de terminaison du CUT du locus *IMD2*.

A. Vingt séquences ont été sélectionnées, séquencées et alignées. Le nombre de mutations dans une fenêtre de 5 nucléotides est représenté en fonction de la position sur le CUT. Les mutations déjà connues pour affecter la terminaison de la transcription sont indiquées par des flèches (Kuehner et Brow, 2008).
B. Séquence du CUT du locus *IMD2*. Les motifs de fixation de Nrd1 et Nab3 ainsi que les motifs riches en A-T sont surlignés en respectant le code couleur de la figure précédente.

L'un des deux sites Nab3 identifié n'est pas un site canonique car il présente la séquence CTTG (figure 2.1.3.B.). Des résultats récents (Creamer et al., 2011; Porrua et al., 2012; Wlotzka et al., 2011) ont montré l'importance du dernier G et confirmé que CTTG est aussi un site efficace de fixation de Nab3.

Lorsque j'ai effectué ces travaux entre 2010 et 2011, la séquence A-T n'avait pas encore été identifiée comme un élément participant à la terminaison par le NNS. Plus tard d'autres travaux du laboratoire ont retrouvé un motif identique (TATATAA) dans un terminateur synthétique dépendant du NNS contenant également deux sites Nab3

et un site Nrd1 (Porrua et al., 2012). Le rôle de ce motif était mystérieux puisqu'il ne ressemble pas aux sites de fixation de Nrd1 et Nab3 et qu'il est plutôt semblable aux séquences A-U responsable du recrutement du complexe CPF-CF (chapitre I.2.4.1.). En effet, la séquence UA(C/U)(A/G)UA correspond à l'élément de positionnement qui est reconnu par la sous-unité du CPF-CF Hrp1. Cependant il a été montré sur le CUT du locus *IMD2* et sur le terminateur synthétique que le motif A-T (A-U sur l'ARN) n'est pas reconnu par des sous-unités du CPF-CF. En effet, la mutation de *HRP1* n'a pas d'effet sur la terminaison du CUT du locus *IMD2* (Kuehner and Brow, 2008) et la déplétion de Hrp1 n'affecte pas la terminaison au niveau du terminateur synthétique contenant le motif A-T (Porrua et al., 2012). Une autre sous-unité du CPF-CF, Rna15, est également capable de se lier à un motif riche en A et U, l'élément de positionnement. Il a été montré que l'inactivation de Rna15 n'a pas non plus d'effet sur le terminateur synthétique (Porrua et al., 2012).

Ce motif A-T semble plutôt participer au recrutement du NNS puisqu'il favorise, *in vitro*, la fixation du dimère Nrd1-Nab3 sur un transcrit contenant le terminateur synthétique (Porrua et al., 2012). Cependant il n'est pas exclu qu'*in vivo* cette séquence soit reconnue par un autre facteur participant à la voie de terminaison par le NNS.

II.1.2.3. Limites de l'expérience et perspectives

Comme présenté sur la figure 2.1.3. cette méthode m'a permis d'identifier cinq régions particulièrement mutées. Elles atteignent toutes les cinq le palier « mutations majeures ». Ce palier a été déterminé arbitrairement pour qu'il contienne tous les pics de mutation correspondant aux régions dont l'implication dans la terminaison était déjà connue (les deux sites Nab3). Par déduction, il est vraisemblable que les régions qui présentent un niveau de mutation au moins égal soient elles aussi impliquées dans la terminaison, comme ceci a été vérifié pour la région A-T. Il est aussi envisageable que certaines régions moins mutées, notamment celles du palier « mutations intermédiaires », soient aussi requises pour la terminaison.

Deux solutions sont concevables pour affiner cette analyse. D'une part, on pourrait utiliser un échantillon plus grand de séquences sélectionnées ce qui permettrait d'établir des différences statistiquement significatives entre le niveau de mutation des régions impliquées dans la terminaison et le bruit de fond de la mutagenèse.

Un second levier pour améliorer la résolution de cette expérience pourrait être la diminution du taux de la mutagenèse (et donc du bruit de fond). Ceci est envisageable, mais ça impliquerait d'augmenter de façon significative le nombre de séquences à analyser.

La fonction des régions identifiées dans ce crible reste à valider. Il conviendrait de confirmer l'implication dans la terminaison de ces motifs identifiés par mutagenèse ciblée du CUT du locus *IMD2* et d'analyser la production d'ARN instables dérivés de ces séquences mutées. Concernant les deux motifs A-U, il pourrait être intéressant de tester leur rôle dans le recrutement du dimère Nrd1-Nab3. Pour cela des expériences de fixation *in vitro* pourraient être réalisées entre le dimère Nrd1-Nab3 et des fragments du CUT contenant des combinaisons variées des cinq motifs identifiés, mutés ou non. A cause de mon implication dans d'autres projets, et de la publication d'études adressant l'importance de certains des motifs identifiés, ces expériences n'ont pas été poursuivies.

II.2. Analyse des sites de fixation du dimère Nrd1-Nab3 sur l'ARN par SELEX

II.2.1. Introduction

Comme détaillé précédemment, il existe trois motifs requis pour la terminaison par le complexe Nrd1-Nab3-Sen1. Lorsque j'ai démarré ce travail, la fonction du motif A-U n'était pas clairement définie. Dans une certaine mesure il n'est toujours pas clair si ce motif contribue spécifiquement à la liaison du dimère Nrd1-Nab3, à la liaison d'autres facteurs ou à d'autres étapes de la terminaison. Il a été montré qu'*in vitro* ce motif favorise la liaison à l'ARN du dimère Nrd1-Nab3, et que sa fonction dans la terminaison *in vivo* n'est pas liée à sa reconnaissance par Hrp1 ou Rna15, deux sous-unités du CPF-CF capables de se fixer à des motifs similaires (Porrua et al., 2012). Pour isoler les déterminants de liaison du complexe NNS des autres déterminants de la terminaison nous avons voulu analyser plus en détail la structure d'un site de fixation optimal du dimère Nrd1-Nab3. J'ai donc effectué une expérience de SELEX (systematic evolution of ligands by exponential enrichment). Une question importante adressée par cette approche a été de déterminer si le motif riche en A-U est retrouvé dans ce site optimal.

II.2.2. Résultats

II.2.2.1. Le SELEX Nrd1-Nab3

Le SELEX est une technique très efficace pour identifier les meilleurs sites de fixation d'une protéine sur l'ARN. Elle est basée sur la sélection des molécules d'ARN contenant un site de liaison pour une protéine donnée à partir d'une très large population de séquences différentes. Les complexes protéine-ARN sont séparés des protéines ou ARN isolés grâce à des techniques variées. Les ARN associés aux complexes sont purifiés, amplifiés et soumis à des cycles successifs de sélection. Ceci produit un enrichissement exponentiel de la population d'ARN capables de se lier à la protéine (généralement très minoritaire au début de l'expérience). L'expérience que j'ai menée compte 10 cycles d'enrichissement d'ARN se liant au dimère Nrd1-Nab3.

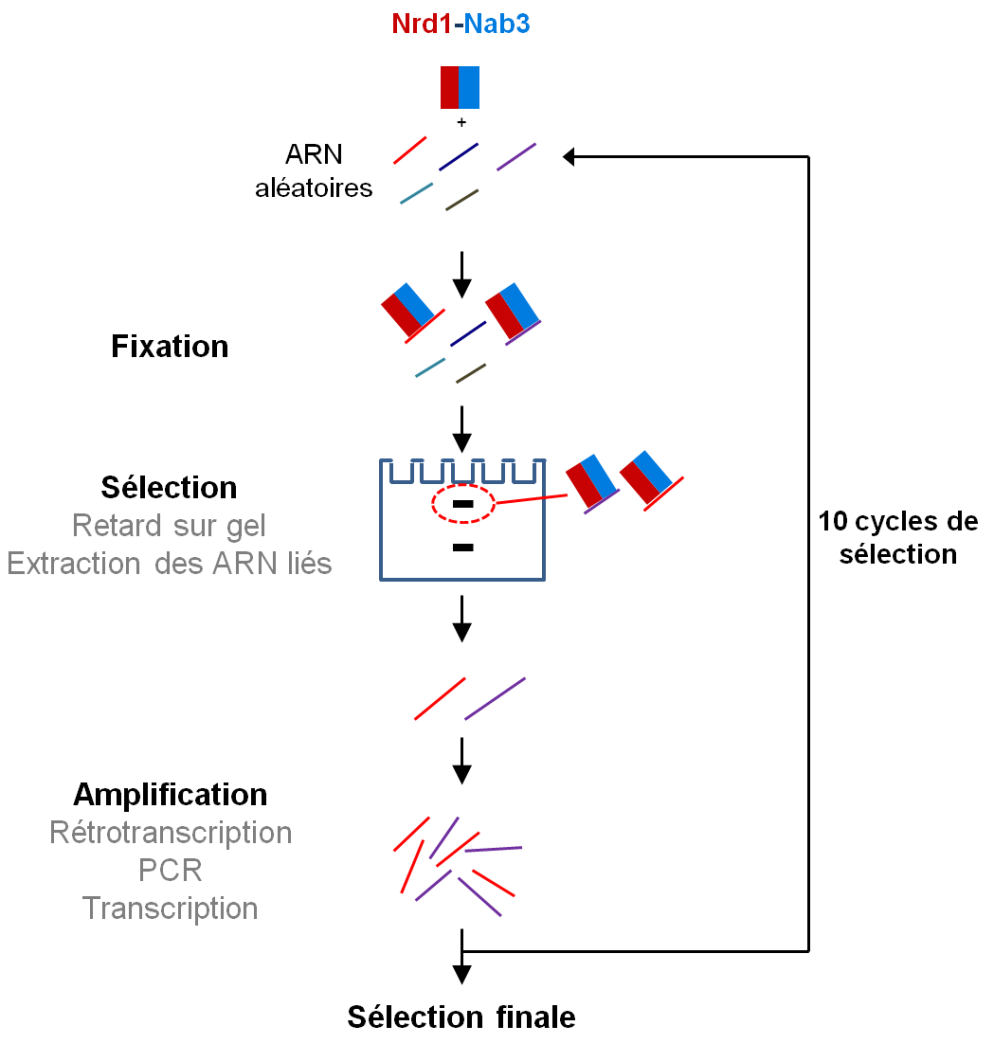

Figure 2.2.1. Procédure expérimentale pour identifier les sites de fixation du dimère Nrd1-Nab3 sur l'ARN par SELEX.
Fixation : Des ARN radiomarqués contenant 40 nucléotides aléatoires et deux extrémités constantes sont incubés avec le dimère Nrd1-Nab3 reconstitué à partir de protéines recombinantes.
Sélection : Les complexes ARN-Nrd1-Nab3 sont isolés par retard sur gel de polyacrylamide. Les ARN sont ensuite purifiés.
Amplification : Rétrotranscription des ARN purifiés, amplification par PCR des ADNc avec une amorce contenant le promoteur de l'ARN pol T7 puis transcription des ADNc amplifiés. Les ARN synthétisés sont soumis à un nouveau cycle de sélection.

Chaque cycle est composé de 5 étapes (figure 2.2.1.). La première étape consiste à mélanger le dimère Nrd1-Nab3 recombinant avec des ARN radiomarqués contenant une région de 40 nucléotides de séquence aléatoire flanquée de deux régions constantes. Au cours de cette étape Nrd1 et Nab3 se fixent aux séquences d'ARN qui correspondent à leurs sites de fixation.

La deuxième étape correspond à la séparation des ARN, des protéines libres et des complexes ARN-Nrd1-Nab3 par migration sur gel de polyacrylamide (EMSA,

electrophoretic mobility shift assay). Les complexes sont identifiés grâce au radiomarquage de l'ARN, et extraits du gel par électro-élution.

Les ARN sont ensuite purifiés à partir des complexes, reverse-transcrits en ADN complémentaire (ADNc) et amplifiés par PCR. Une des amorces utilisées pour la PCR contient un promoteur de l'ARN polymérase T7, ce qui permet d'obtenir un nouveau répertoire d'ARN à partir des ADNc amplifiés.

Les ARN ainsi sélectionnés et amplifiés sont ensuite soumis à un nouveau cycle d'enrichissement en présence de Nrd1 et Nab3. Au total, 10 cycles ont été réalisés et la progression de la sélection est observable par EMSA (figure 2.2.2.). On remarque clairement que la proportion d'ARN liés au dimère Nrd1-Nab3 augmente au fil des cycles ce qui signifie que l'expérience a fonctionné.

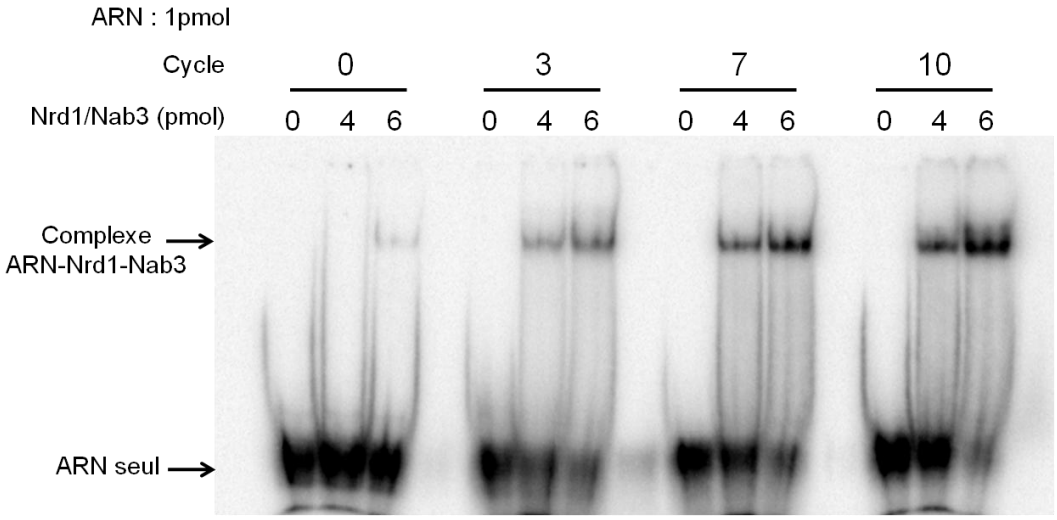

Figure 2.2.2. Autoradiographie d'un retard sur gel présentant l'évolution de l'affinité du dimère Nrd1-Nab3 pour les ARN sélectionnés au cours du SELEX.
Les ARN radiomarqués seuls ou fixés au complexe Nrd1/Nab3 sont indiqués par les flèches.

II.2.2.2. Séquençage et analyse des séquences sélectionnées

Les ARN sélectionnés ont été reverse-transcrits en ADNc et les répertoires issus des septième et dixième cycles de SELEX ont été séquencés en même temps que le répertoire initial pour les analyses statistiques. Les fragments d'ADN à séquencer étant courts (191pb, en comptant les extrémités adaptatrices) ils ont été séquencés à grande échelle par la technologie Illumina/Solexa permettant d'obtenir des centaines de milliers de séquences uniques pour chacun des groupes.

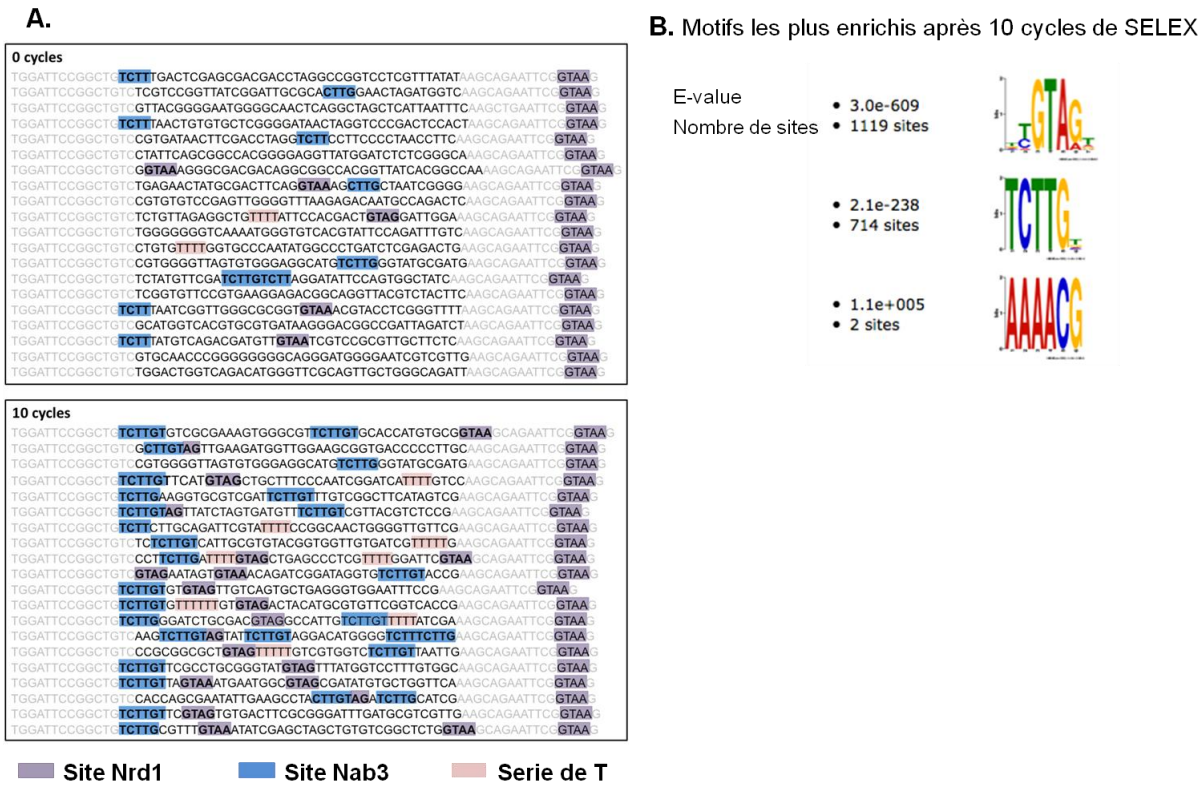

Figure 2.2.4. Analyse des séquences obtenues après 10 cycles de SELEX.

A. Echantillon de séquences obtenues avant et après 10 cycles de SELEX. Les extrémités constantes sont grisées et les 40 nucléotides aléatoires sont en noir. Les sites de liaison des protéines Nrd1 (violets) et Nab3 (bleus) sont clairement enrichis après 10 cycles de SELEX.

B. Analyse statistique des motifs les plus enrichis après 10 cycles de SELEX. Logos MEME (multiple expectation maximization for motif elicitation, Bailey and Elkan, 1994) représentant les motifs les plus enrichis parmi 1500 séquences obtenues après 10 cycles de SELEX et comparées à la composition des séquences initiales. Les motifs sont classés dans l'ordre décroissant de leur E-value (probabilité d'observer ce résultat par chance), représentant leur enrichissement significatif. Seuls les deux premiers motifs, correspondant aux sites de fixation de Nrd1 et Nab3, présentent un enrichissement significatif. Seuls les 40 nucléotides aléatoires ont été analysés.

L'analyse bioinformatique comparée des séquences initiales et de 1500 séquences sélectionnées après 10 cycles de SELEX montre qu'il existe deux motifs de 6 à 8 nucléotides fortement enrichis après sélection. Ils contiennent soit la séquence GUAG, le site de fixation de Nrd1, soit la séquence UCUUG, le site Nab3 (figure 2. 2. 4. B). Cela confirme les données déjà acquise et on remarque une préférence de Nrd1 pour le site GUAG par rapport à GUAA.

II.2.2.3. Les sites de fixation de Nrd1 et Nab3 ne sont pas positionnés de façon aléatoire.

Nous avons voulu déterminer si une organisation spécifique des sites de fixation de Nrd1 et Nab3 les uns par rapport aux autres pourrait favoriser la fixation du dimère Nrd1-Nab3. En effet, il est possible en que le dimère contacte l'ARN de trois façon différentes : par les protéines Nrd1 ou Nab3 seules ou par l'interaction concomitante des deux protéines. Dans ce dernier cas, une disposition préférentielle des sites Nrd1 et Nab3 est envisageable et, si elle existe, elle devrait être enrichie dans les séquences sélectionnées par rapport aux séquences initiales.

Pour analyser la disposition des sites Nrd1 et Nab3 les uns par rapport aux autres nous avons comparé les différentes distances qui séparent un site Nrd1 d'un site Nab3, et inversement. Si une distance spécifique entre ces sites facilite la fixation du dimère Nrd1-Nab3, on devrait la trouver enrichie dans les séquences sélectionnées. Si aucune organisation spécifique n'existe, aucun enrichissement pour une distance particulière ne devrait être observé.

La comparaison des différentes distances séparant les sites Nrd1 et Nab3 parmi les séquences sélectionnées après 10 cycles de SELEX (figure 2.2.5.) a été effectuée par Tito Candelli, un bio-informaticien du laboratoire. Pour la distance entre un site Nrd1 suivi d'un site Nab3 (courbe pointillée rouge) on n'observe pas d'enrichissement pour une distance particulière, suggérant qu'aucune disposition préférentielle n'existe dans ce cas. A l'inverse, de manière remarquable, une disposition contenant un site Nab3 séparé de 4 à 10 nucléotides d'un site Nrd1 est significativement préférée par rapport à d'autres possibilités. Le pic à la position 0 est très probablement un artefact de l'analyse. Le site ainsi formé, TCTTGATG, contient un site de fixation de Nab3 complet UCUUG, ce qui rend plus probable la présence d'un site Nrd1 une fois « fixé » le site Nab3.

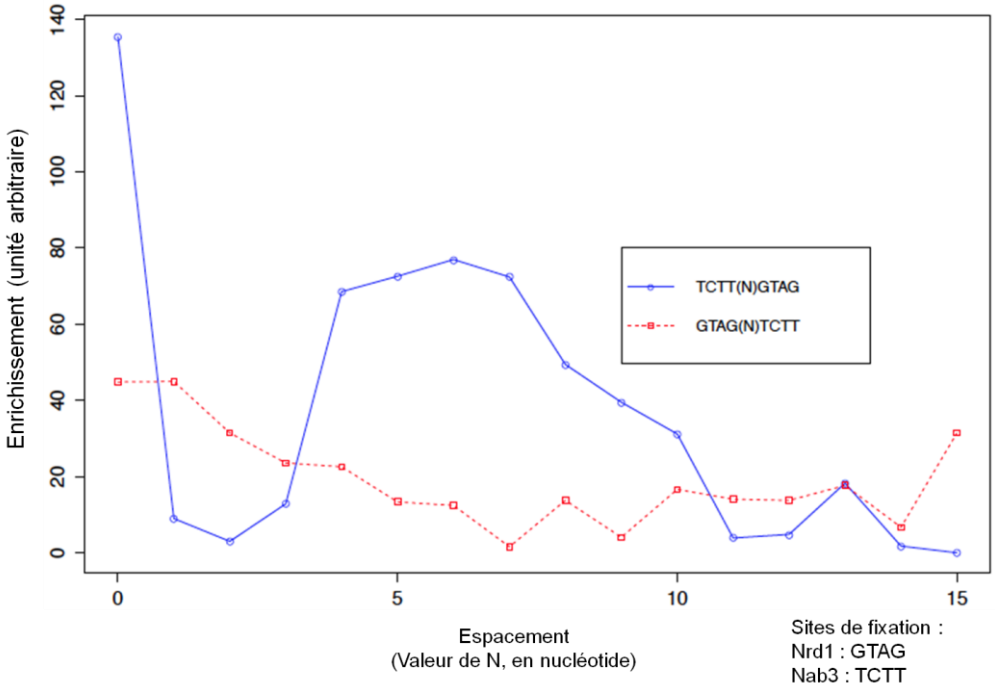

Figure 2.2.5. Analyse statistique de l'espacement entre les sites de fixation de Nrd1 et Nab3 dans les séquences obtenues après 10 cycles de SELEX.
La courbe bleue représente la distance, en nucléotides, entre un site Nab3 et un site Nrd1 en fonction de leur enrichissement dans les séquences obtenues après 10 cycles de SELEX. De la même manière, la courbe en trait discontinu rouge représente les distances observées entre un site Nrd1 et un site Nab3 après 10 cycles de SELEX. L'axe en ordonné représente l'abondance d'un certain motif dans les séquences issues du SELEX par rapport aux séquences initiales (voir Matériels et Méthodes).

Discussion

Le nombre de sites de fixation de Nrd1 et Nab3 est très variable, de 3 à 15 parmi les terminateurs NNS connus. Dans les terminateurs possédant peu de sites Nrd1 et Nab3 il est donc possible qu'une organisation optimale des sites de fixation de Nrd1 et Nab3 permette de contrebalancer leur faible nombre. Cette organisation optimale pourrait notamment permettre une fixation plus forte de ces deux facteurs sur leurs sites respectifs.

L'analyse statistique des motifs les plus enrichis après 10 cycles de SELEX montre que les deux seuls sites enrichis à l'issu du SELEX sont les sites Nrd1 et Nab3 (figure 2.2.4.B). Bien que quelques séries de T aient été identifiées dans l'échantillon de séquences sélectionnées après 10 cycles, aucun motif riche en A-T n'a été retrouvé statistiquement enrichi dans une population de séquence plus importante. Ce résultat semble suggérer que la séquence A-U ne favorise pas la fixation du

dimère Nrd1-Nab3 sur l'ARN, du moins dans les conditions de l'expérience. On peut donc imaginer qu'*in vivo* ce motif participe à la terminaison par un autre moyen, peut être par le recrutement d'un autre facteur ou par un effet sur l'élongation de la transcription.

L'analyse statistique de la distance entre les sites de fixation de Nrd1 et Nab3 dans les séquences sélectionnées (figure 2.2.5.) suggère qu'un site Nab3 suivi d'un site Nrd1 séparé par 4 à 10 nucléotides constitue le site préférentiel d'interaction du dimère, où les deux protéines pourraient interagir de façon coopérative. Cette organisation pourrait refléter la géométrie du complexe, avec Nab3 situé en amont et Nrd1 en aval sur l'ARN. Il faudrait cependant une étude structurale de la liaison du dimère Nrd1-Nab3 à l'ARN pour confirmer cette hypothèse. Il serait également intéressant de savoir si une telle organisation des sites Nab3 et Nrd1 est retrouvée de manière statistiquement enrichie au niveau des terminateurs naturels dépendants du NNS.

Il est aussi possible qu'une interaction plus forte du complexe Nrd1-Nab3 n'implique pas forcement que la terminaison de la transcription soit plus efficace, et il sera important de vérifier ceci avec des systèmes modèle *in vivo*. Ces expériences ont été prévues, mais il n'a pas été possible de les effectuer par manque de temps. Elles seront très probablement continuées au laboratoire.

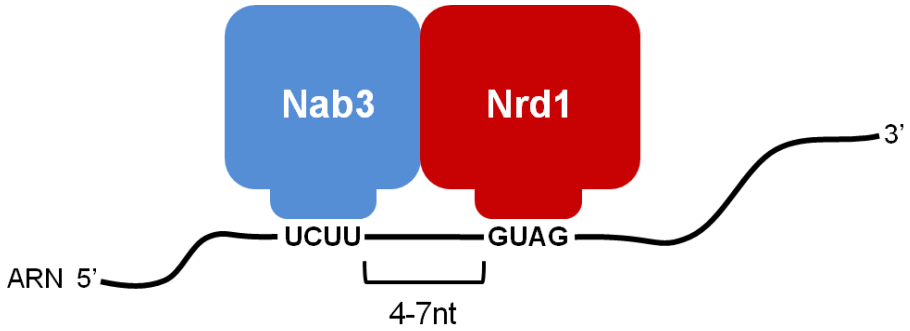

Figure 2.2.6. Modèle de fixation optimale du dimère Nab3-Nrd1 sur l'ARN.

II.3. Rap1 induit la terminaison de la transcription par un mécanisme de « road-block »

Comme détaillé dans l'introduction, la transcription par l'ARN pol II est terminée par deux voies principales chez *S. cerevisiae*. Ces deux voies font alternativement intervenir le complexe CPF-CF (voir chapitre I.2.4.1.) ou le complexe NNS (voir chapitre I.2.4.2.). Ces deux complexes reconnaissent des signaux spécifiques sur l'ARN naissant et interagissent avec le CTD de la grande sous-unité de l'ARN pol II, Rpb1, pour provoquer la libération de la polymérase de sa matrice ADN. Une troisième voie de terminaison qui a des caractéristiques différentes a été récemment décrite au laboratoire. Il a été montré que Reb1, un facteur de transcription important pour la formation des NDR, est capable de provoquer la terminaison de la transcription de l'ARN pol II par un mécanisme de road-block (voir chapitre I.3.5.1., Colin et al., 2014). Contrairement aux deux voies de terminaison canoniques la terminaison road-block n'est pas provoquée par des facteurs se fixant à l'ARN naissant mais fait intervenir un facteur de liaison à l'ADN capable de bloquer la progression de l'ARN pol II.

Les travaux présentés dans la suite de ce manuscrit montrent que le facteur Rap1 peut aussi terminer la transcription par un mécanisme analogue. Des résultats préliminaires suggèrent que cette caractéristique peut être étendue au facteur Abf1. Ces résultats donnent une dimension plus générale à la terminaison road-block dans le contrôle de qualité du transcriptome.

II.3.2. Résultats

Une expérience de sélection *in vivo* de terminateurs à partir de séquences aléatoires a été menée au laboratoire dans le but d'identifier de nouveaux motifs de terminaison de la transcription par l'ARN pol II. Comme détaillé précédemment, ce crible a permis d'identifier de nouveaux déterminants de séquence pour les voies de terminaison connues (Porrua et al., 2012) et a également mené à la découverte de la voie de terminaison road-block dépendante de Reb1 (Colin et al., 2014). Parmi les séquences issues de l'expérience de sélection de terminateurs, le site de fixation d'un deuxième facteur de transcription, Rap1, était aussi enrichi, suggérant que ce site puisse favoriser la terminaison de la transcription.

La caractérisation de la fonction de cette séquence et du mécanisme de terminaison par Rap1 a été le projet majeur de mon doctorat.

II.3.2.1. L'expérience de sélection de terminateurs à partir de séquences aléatoires

Pour réaliser la sélection de terminateurs *in vivo,* un système similaire à celui décrit dans le chapitre II.1.2.1. (figure 2.3.1.) a été utilisé. Un répertoire de fragments de séquence aléatoire a été inséré dans une région codante dérivée du gène *HSP104* située entre le promoteur répressible par la tetracycline (P_{TET}) et le promoteur du gène *GAL1*. Ce dernier contrôlant la transcription du gène *CUP1* dont l'expression est nécessaire pour la croissance de la levure en présence de cuivre.

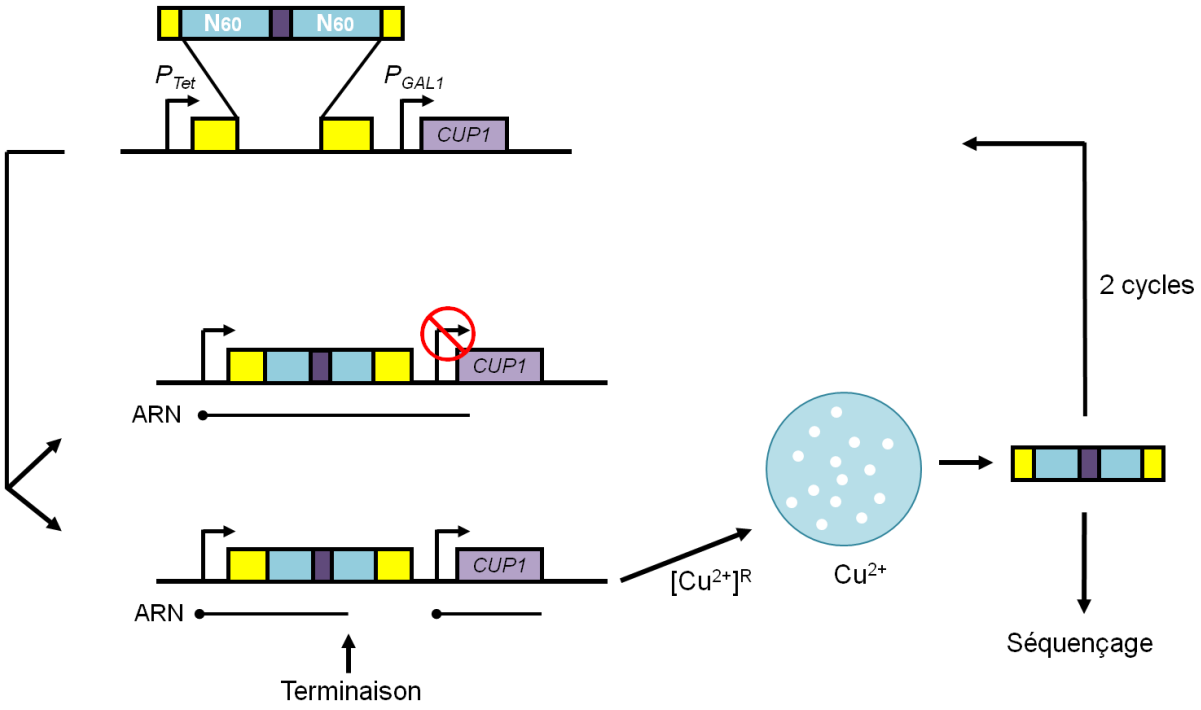

Figure 2.3.1. Procédure expérimentale pour sélectionner des terminateurs à partir de séquences aléatoires *in vivo*.
Une séquence contenant deux régions aléatoires de 60 nt (rectangles bleus clairs) flanquées de régions constantes est insérée dans une portion de l'ORF du gène *HSP104* (en jaune) située entre un promoteur répressible par la tétracycline (P_{Tet} TetO7) et le promoteur du gène *GAL1* (P_{GAL1}). Si la séquence aléatoire permet la terminaison, alors le gène *CUP1* est transcrit à partir du promoteur *GAL1*. Cela permet aux levures de croître en présence de cuivre. Après deux cycles de sélection, les inserts sélectionnés ont été séquencés.

Comme décrit au chapitre II.1.2.1., si la séquence insérée entre les deux promoteurs termine la transcription initiée par le promoteur Tet, le promoteur *GAL1* est actif, le

gène *CUP1* est transcrit et la levure sera résistante au cuivre. Ce système permet donc de sélectionner des terminateurs à partir de séquences aléatoires. Pour réduire l'impact des réarrangements du plasmide portant le système de sélection, les régions contenant la séquence aléatoire ont été amplifiés en vrac après un cycle de sélection et reclonés pour une deuxième étape de sélection (Porrua et al., 2012). Parmi les séquences obtenues, une large fraction provoquait la terminaison en faisant intervenir le complexe NNS (Porrua et al., 2012). Parmi les séquences restantes, environ 15% contiennent des motifs correspondant aux sites de liaison pour Reb1 et Rap1. La terminaison dépendante de Reb1 a été étudiée au laboratoire et j'ai effectué un travail parallèle sur la fonction de Rap1 dans la terminaison. La figure 2.3.2. décrit un logo représentant un motif retrouvé dans 7 séquences du répertoire sélectionné contenant des sites Rap1.

Figure 2.3.2. Motif trouvé parmi 7 terminateurs sélectionnés : site de fixation pour Rap1.
Logo MEME (Bailey and Elkan, 1994) d'un motif enrichi parmi les terminateurs sélectionnés. Ce motif correspond au site de fixation de la protéine Rap1 (voir chapitre I.3.5.2.).

Nous avons utilisé l'une des séquences sélectionnées contenant un site de fixation de Rap1 (figure 2.3.3. : X118) pour étudier plus en détail l'implication de cette protéine dans la terminaison. Nous avons montré que la résistance au cuivre est provoquée par la terminaison de la transcription puisqu'un transcrit court peut être détecté, dont l'extrémité 3' se situe au niveau de la séquence insérée. Tout comme les transcrits terminés par le complexe NNS (figure 2.3.3. : X9, terminateur dependant du complexe NNS), ce transcrit est instable dans une cellule sauvage et il est stabilisé lorsque *RRP6,* le gène codant une sous-unité catalytique de l'exosome nucléaire, est délété (figure 2.3.3.A). Cependant, à l'inverse des CUT, la terminaison induite par la séquence X118 ne dépend pas du complexe NNS car la déplétion de Nrd1 n'a pas d'effet (figure 2.3.3.B).

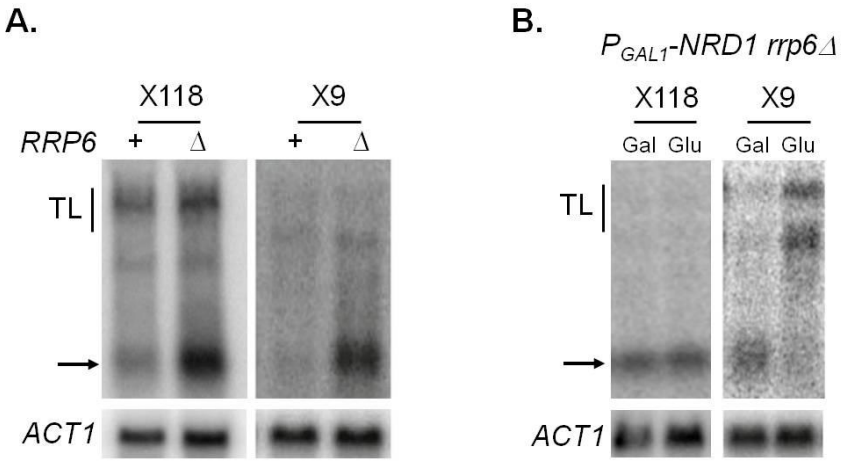

Figure 2.3.3. Analyse par northern blot des transcrits dérivés du clone X118 et du clone X9. Le terminateur du clone X118 contient un site Rap1 et X9 est un terminateur utilisé comme contrôle issu de la même expérience de sélection mais dépendant du complexe NNS (Porrua et al., 2012). Les constructions utilisées correspondent au système rapporteur tel qu'il est représenté figure 2.3.1. L'analyse a été effectuée sur des transcrits issus de souches portant une délétion de *RRP6* (A) ou permettant la déplétion métabolique de Nrd1 en présence de glucose dans une souche P_{GAL1}-*NRD1 rrp6Δ* (B). Les flèches indiquent les transcrits terminés au niveau des terminateurs sélectionnés, TL indique les transcrits issus de la translecture des terminateurs. La sonde utilisée est complémentaire de l'extrémité 5' de la séquence d'*HSP104* dans laquelle les terminateurs sont insérés. La même sonde est utilisée dans les figures suivantes pour détecter tous les ARN issus de cette construction.

II.3.2.2. Le facteur de transcription Rap1 est impliqué dans la terminaison de la transcription

Le site de liaison de Rap1 semblant se comporter comme un terminateur, j'ai voulu déterminer si le facteur Rap1 est impliqué dans la terminaison de la transcription. *RAP1* étant un gène essentiel, pour inactiver Rap1 j'ai utilisé une souche contenant l'allèle thermo-sensible *rap1-2*.

Comme cela est visible sur la figure 2.3.4., lorsque le facteur Rap1 est inactivé en cultivant le mutant *rap1-2* à 37°C (température restrictive), le niveau du transcrit court correspondant à la terminaison au niveau de la séquence contenant le site de liaison de Rap1 diminue fortement. Cette diminution est accompagnée par l'accumulation d'un transcrit plus long, dérivant d'une translecture du terminateur. Ces résultats indiquent que Rap1 est requis pour la terminaison de la transcription par l'ARN pol II au niveau de la séquence étudiée.

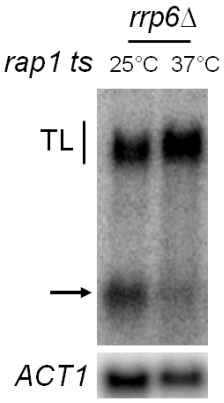

Figure 2.3.4. Rap1 est impliqué dans la terminaison du transcrit dérivé du clone X118.
Analyse par northern blot des transcrits dérivés du clone X118 lorsque *RAP1* est inactivé par l'utilisation d'un allèle thermosensible de *RAP1* à température restrictive.

II.3.2.3. Le site de liaison de Rap1 est nécessaire et suffisant pour la terminaison de la transcription

Afin de déterminer si le site de fixation de Rap1 est nécessaire pour la terminaison, je l'ai supprimé du terminateur sélectionné.

Comme le montre la figure 2.3.5.A, la séquence X118 sans le site de fixation de Rap1 n'induit pas terminaison de la transcription, démontrant que se site est nécessaire.

Pour établir si le site de liaison de Rap1 est suffisant pour la terminaison de la transcription, j'ai cloné deux versions du site de fixation de Rap1 dans le système rapporteur. L'un des motifs est une version réduite du site et l'autre est une version allongée en 5', correspondant au consensus complet du site de fixation de Rap1 allongée de 2nt.

On constate sur la figure 2.3.5.B que l'insertion des deux versions du site de fixation de Rap1 permet une résistance au cuivre, ce qui indique que les deux versions du site sont suffisantes pour terminer la transcription. Cependant, la résistance au cuivre est plus forte pour la version allongée du site de fixation de Rap1, indiquant que ce site provoque la terminaison de manière plus efficace, probablement grâce à une fixation plus forte de Rap1.

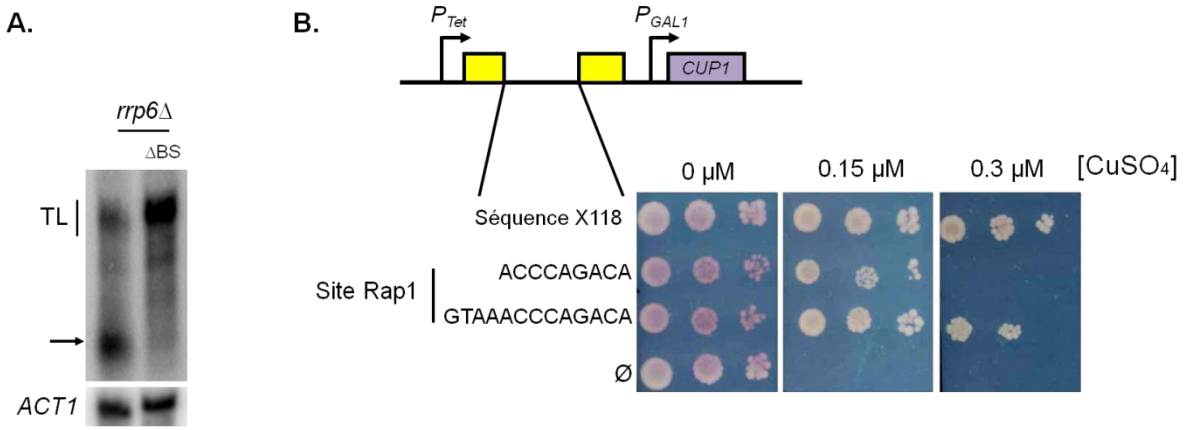

Figure 2.3.5. Le site de fixation de Rap1 est nécessaire et suffisant pour la terminaison.

A. Analyse par northern blot des transcrits dérivés du clone X118 après avoir supprimé le site de fixation de Rap1 du terminateur (ΔBS).
B. Analyse de l'efficacité de terminaison par la séquence X118 et le site Rap1 isolé par test de croissance. Le test de croissance a été effectué avec le système rapporteur utilisé lors de l'expérience de sélection de terminateur. Les effets de la séquence X118, de deux versions du site Rap1 issu de la séquence X118 et d'un témoin sans terminateur ont été testés.

II.3.2.4. La terminaison par Rap1 produit des ARN instables et polyadénylés par Trf4

La terminaison par Rap1 produit des transcrits instables, dégradés par l'exosome nucléaire. Dans le cas d'une terminaison par le complexe NNS, la dégradation est stimulée suite à la polyadénylation des transcrits par Trf4. J'ai donc voulu savoir si les ARN terminés par la séquence X118 suivent la même voie de dégradation que les CUT.

La séparation des fractions d'ARN polyadénylés et non polyadénylés montre qu'une fraction des ARN terminés au niveau de la séquence X118 sont polyadénylés et fortement stabilisés en absence de Rrp6 (figure 2.3.6., comparer WT A+ et *rrp6*Δ A+). Cette fraction polyadénylée disparaît lorsque le gène *TRF4* est délété. Cela montre que Trf4, probablement associée au complexe TRAMP, est responsable de la polyadénylation de ces ARN.

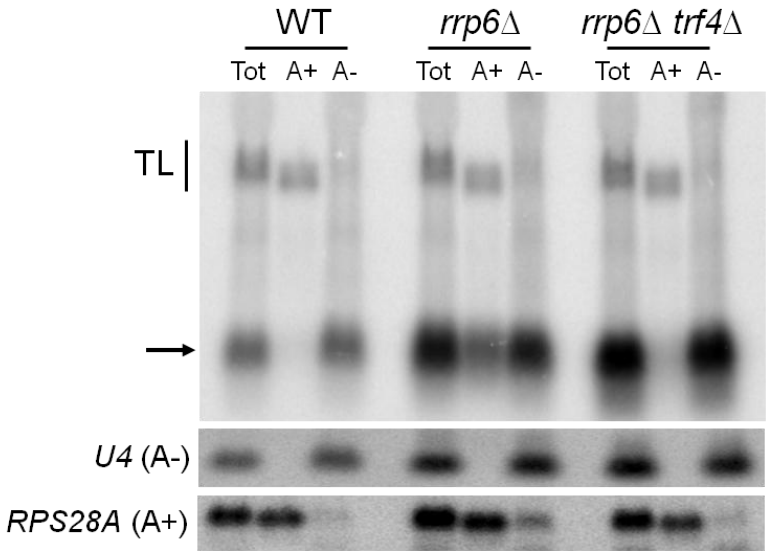

Figure 2.3.6. L'ARN terminé au niveau de X118 est polyadénylé par Trf4.
Analyse par northern blot des fractions d'ARN polyadénylés et non polyadénylés derivés du clone X118 dans des souches sauvage (WT), délétée du gène *RRP6* et délétée à la fois de *RRP6* et de *TRF4*. Les ARN totaux (Tot), polyadénylés (A+) et non polyadénylés (A-) extraits de chaque souche sont montrés. Deux ARN témoins sont utilisés, U4 est un ARN non polyadénylé et RPS28A est un ARN polyadénylé.

Des ARN non polyadénylés sont aussi détectés de façon majoritaire et sont aussi stabilisés en l'absence de *RRP6*, ce qui indique que la polyadénylation n'est pas absolument nécessaire pour la dégradation. Une fraction de ces ARN non adenylés est aussi observable dans une souche sauvage (figure 2.3.6., WT A-). La présence de ces ARN stables pourrait indiquer que la dégradation des formes non adénylées est moins efficace, ou bien il pourrait s'agir d'ARN naissants dont les extrémités 3', encore protégées par l'ARN pol II, ne sont pas accessibles à l'exosome (voir le mécanisme de terminaison, figure 2.3.17).

La terminaison de la transcription par Rap1 produit donc des ARN instables qui sont en partie polyadénylés par Trf4 et dégradés par l'exosome nucléaire.

II.3.2.5. Rap1 termine la transcription en se liant à l'ADN

Les deux voies canoniques de terminaison de la transcription par l'ARN pol II reposent sur la reconnaissance de motifs spécifiques et la fixation de complexes de terminaison (NNS ou CPF-CF) sur l'ARN naissant. Cependant Rap1 est une protéine de liaison à l'ADN, on peut donc imaginer un mécanisme de terminaison différent. A

l'image de Reb1, Rap1 pourrait bloquer l'avancée de l'ARN pol II en étant fixé à l'ADN.

Pour vérifier que Rap1 se fixe à l'ADN plutôt qu'à l'ARN, j'ai comparé l'affinité de Rap1 pour l'ADN et l'ARN. La figure 2.3.7. montre que la protéine recombinante Rap1 lie un ADN double brin contenant le site de fixation de Rap1 issu de la séquence X118 avec une affinité d'environ 60nM. Par contre, la fixation de Rap1 sur la version ADN simple brin ou ARN de cette même séquence est beaucoup plus faible sinon nulle. Pour ces deux molécules, le retard sur gel observable aux plus hautes concentrations de Rap1 correspond vraisemblablement à une fixation non spécifique.

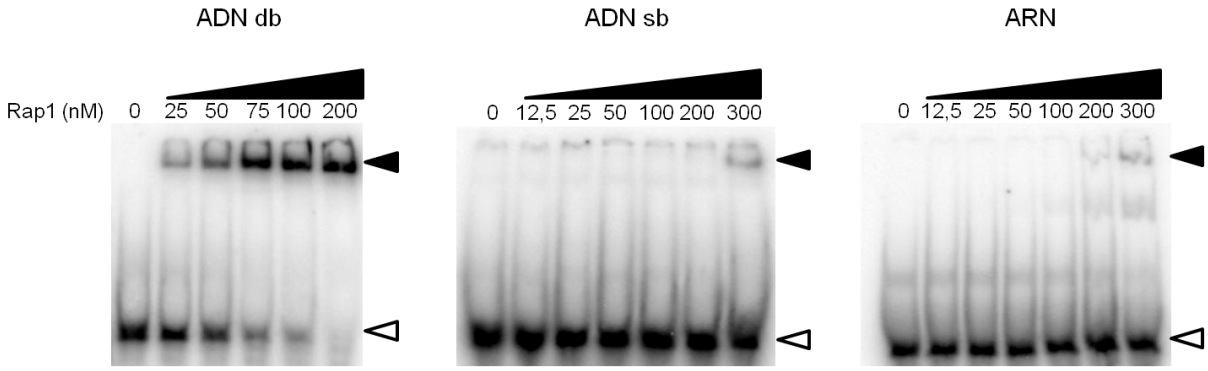

Figure 2.3.7. Rap1 se fixe à l'ADN double brin mais pas à l'ADN simple brin ni à l'ARN.
L'expérience de retard sur gel a été menée en utilisant des quantités croissantes de protéine Rap1 recombinante et une quantité constante de sonde radiomarquée (50 fmol). Les complexes Rap1-sonde sont indiqués par la flèche noire et les sondes seules sont indiquées par la flèche blanche

Cela suggère donc que la fixation de Rap1 sur l'ADN provoque la terminaison de la même manière que Reb1, par un mécanisme de « road-block ».

II.3.2.6. Le site de liaison de Rap1 n'est pas transcrit dans les terminateurs dépendants de Rap1

Dans le cas d'un mécanisme de terminaison par road-block, le site de liaison de Rap1 ne serait pas transcrit et l'ARN se terminerait quelques nucléotides en amont du site Rap1. Cette distance de quelques nucléotides entre l'extrémité 3' de l'ARN et le site Rap1 correspondrait à l'empreinte de Rap1 et de l'ARN pol II sur l'ADN. Pour étudier le mécanisme de terminaison utilisé par Rap1, j'ai cartographié les extrémités

3' des ARN terminés au niveau de plusieurs terminateurs artificiels contenant un site de liaison pour Rap1.

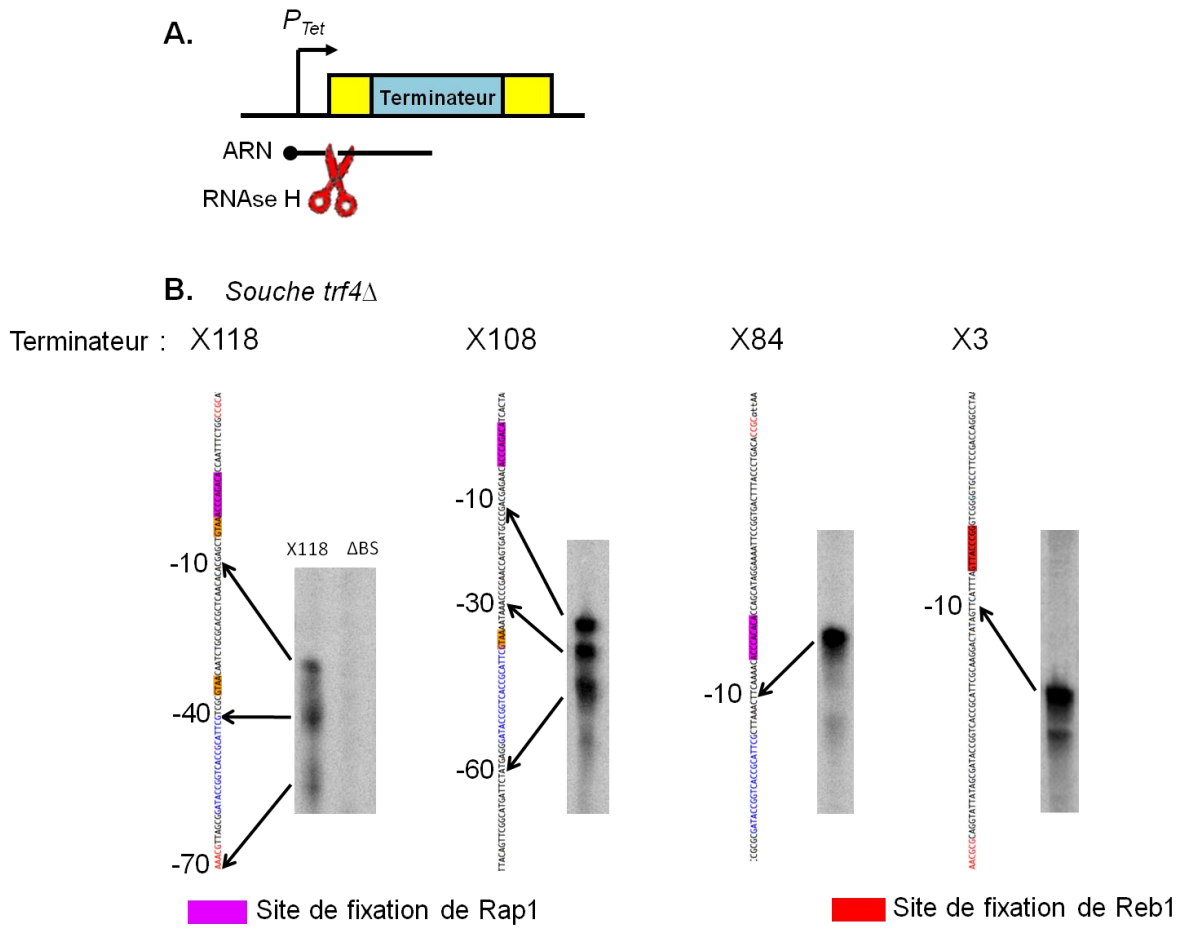

Figure 2.3.8. Les transcrits sont terminés en amont de Rap1.

A. Schéma représentant l'expérience réalisée pour cartographier les extrémités 3' des ARN issus des clones X118, X108, X84 et X3. Les ARN terminés par les terminateurs sélectionnés sont clivés par une digestion à la RNAseH pour avoir un 5' fixe, migrés sur un gel à haute résolution et analysés par northern blot.
B. Analyse par northern blot des ARN obtenus après digestion à la RNAse H. Les terminateurs X118, X108 et X84 contiennent un site de fixation de Rap1 (surligné en violet) et le terminateur contrôle X3 contient un site Reb1. Les ARN issus du terminateur X118 dans lequel le site Rap1 a été supprimé (ΔBS) ont également été analysés. Les positions des extrémités 3' des transcrits sont indiquées par les flèches. La distance approximative en nucléotides (à 5 nucléotides près) entre les extrémités 3' et les sites Rap1 ou Reb1 est indiquée. Les ARN sont extraits d'une souche délétée de *TRF4*.

Sur la figure 2.3.8., on constate pour chacune des séquences étudiées que les sites de terminaison les plus distaux se situent toujours une dizaine de nucléotides en amont des sites Rap1, ce qui a été confirmés par 3' RACE (résultats non montrés). Des résultats identiques ont été obtenus avec trois autres terminateurs artificiels

contenant un site Rap1 (résultats non montrés). Comme contrôle j'ai utilisé la séquence X3, contenant un site Reb1. Dans ce cas la transcription se termine environ 10 nucléotides en amont du site de fixation de Reb1 (Colin et al., 2014 et figure 2.3.8.).

Ces résultats sont compatibles avec l'hypothèse que les sites de fixation de Rap1 ne sont pas transcrits, et confortent l'hypothèse d'un mécanisme de terminaison de la transcription par road-block par Rap1.

De façon intéressante on a observé que, bien que les extrémités les plus distales des ARN analysés soient toujours situées quelques nucléotides en amont des sites Rap1, d'autres espèces plus courtes sont aussi détectées et leur présence dépends du site Rap1 (figure 2.3.8.B, séquence X118).

II.3.2.7. Le mécanisme de désassemblage des ARN pol II bloquées par Rap1

Par analogie avec le mécanisme de terminaison road-block de Reb1 (Colin et al., 2014) nous avons suspecté que les ARN pol II bloquées par Rap1 soient elles aussi éliminées *via* un système d'ubiquitination/dégradation.

L'accumulation d'ARN pol II arrêtées représente un danger pour la cellule et il a été montré que les polymérases bloquées sont éliminées par un mécanisme d'ubiquitination, désassemblage et dégradation (chapitre I.2.3.3.). La mono-ubiquitination de la grande sous-unité de l'ARN pol II, Rpb1, est catalysée par Rsp5, et cette étape est un pré-requis à la formation d'une chaîne de polyubiquitine qui mène à la dégradation de Rpb1 et le démantèlement du complexe d'élongation.

Afin de savoir si ce mécanisme d'ubiquitination est impliqué dans la terminaison par Rap1, j'ai analysé les transcrits terminés au niveau de la séquence X118 dans une souche *rsp5-1* dans laquelle le processus d'ubiquitination est altéré. On peut s'attendre à ce qu'en absence d'ubiquitination de Rpb1, les polymérases s'accumulent en amont de Rap1. Cela devrait coïncider avec une augmentation de la quantité d'ARN naissants associés aux ARN pol II accumulées et donc se traduire par une augmentation des transcrits détectés.

Conformément à cette hypothèse, on observe une accumulation de l'ARN terminé en amont de Rap1 dans la séquence X118 dans une souche *rsp5-1* (figure 2.3.9.). L'accumulation de ces ARN est probablement due à la protection de leurs extrémités 3' contre l'exosome nucléaire puisque ces ARN naissants restent associés aux complexes d'élongation bloqués qui ne sont pas éliminés en absence d'ubiquitination. On observe aussi une augmentation du transcrit correspondant à la translecture du terminateur de l'ARN court. Cette augmentation pourrait correspondre à une augmentation de la quantité de translecture du terminateur provoquée par la diminution de la dégradation des polymérases arrêtées par Rap1.

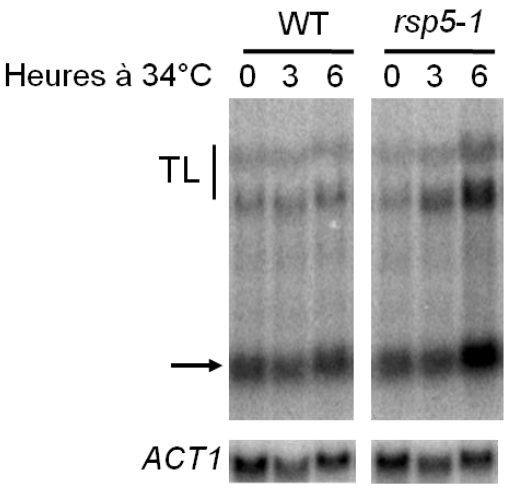

Figure 2.3.9. L'ARN terminé par Rap1 s'accumule lorsque l'ubiquitination des ARN pol II arrêtées est inhibée.
Analyse par northern blot des ARN issus du terminateur X118 dans une souche sauvage (WT) ou possédant un allèle thermo-sensible de *RSP5*. Un défaut d'élimination des ARN pol II arrêtées par Rap1 mène à une accumulation de l'ARN naissant (indiqué par une flèche).

Ces résultats indiquent que les ARN pol II arrêtées par Rap1 sont très probablement ubiquitinées et dégradées par un mécanisme qui fait intervenir Rsp5.

II.3.2.8. Rap1 termine la transcription dans le génome de S. cerevisiae

Toutes les expériences de caractérisation du mécanisme de terminaison ont été réalisées avec des constructions artificielles. Pour chercher des transcrits terminés par Rap1 chez *S. cerevisiae*, nous avons croisé les données de distribution de Rap1 et de la cartographie des transcrits instables. Les positions précises des sites liés par Rap1 sur l'ensemble du génome ont été cartographiées par une expérience de ChIP-

exo (Rhee and Pugh, 2011) et nous avons comparé ces données à une cartographie des transcrits instables par tiling arrays (puces ADN à haute densité) (Neil et al., 2009). A l'aide de Tito Candelli, bio-informaticien du laboratoire, j'ai identifié de nombreuses cibles putatives de la voie de terminaison par Rap1.

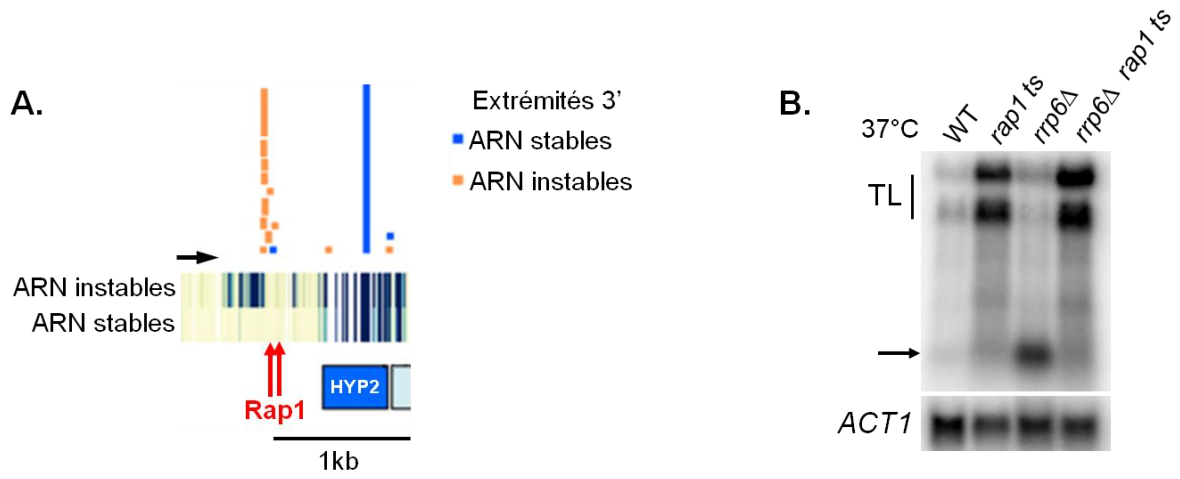

Figure 2.3.10. Exemple d'un transcrit naturel terminé par Rap1en amont du gène *HYP2*.

A. Transcrits stables et instables issus de la région du gène *HYP2*. Analyse des régions transcrites par tilling array et cartographie des extrémités 3' par 3' SAGE (Neil et al., 2009). Le sens de la transcription est indiqué par la flèche noire. Un transcrit instable en amont de *HYP2* est terminé au niveau de deux sites de fixation de Rap1 (flèche rouge).
B. Test de terminaison par les deux sites Rap1 extraits en amont de *HYP2*. La séquence contenant les deux sites de fixation de Rap1 a été insérée dans le système rapporteur permettant de tester la terminaison. La construction a été placée dans des souches : sauvage (WT), possédant un allèle thermo-sensible de *RAP1*, une délétion de *RRP6* ou une combinaison des deux mutations. L'insertion des deux sites Rap1 permet la terminaison d'un transcrit instable (flèche noire) qui disparaît lorsque la protéine Rap1 est inactivée.

La figure 2.3.10.A. montre l'une de ces cibles. Il s'agit d'un ARN instable transcrit en amont du gène *HYP2*, nommé *uHYP2* pour « upstream » *HYP2.* Les résultats de tiling arrays montrent l'existence d'un ARN instable se terminant au niveau de deux sites de fixation de Rap1 espacés de 20 nucléotides. On observe également une forte accumulation d'extrémités 3' de transcrits au niveau de ces sites de fixation de Rap1 par 3' SAGE (serial analysis of gene expression). Rap1 pourrait donc être responsable de la terminaison de cet ARN.

J'ai testé le rôle de Rap1 dans la terminaison de cet ARN. Puisque ce transcrit était très difficile à détecter par northern blot, j'ai inséré un fragment d'ADN contenant les deux sites Rap1 du locus *uHYP2* dans le système rapporteur permettant de tester la

terminaison. J'ai placé cette construction dans quatre souches : sauvage, possédant un allèle thermo-sensible de *RAP1*, possédant une délétion de *RRP6* ou possédant une combinaison des deux mutations. L'analyse par northern blot permet d'observer que l'insertion des deux sites Rap1 provoque la production d'un transcrit instable qui disparaît lorsque la protéine Rap1 est inactivée (figure 2.3.10.B).

Afin de valider l'implication de Rap1 dans la terminaison du transcrit naturel *uHYP2*, j'ai mesuré la quantité d'ARN transcrit en amont et en aval de ces deux sites Rap1 en présence et en absence de la protéine Rap1 par PCR quantitative. La protéine Rap1 a été déplétée en utilisant un système « Anchor-away » (Haruki et al., 2008) permettant la déplétion nucléaire de Rap1 en présence de rapamycine.

Lorsque Rap1 est déplété, on constate sur la figure 2.3.11. B. une augmentation de la quantité d'ARN en aval des sites de fixation de Rap1 par rapport à la quantité d'ARN en amont de ces sites suggérant fortement que Rap1 est bien impliqué dans la terminaison de cet ARN. On remarque également que la déplétion de Rap1 engendre une augmentation de la quantité de transcrit en amont du site Rap1. Cette augmentation pourrait être provoquée par un mode de terminaison différent de ces transcrits en absence de Rap1 menant à leur stabilisation et donc à leur accumulation.

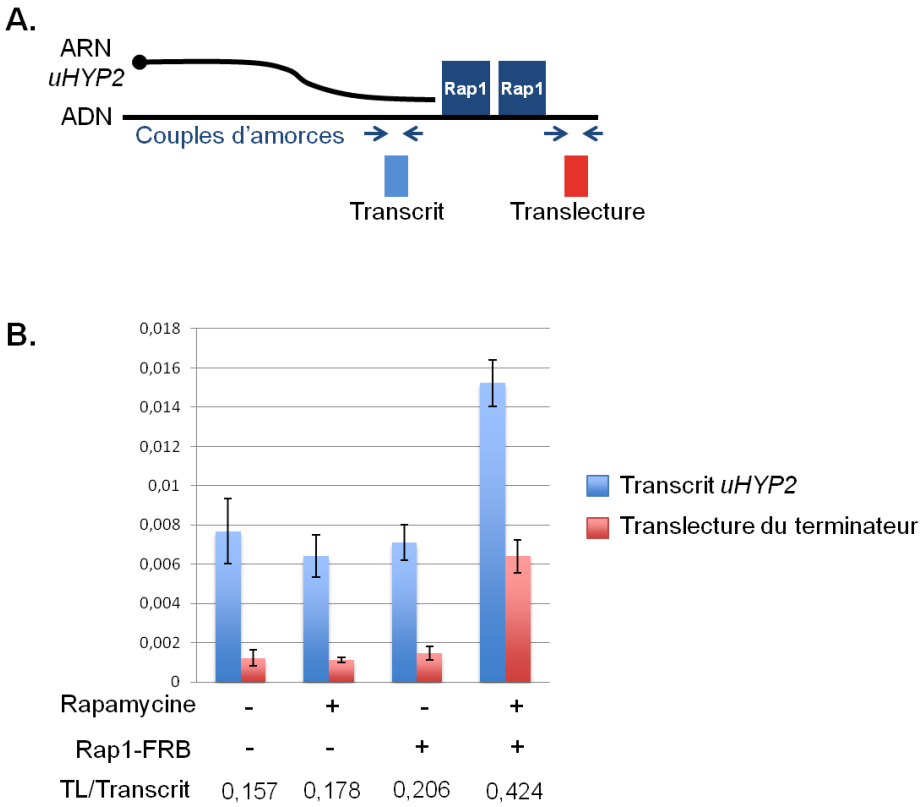

Figure 2.3.11. Rap1 est impliqué dans la terminaison d'*uHYP2*.

A. Description des couples d'amorces utilisés pour détecter la transcription avant et après les deux sites Rap1 identifiés en amont de *HYP2* par rétro-transcription et PCR quantitative (RT-qPCR).

B. Analyse par RT-qPCR de la transcription au locus du transcrit *uHYP2*. L'analyse est effectuée dans une souche modifiée pour utiliser un système anchor-away, celui-ci permet la déplétion nucléaire de Rap1 lors de la présence combinée de rapamycine dans le milieu de culture et d'une étiquette FRB (FKBP12-rapamycin binding) sur Rap1. Le ratio de la quantité d'ARN détecté en aval des deux sites Rap1 (TL) comparé à la quantité de transcrit *uHYP2* est indiqué pour les 4 conditions de l'expérience. Lorsque Rap1 est déplété le taux de translecture augmente indiquant un défaut de terminaison. Les barres d'erreurs représentent l'écart type en les résultats de trois expériences indépendantes.

Pour comparer le rôle des deux sites de fixation de Rap1 dans la terminaison d'uHYP2, j'ai réutilisé le système rapporteur permettant de tester la terminaison dans lequel j'ai inséré un fragment d'ADN contenant les deux sites Rap1 (figure 2.3.12. A et B). La présence de ces deux sites permet aux cellules de pousser en présence de cuivre, ce qui confirme leur capacité à terminer la transcription.

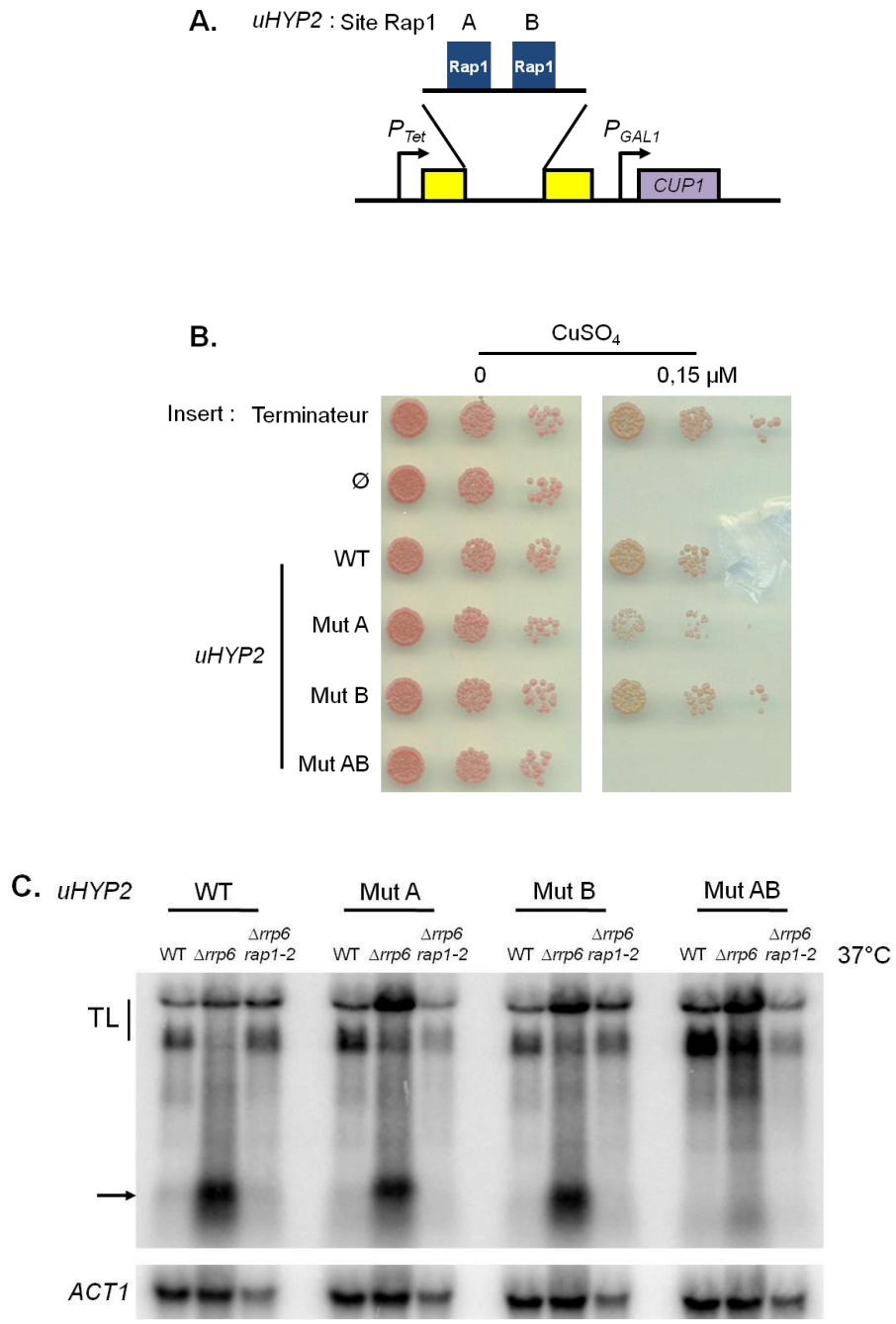

Figure 2.3.12. Chacun des deux sites Rap1 d'*uHYP2* est suffisant pour la terminaison.

A. Schéma du système rapporteur et du fragment d'*uHYP2* inséré. Cette construction est utilisée dans les deux expériences suivantes.
B. Analyse de l'efficacité de terminaison de 4 versions du terminateur d'*uHYP2* par test de croissance. La terminaison par les terminateurs sauvage (WT) et mutés sur les sites Rap1 A et/ou B a été testée. Les deux premières lignes correspondent à des constructions témoins contenant, ou non, un terminateur. La mutation des deux sites Rap1 altère la terminaison.
C. Analyse par northern blot des ARN synthétisés à partir des constructions utilisées pour le test de croissance. Les constructions ont été placées dans des souches sauvage, délétée de *RRP6* et à la fois délétée de *RRP6* et possédant un allèle thermo-sensible de *RAP1*. Tout comme décrit dans la figure 2.3.11.A., les deux sites Rap1 mènent à la production d'un ARN court (flèche noire) qui disparaît lorsque Rap1 est inactivée ou lorsque les deux sites Rap1 sont mutés.

Les levures sont également capables de pousser quand un seul des deux sites de fixation est fonctionnel. Cependant en présence d'un seul site de fixation de Rap1 les colonies sont plus petites en présence de cuivre, ce qui indique une résistance au cuivre plus faible due à une terminaison moins efficace. Un seul site Rap1 est donc suffisant pour la terminaison de la transcription mais l'efficacité de cette dernière est plus élevée lorsque les deux sites sont présents.

Ces résultats sont confirmés par une analyse par northern blot des transcrits produits à partir des constructions testées sur cuivre (figure 2.3.12.C).

La figure 2.3.13.A. présente un second exemple de terminaison naturelle par Rap1. Au niveau de ce locus, le site de liaison de Rap1 se situe en aval du gène *STF1*, à l'extrémité 3' d'un transcrit instable qui semble être initié par le promoteur du gène *STF1*. Cet ARN est appelé *dSTF1* pour « downstream » *STF1*.

J'ai aussi testé l'implication de Rap1 dans la terminaison au locus *dSTF1* en comparant la quantité d'ARN transcrit en amont et en aval de ce site Rap1 en présence et en absence de la protéine Rap1. Quand Rap1 est déplété, on remarque sur la figure 2.3.13.B. qu'il y a une augmentation de la proportion d'ARN passant à travers le site de fixation de Rap1 par rapport à la quantité d'ARN en amont de ce site, le ratio de la quantité d'ARN détecté en aval du site Rap1 comparé à la quantité de transcrit *dSTF1* augmente d'un facteur 10. Cela suggère que Rap1 est bien impliqué dans la terminaison de cet ARN. Tout comme dans l'exemple du transcrit *uHYP2*, la déplétion de Rap1 engendre aussi une augmentation de la quantité de transcrit en amont du site Rap1 probablement parce que celui-ci est stabilisé grâce à l'utilisation d'une voie de terminaison de la transcription différente en absence de Rap1.

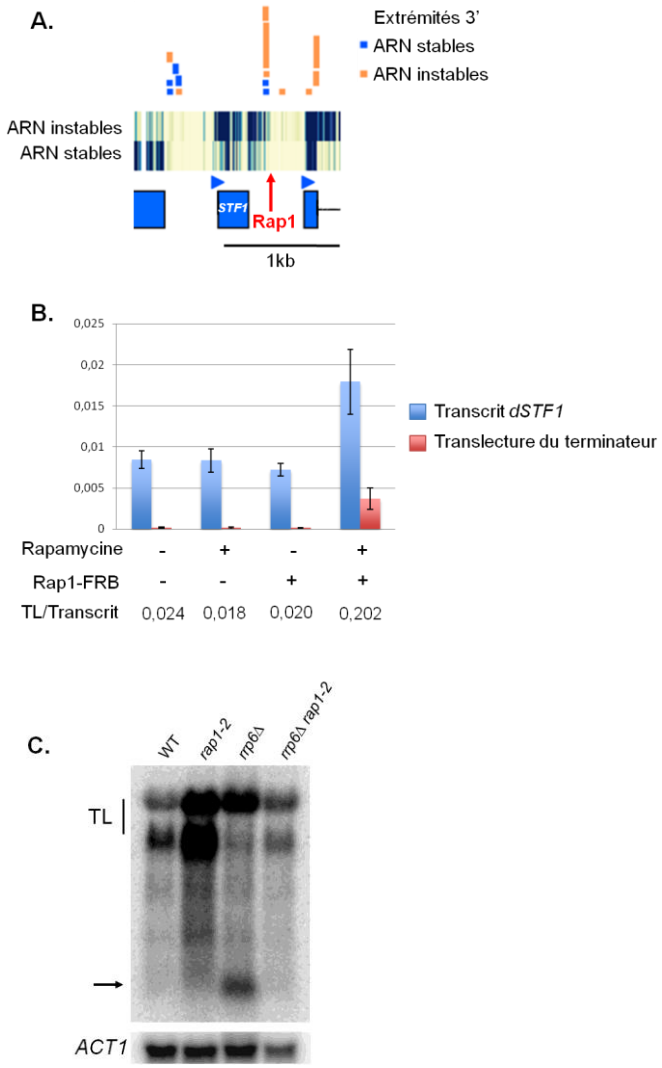

Figure 2.3.13. Exemple de transcrit naturel terminé par Rap1 en aval du gène *STF1*.

A. Transcrits stables et instables dans la région de *STF1*. Analyse de la transcription d'ARN stables et instables par puces à haute densité et cartographie des extrémités 3' par 3' SAGE (Neil et al., 2009). Le sens de la transcription est indiqué par les flèches bleues. Un transcrit chevauchant *STF1* est terminé au niveau d'un site Rap1 en aval de *STF1* (flèche rouge).

B. Analyse par RT-qPCR de la transcription autour du site Rap1 situé en aval de *STF1*. Les deux couples d'amorces ont été choisis de la même manière que pour *uHYP2* (figure 2.3.11.A.), les deux couples sont soit avant, soit après le site Rap1. L'analyse est effectuée dans une souche modifiée pour utiliser un système anchor-away, celui-ci permet la déplétion nucléaire de Rap1 lors de la présence combinée de rapamycine dans le milieu de culture et d'une étiquette FRB (FKBP12-rapamycin binding) sur Rap1. Le ratio de la quantité d'ARN détecté en aval du site Rap1 (TL) comparé à la quantité de transcrit *dSTF1* est indiqué pour les 4 conditions de l'expérience. Lorsque Rap1 est déplétée le taux de translecture augmente. Les barres d'erreurs représentent l'écart type en les résultats de trois expériences indépendantes.

C. Test de terminaison par le site Rap1 extrait en aval de *STF1*. La séquence contenant le site Rap1 a été insérée dans le système rapporteur permettant de tester la terminaison. La construction a été placée dans des souches sauvages (WT), possédant un allèle thermo-sensible de *RAP1*, une délétion de *RRP6* ou une combinaison des deux mutations. Le site Rap1 permet la terminaison d'un transcrit instable (flèche noire) qui disparaît lorsque Rap1 est inactivée.

Pour confirmer ce résultat j'ai testé la terminaison par ce site Rap1 dans notre système rapporteur. On voit sur la figure 2.3.13.C. que cette séquence induit la terminaison et que cette dernière est abolie lorsque Rap1 est inactivé dans une souche *rap1-2* à température restrictive. De plus, dans la souche *rap1-2* on constate une forte augmentation du transcrit correspondant à une translecture du terminateur dépendant de Rap1. Cette augmentation peut être expliquée par l'augmentation de la quantité de transcrit terminé en aval du site Rap1 quand Rap1 est inactivé. De manière étonnante, l'augmentation de ce transcrit n'est pas observable dans une souche *rrp6Δ rap1-2*, peut être parce que le niveau de transcription du système rapporteur est plus faible dans cette souche.

II.3.2.9. Accumulation d'ARN pol II en amont des sites de fixation de Rap1 dans le génome

Une des principales caractéristiques de la terminaison par road-block est qu'elle provoque une accumulation d'ARN pol II arrêtées aux sites de terminaison (Colin et al., 2014). Puisque Rap1 semble terminer la transcription par ce même mécanisme, on s'attend à observer une accumulation d'ARN pol II en amont des sites de liaison de Rap1.

J'ai donc comparé la distribution des ARN pol II en cours d'élongation à la position des sites de liaison de Rap1 dans le génome de *S. cerevisiae*. Pour analyser la distribution de l'ARN pol II, j'ai utilisé les résultats de NET-Seq (native elongating transcripts sequencing) de Churchman et Weissman (Churchman and Weissman, 2011). Cette technique repose sur le séquençage des extrémités 3' des ARN naissants associés à l'ARN pol II et permet donc le positionnement précis, au nucléotide près, des ARN pol II en cours de transcription.

Dans le cas du locus du transcrit instable *uHYP2*, on observe une forte accumulation de l'ARN pol II une vingtaine de paires de bases en amont du premier site de fixation de Rap1 (figure 2.3.14.A).

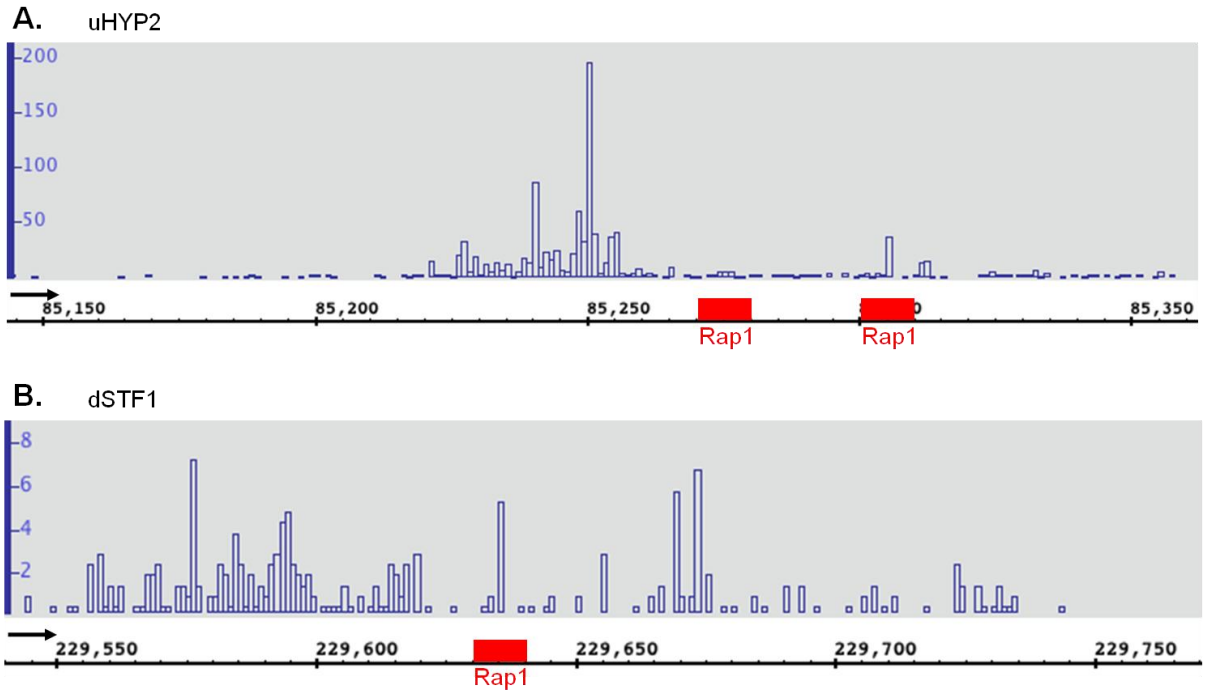

Figure 2.3.14. Distribution de l'ARN pol II au niveau des sites Rap1 des loci *uHYP2* et *dSTF1*.
La distribution de l'ARN pol II est révélée par NET-Seq (Churchman et Weissman, 2011). Les coordonnées sur les chromosome 5 (A) et 4 (B) sont indiquées en abscisse (en pb). La position des sites Rap1 est indiquée par les rectangles rouges. Le sens de la transcription est indiqué par les flèches noires.

Dans le cas du site de terminaison du transcrit *dSTF1*, l'accumulation d'ARN pol II en amont du site de fixation de Rap1 est aussi détectable, mais plus faible (figure 2.3.14.B).

Pour estimer l'occurrence de la terminaison par Rap1 dans l'ensemble du génome, avec l'aide de Tito Candelli, bio-informaticien au laboratoire, nous avons comparé la distribution de l'ARN Pol II à la distribution de Rap1 sur le génome déterminée par immunoprécipitation de chromatine (Harbison et al., 2004; MacIsaac et al., 2006). Dans cette analyse nous avons utilisé une distribution de l'ARN pol II déterminée par PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) (Schaughency et al., 2014), une analyse plus récente et possédant une plus haute résolution que la distribution de l'ARN pol II déterminée par NET-Seq utilisée pour l'analyse précédente. Tout comme le NET-Seq, Le PAR-CLIP permet notamment d'identifier l'extrémité 3' des transcrits naissants associés à l'ARN pol II et ainsi de cartographier précisément les ARN pol II en cours de transcription.

Cette analyse révèle une augmentation de la fréquence de pic d'ARN pol II en amont des sites de fixation de Rap1 correspondant une accumulation flagrante de l'ARN pol II entre 10 et 60 pb en amont de ces sites par rapport à la quantité de polymérase dans les environs (figure 2.3.15). Les pics d'ARN pol II sont définis comme des signaux significativement supérieurs à la moyenne locale des signaux dans une fenêtre de 200 nucléotides. Les résultats obtenus à partir des données de PAR-CLIP montrent une accumulation d'ARN pol II aussi bien en amont du site Rap1 dans l'orientation « riche en C » (qui correspond à l'orientation du site identifié dans le crible) qu'en amont de sites dans l'orientation inverse, « riche en G » avec des différences quantitatives mineurs. L'ARN pol II est donc bloquée par Rap1 dans le génome dans les deux orientations, même si ce blocage pourrait être un peu plus efficace lorsque Rap1 est fixé sur un site dans l'orientation « riche en C ».

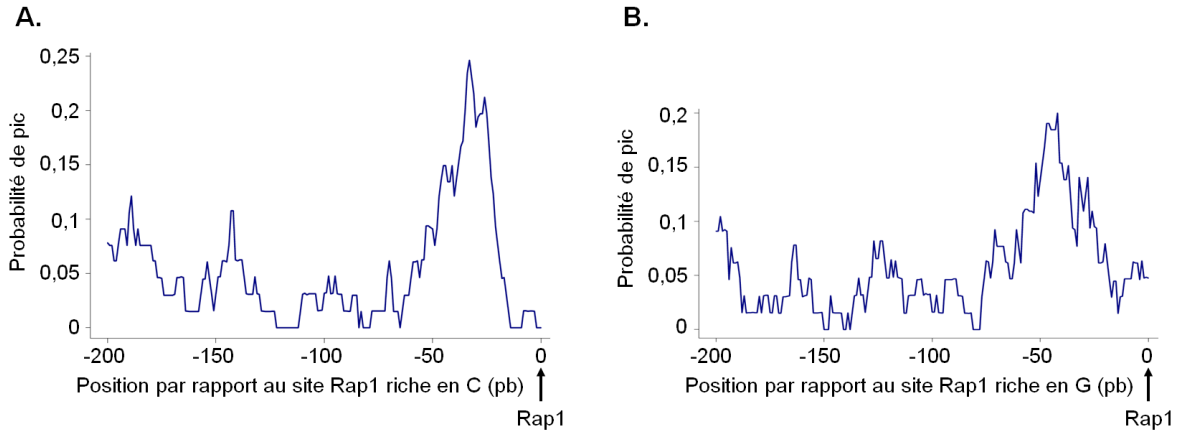

Figure 2.3.15. Analyse globale de la pause de l'ARN pol II en amont d'un métasite de fixation de Rap1.
La fréquence de pic d'ARN pol II, révélée par PAR-CLIP (Schaughency et al., 2014), est représentée dans une fenêtre de 200 pb en amont des sites de fixation de Rap1 trouvés dans l'ensemble du génome (Harbison et al., 2004 ; MacIsaac et al., 2006). La fréquence de pic a été calculée en amont de 68 sites Rap1, soit dans l'orientation riche en C (A.), soit dans l'orientation riche en G (B.).

Pour résoudre la pause transcriptionnelle, les ARN pol II bloquées reculent (chapitre I.2.3.3.) et l'extrémité 3' de l'ARN naissant sort du site catalytique de l'ARN pol II. Afin de réaligner l'extrémité 3' de l'ARN naissant avec le site catalytique de la polymérase, ce qui est indispensable à une éventuelle reprise de l'élongation, le facteur d'élongation TFIIS/Dst1 stimule le clivage de l'ARN par la polymérase. Dans une souche sauvage, la position de l'extrémité 3' d'un ARN associé à une ARN pol II

bloquée dépend donc du site de ce clivage et donc de la distance parcourue par l'ARN pol II lorsqu'elle a reculé. Cette distance étant variable, la position de l'extrémité 3' de l'ARN n'est donc pas fixe. En absence de TFIIS/Dst1, le clivage de l'ARN est inhibé et l'extrémité 3' du transcrit correspond précisément à la dernière position occupée par l'ARN pol II avant qu'elle ne recule, c'est-à-dire au point de collision avec l'obstacle dans le cas d'un road-block.

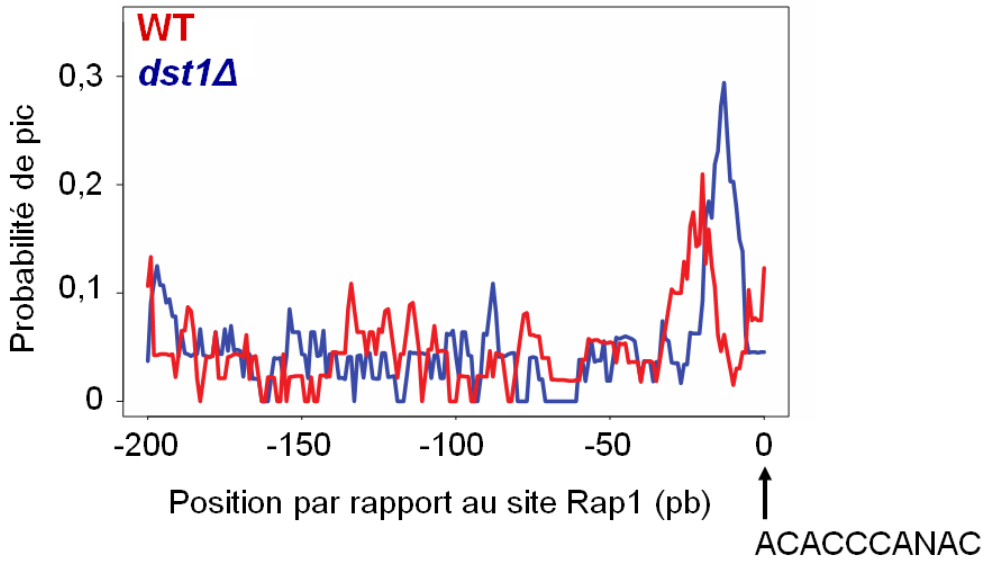

Figure 2.3.16. Analyse globale de la pause de l'ARN pol II en amont d'un métasite de fixation de Rap1 dans une souche sauvage et une souche *dst1*Δ.
La fréquence de pic d'ARN pol II, révélée par NET-Seq (Churchman et Weissman, 2011), est représentée dans une fenêtre de 200 pb en amont des séquences ACACCCANAC trouvées dans l'ensemble du génome. Cette séquence correspond à un site de fixation de Rap1. Cette analyse a été effectuée dans une souche sauvage (rouge) et en absence de TFIIS (*dst1*Δ, en bleu).

Pour confirmer l'arrêt de la transcription en amont des sites de fixation de Rap1 dans le génome nous avons comparé le niveau de pause globale de l'ARN pol II en amont des sites Rap1 dans une souche sauvage et dans une souche délétée du gène codant TFIIS (*dst1*Δ, figure 2.3.16., courbe bleue). En absence de TFIIS, on remarque que le pic d'ARN pol II est plus proche du site de fixation de Rap1 et moins dispersé, suggérant qu'il représente la dernière position occupée par la polymérase en amont de Rap1. Ces observations sont tout à fait compatibles avec un modèle de terminaison par road-block dans lequel TFIIS stimulerait le clivage des ARN associés aux ARN pol II qui auraient reculé suite à une collision avec Rap1.

L'ensemble de ces résultats montre qu'il existe des transcrits terminés par Rap1 chez *S. cerevisiae* et que cette terminaison repose sur un mécanisme de road-block.

II.3.3. Discussion

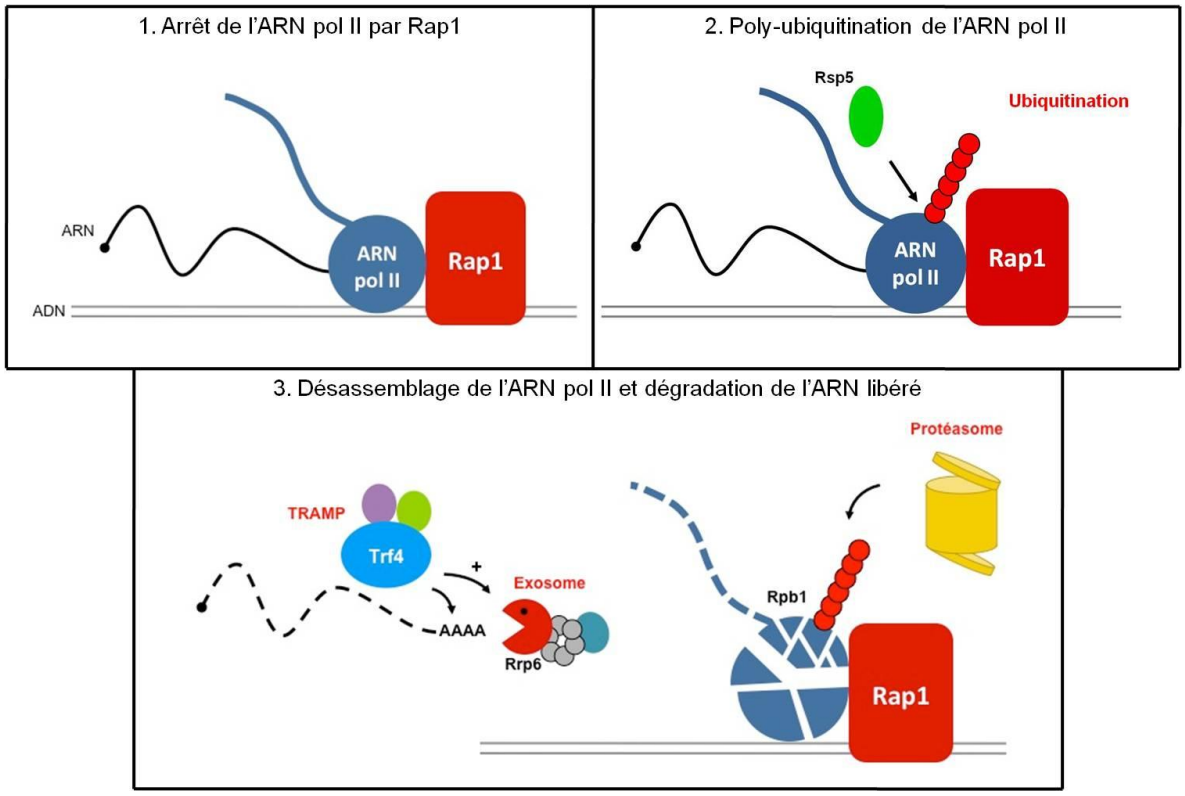

Figure 2.3.17. Modèle de terminaison de la transcription par Rap1.

1. Tout comme Reb1, Rap1 fixée sur l'ADN bloque la progression de l'ARN pol II ce qui provoque un arrêt de la transcription.
2. L'ARN pol II arrêtée serait poly-ubiquitinée. L'ubiquitine ligase Rsp5 mono-ubiquitine Rpb1 puis la chaîne d'ubiquitine est vraisemblablement allongée par l'action d'une autre ubiquitine ligase.
3. L'ARN pol II poly-ubiquitinée serait désassemblée et Rpb1 serait dégradée par le protéasome 26S. L'ARN libéré est polyadénylé par la sous-unité du complexe TRAMP, Trf4. Trf4 favorise également la dégradation de l'ARN par l'exosome nucléaire.

Le génome de *S. cerevisiae* est très compact et largement transcrit, jusqu'à environ 85% du total en milieu riche (David et al., 2006). La terminaison de la transcription est donc particulièrement importante pour assurer la délimitation des unités de transcription. En plus de permettre la production d'extrémités 3' de transcrits fonctionnels ou de favoriser la dégradation de certains transcrits, une terminaison efficace contribue à éviter l'interférence transcriptionnelle et à contrôler la transcription envahissante. Les résultats présentés dans cette étude montrent que

Rap1, qui est un facteur de transcription et un élément structural des télomères, est également capable de terminer la transcription (figure 2.3.17.). Rap1 pourrait donc, par ce rôle, participer au contrôle de la stabilité de l'expression génique.

II.3.3.1. Un mécanisme de terminaison de la transcription par road-block conservé

Lorsque le facteur Rap1 est fixé sur l'ADN, il est capable de bloquer la progression de l'ARN pol II par un mécanisme de road-block. C'est un mécanisme conservé, retrouvé à la fois chez les eucaryotes et les procaryotes. Il existe des exemples bactériens de terminaison par road-block décrits chez *E. coli* (Chintakayala et al., 2013; Deuschle et al., 1986; He and Zalkin, 1992) et chez *B. subtilis* (Belitsky and Sonenshein, 2011; Choi and Saier, 2005). Chez les eucaryotes, les trois ARN polymérases sont aussi susceptibles d'être bloquées par des facteurs liés à l'ADN, c'est le cas par exemple de Reb1 qui peut, *in vitro*, bloquer l'élongation de la transcription par les trois ARN polymérases (Lang et al., 1994). Il a été montré récemment au laboratoire que Reb1 bloque *in vivo* la transcription par l'ARN pol II chez *S. cerevisiae* (Colin et al., 2014) et son paralogue, Nsi1 (TTF1 chez les mammifères), provoque la terminaison de la transcription de l'ADNr par l'ARN pol I, très probablement via un mécanisme de road-block (Evers and Grummt, 1995; Kuhn et al., 1990; Reiter et al., 2012). Parallèlement à la découverte du mécanisme de road-block par Rap1 présentée dans ce manuscrit, un cas probable de terminaison road-block par Rap1 a été décrit chez *S. cerevisiae*. Il a été montré qu'un site de liaison de Rap1, apparu suite à la mutagenèse aléatoire du gène de la reverse transcriptase du rétro-transposon Ty1, pouvait provoquer la terminaison prématurée de la transcription de ce gène (Yarrington et al., 2012).

Certains facteurs liés à l'ADN semblent donc constituer un obstacle pour les ARN polymérases. D'après une étude menée chez la bactérie, l'efficacité de ce blocage est notamment corrélée à l'affinité de ces facteurs de terminaison à l'ADN (Hao et al., 2014). Plus un facteur est fortement lié à l'ADN, plus il bloque efficacement les ARN polymérases. D'après cette caractéristique le facteur Rap1 pourrait être un terminateur efficace puisqu'il est capable de se lier de façon très stable sur certains sites, avec des constantes de dissociation inférieures à 60 nM (Piña et al., 2003; Vignais et al., 1987).

Puisque Rap1 est un obstacle à la transcription, ses sites de liaison ne devraient généralement pas être retrouvés au niveau de séquences codantes. Conformément à cela, Rap1 a une très forte spécificité de fixation pour les régions promotrices par rapport aux ORF. En effet, bien qu'il existe des motifs de fixation de Rap1 dans les ORF, ceux ci sont beaucoup moins fréquemment occupés par Rap1 que dans les promoteurs (Lieb et al., 2001) ce qui suggère que d'autres facteurs pourraient contribuer à la spécificité de fixation de Rap1.

II.3.3.2. Le mécanisme de la terminaison de la transcription par road-block par Rap1

Rap1 bloque la progression de l'ARN pol II sur l'ADN cependant l'efficacité de ce blocage semble dépendre de l'orientation de Rap1 sur l'ADN. L'orientation du site Rap1 « riche en C » par rapport au sens de la transcription parait plus favorable à la terminaison que l'orientation inverse, « riche en G ». On peut remarquer que les sites Rap1 sélectionnés dans le crible initial sont toujours orientés dans le sens du site « riche en C ». Dans les conditions du système rapporteur, seul le site Rap1 dans cette orientation semble permettre la terminaison (figure 2.3.2.). Cependant, l'analyse de la pause de l'ARN pol II en amont des sites de fixation de Rap1 *in vivo* montre que bien que l'ARN pol II s'arrête plus fréquemment devant les sites Rap1 dans l'orientation « riche en C », un arrêt de l'ARN pol II est observable pour les deux orientations du site Rap1 (figure 2.3.15). Cela suggère que l'efficacité du blocage de l'ARN pol II par Rap1 dépend de l'orientation de Rap1 sur l'ADN. On peut imaginer qu'en aval d'un promoteur fort, comme le promoteur Tet du système rapporteur, Rap1 ne serait capable de bloquer la transcription suffisamment efficacement que dans son orientation optimale. Au contraire, dans des régions moins transcrites, les deux orientations du site Rap1 permettraient la terminaison même avec un blocage de l'ARN pol II moins efficace. Cette hypothèse est crédible puisqu'il a été montré chez *E. coli* que des ARN polymérases successives, initiées en série à partir d'un promoteur fort, peuvent collaborer pour dépasser un facteur de road-block (Hao et al., 2014).

L'étude de la structure du domaine de liaison à l'ADN de Rap1 donne des indices expliquant la corrélation entre l'efficacité de la terminaison par Rap1 et l'orientation de son site sur l'ADN. Le domaine de liaison à l'ADN de Rap1 contient deux

domaines de liaison Myb-like, ils lient l'ADN en tandem au niveau du grand sillon. Sur un site dans le sens « riche en C » : 5'-ACACCCANAC-3', le domaine aminoterminal lie la partie 5' (en bleu) du site et le domaine carboxyterminal lie la partie 3' (en violet). La région voisine en C-ter des deux domaines Myb-like se replie vers le premier domaine Myb-like et s'enroule autour de l'ADN (König et al., 1996; Matot et al., 2012). Ce repliement autour de l'ADN permet une fixation plus efficace de Rap1 à la partie 5' du site (en bleu) dans l'orientation « riche en C », ce qui pourrait expliquer le blocage plus efficace de l'ARN pol II lorsque Rap1 présente cette orientation (voir figure 2.3.18.). Au contraire le site « riche en G » menerait a une fixation plus faible de Rap1 sur la partie 5' du site, ce qui diminuerait l'efficacité de blocage de l'ARN pol II.

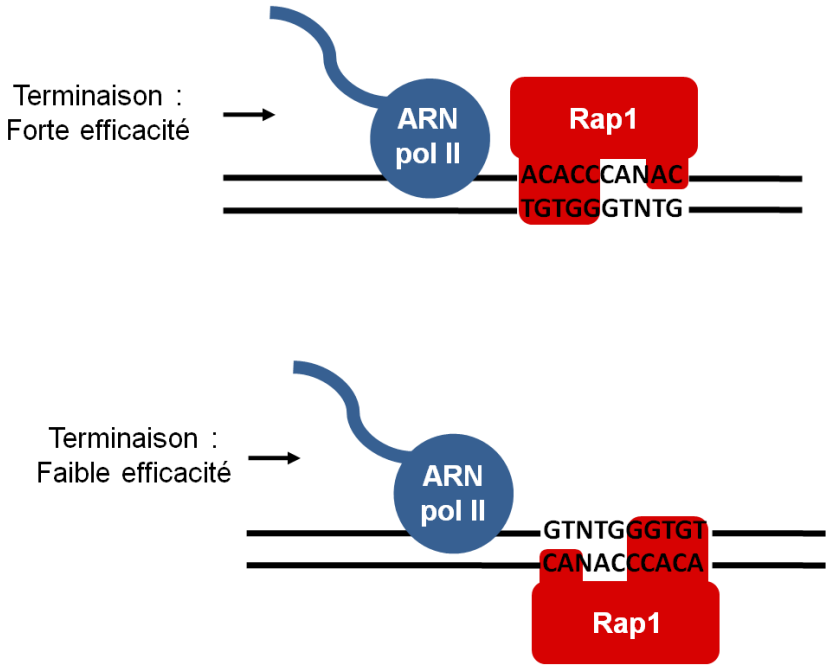

Figure 2.3.18. Schéma représentant l'efficacité de la terminaison par Rap1 en fonction de son orientation sur l'ADN.

A. Rap1 est fixé a un site dans l'orientation « riche en C ». Rap1 lie de façon plus efficace la partie 5' du site, ce qui permettrait une terminaison plus efficace.
B. Rap1 est fixé a un site dans l'orientation « riche en G ». Il lie de façon plus efficace la partie 3' du site, ce qui diminuerait l'efficacité de la terminaison.

Figure partiellement adaptée de Yarrington et al. 2012.

De plus, comme cela est visible sur la majorité des northern blot réalisés dans lesquels un transcrit terminé par Rap1 est détecté, au moins un autre transcrit correspondant à une translecture du site Rap1 est observable (par exemple : figure 2.3.3.A., TL). La présence de ce transcrit suggère que la terminaison par Rap1 n'a

pas une efficacité optimale en aval d'un promoteur qui a une fréquence d'initiation élevée.

Tout comme Rap1, Reb1 possède également deux domaines Myb-like en tandem et l'efficacité de la terminaison par Reb1 dépend de l'orientation de son site (Colin et al., 2014). On peut penser que dans le cas de cette protéine aussi, un des deux domaines Myb-like soit plus fortement fixé à l'ADN, ce qui pourrait influencer l'efficacité du blocage de l'ARN pol II en fonction de l'orientation de Reb1. Il est aussi intéressant de remarquer que, chez *S. cerevisiae*, les trois seules protéines qui possèdent deux domaines Myb-like sont Rap1, Reb1 et Nsi1 et que ces protéines sont toutes trois capables de bloquer les ARN pol I et/ou II.

La distance entre l'extrémité 3' des transcrits plus longs générés par la terminaison dépendante de Rap1 et le site de fixation de Rap1 est compatible avec la distance entre le site actif de la polymérase, contenant l'extrémité 3' de l'ARN, et l'avant de la polymérase qui bute contre Rap1 (Saeki and Svejstrup, 2009). La figure 2.3.8.B. montre qu'il existe aussi des ARN plus courts, la production desquels dépend aussi de Rap1. Ces extrémités 3' espacées de 20 à 30 nucléotides pourraient être le reflet de la collision d'ARN pol II successives en amont de Rap1, l'espacement entre les extrémités étant compatible avec un tel modèle (Saeki and Svejstrup, 2009). On peut également imaginer que ces extrémités 3' plus éloignées du site Rap1 soient le fruit d'évènements de terminaison plus précoces provoqués par une multimérisation de Rap1 sur l'ADN. En effet, une étude *in vitro* récente suggère que Rap1 pourrait former des complexes multimériques liés à l'ADN (Feldmann and Galletto, 2014). Ces évènements de terminaison précoces pourraient aussi être provoqués par d'autres facteurs liés à l'ADN et dont le recrutement dépendrait de Rap1.

Le mécanisme de libération des ARN pol II bloquées par Rap1 fait intervenir la voie de poly-ubiquitination de Rpb1 et vraisemblablement de dégradation par le protéasome, qui est aussi employée pour éliminer les ARN pol II arrêtées face à une lésion de l'ADN. Il n'est pas exclu que d'autres facteurs soient impliqués dans la libération des ARN pol II bloquées par Rap1 même si nous avons montré que le site Rap1 seul est suffisant pour la terminaison (figure 2.3.5.).

Bien que la dégradation des polymérases suite à un blocage transcriptionnel par Rap1 ne soit pas formellement montrée, chaque événement de terminaison pourrait

générer la destruction d'une molécule d'enzyme. Ce mécanisme ne paraît pas économique au premier abord. Cependant la terminaison par Rap1 pourrait fonctionner de façon principale comme système de secours employé lorsque les autres voies de terminaison n'ont pas été totalement efficaces (comme démontré pour Reb1 (Colin et al., 2014)), ce qui limite le nombre d'évènements de terminaison par Rap1 menant à la perte d'ARN pol II.

Aussi, le coût évolutif de ce mécanisme est limité puisqu'il utilise des facteurs sélectionnés pour une autre fonction (par exemple : Rap1 pour le contrôle de la transcription) et une quantité d'information faible (un site d'une dizaine de nucléotides) pour définir un site de terminaison. Il est donc possible que le coût énergétique de la perte d'une molécule de polymérase par événement de terminaison soit compensé par la simplicité du mécanisme et le faible cout évolutif de sa mise en place.

Les transcrits dérivés de la terminaison par Rap1 sont polyadénylés par la poly(A)-polymérase du complexe TRAMP, Trf4, et dégradés par l'exosome nucléaire (figure 2.3.6.). Par analogie avec la voie de terminaison NNS, l'intervention de Trf4 promeut très probablement la dégradation. Cependant, et contrairement au mécanisme de terminaison par le NNS, le mode de recrutement du complexe TRAMP est inconnu. Il ne dépend pas du complexe NNS puisque la déplétion de Nrd1 n'a pas d'effet sur la stabilité de l'ARN terminé par Rap1 (figure 2.3.3.B.). Il est possible que le complexe TRAMP soit recruté par un facteur spécifique ou qu'il cible systématiquement les extrémités 3' d'ARN non protégées dans le noyau.

II.3.3.3. La fonction biologique de la terminaison par Rap1

Etant donné que la terminaison par Rap1 mène à la production de transcrits instables, il est peu probable que cette terminaison conduise à la production d'ARN fonctionnels. Par analogie avec la terminaison road-block dépendante de Reb1 nous proposons que la terminaison dépendante de Rap1 participe au contrôle de la transcription envahissante et joue un rôle dans la neutralisation des fuites au niveau de la terminaison par d'autres voies.

En effet, l'étude de la terminaison par Reb1 a montré au moins un exemple de gène, *OST4*, pour lequel Reb1 a un rôle non pas dans l'activation de la transcription mais

dans la protection contre l'interférence transcriptionnelle (Colin et al., 2014). Dans ce cas, Reb1 permet de neutraliser des fuites au niveau de la terminaison du gène en amont d'*OST4*. En absence de Reb1 l'expression d'*OST4* est fortement affectée, mais peut être restaurée par l'expression du seul domaine de liaison à l'ADN de Reb1. Nous n'avons pas encore pu démontrer de façon expérimentale que ce concept est aussi valable pour Rap1. Par exemple il sera important de vérifier que dans le cas des unités de transcription naturellement terminés par Rap1, *uHYP2* et *dSTF1*, l'expression des gènes en aval, *HYP2* et *RPP1B* soit dépendante de Rap1. Puisque Rap1 pourrait activer la transcription de ces gènes, il faudrait savoir si Rap1, dépourvue de toute activité activatrice, reste nécessaire et suffisante pour protéger les gènes *HYP2* et *RPP1B* simplement par sa fonction de terminaison de sauvegarde.

Le positionnement de Rap1 dans le génome est cohérent aussi bien avec une fonction d'activation de la transcription que de « protection » des régions promotrices des interférences transcriptionnelles. Les sites de fixation de Rap1 sur les promoteurs sont fréquemment retrouvés dans l'ADN associé au nucléosome -1 (Koerber et al., 2009), en amont de la boîte TATA et du site d'initiation de la transcription. Au niveau des promoteurs, Rap1 est depuis longtemps connu pour ses activités antagonistes d'activateur ou de répresseur de la transcription (Kurtz and Shore, 1991). En tant qu'activateur, Rap1 participe notamment à la création de NDR à partir desquelles la transcription est initiée.

La cellule pourrait tirer profit de la présence d'une barrière bloquant la transcription au niveau des promoteurs, on peut ainsi imaginer au moins deux effets bénéfiques apportés par Rap1. Tout d'abord, Rap1 pourrait bloquer les ARN polymérases menaçant d'envahir les promoteurs. De plus, Rap1 pourrait également bloquer l'initiation de la transcription divergente en empêchant la transcription à travers le nucléosome -1 auquel il est fixé (voir figure 2.3.20).

Il a été observé qu'entre deux gènes divergents, Rap1 lie de manière sélective les nucléosomes -1 partagés par ces deux gènes, mais pas lorsque le nucléosome -1 d'un gène est aussi le +1 du gène divergent (Koerber et al., 2009; Yarrington et al., 2012, figure 2.3.19.), suggérant l'existence d'une pression évolutive qui empêche la présence de sites de fixation de Rap1 dans les unités de transcription. Son

positionnement sur le nucléosome -1 commun à deux gènes divergeant pourrait participer à déterminer le sens de la transcription. En effet, les promoteurs sont considérés comme étant intrinsèquement bidirectionnels chez la levure, ce qui représente une source majeure de transcrits non-fonctionnels (Neil et al., 2009).

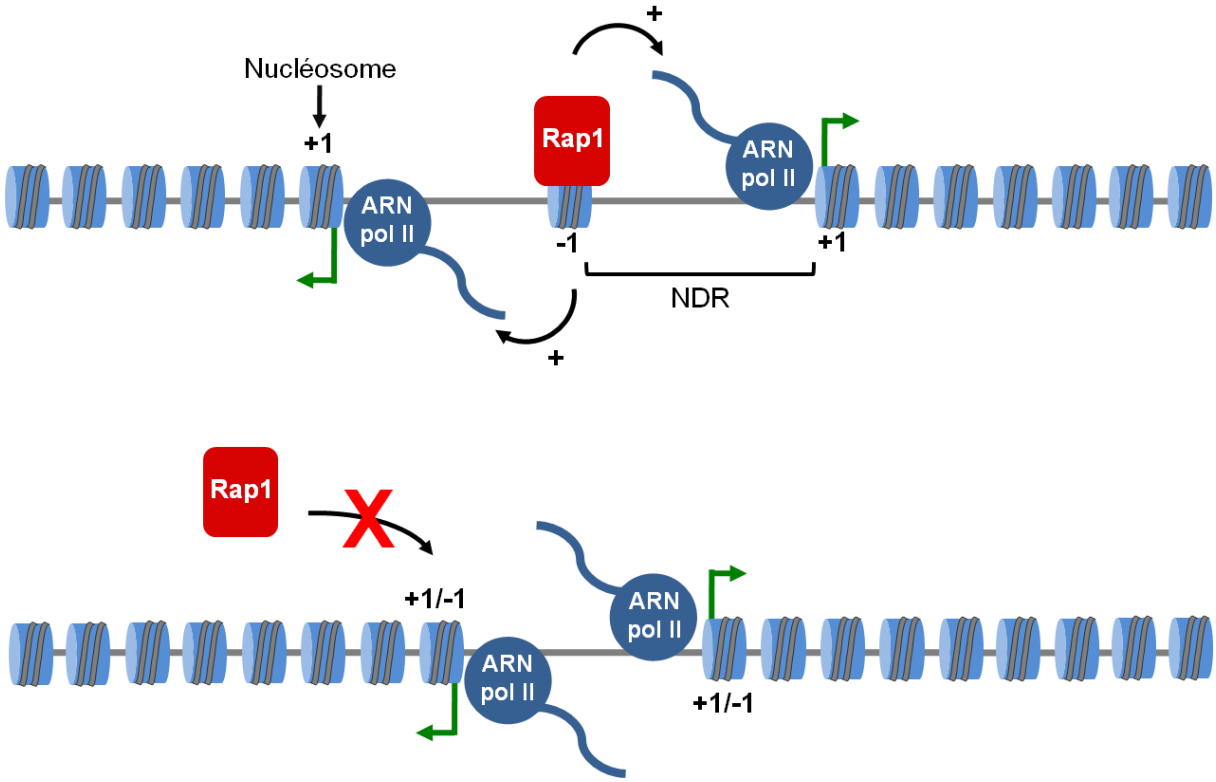

Figure 2.3.19. Rap1 se lie de manière sélective aux nucléosomes -1 qui sont partagés entre deux gènes divergents.
Lorsque Rap1 est fixé sur l'ADN, il peut favoriser la formation de NDR (flèches noires) ce qui permet d'initier la transcription (flèches vertes : TSS).
Figure adaptée de Koerber *et al.*, 2009.

La présence de Rap1 sur le nucléosome -1 pourrait permettre de bloquer la transcription antisense à partir des promoteurs qu'il cible. En empêchant l'initiation bidirectionnelle, Rap1 participerait au contrôle de l'initiation de la transcription envahissante. Cette hypothèse est soutenue par une étude montrant que la déplétion de Rap1 mène à l'apparition de transcrits divergents issus de NDR (van Bakel et al., 2013).

L'impact de Rap1 sur la ponctuation du génome est probablement important puisque plus de 500 sites de fixation de Rap1 ont été identifiés dans le génome (en plus des régions télomériques) (Rhee and Pugh, 2011). L'analyse globale de la pause de

l'ARN pol II en amont des sites de liaison de Rap1 (figure 2.3.16.) montre qu'il y a une accumulation d'ARN pol II en amont de 30% des sites Rap1 autour desquels la transcription est détectable par NET-Seq (98 sites ACCCANAC dans le génome).

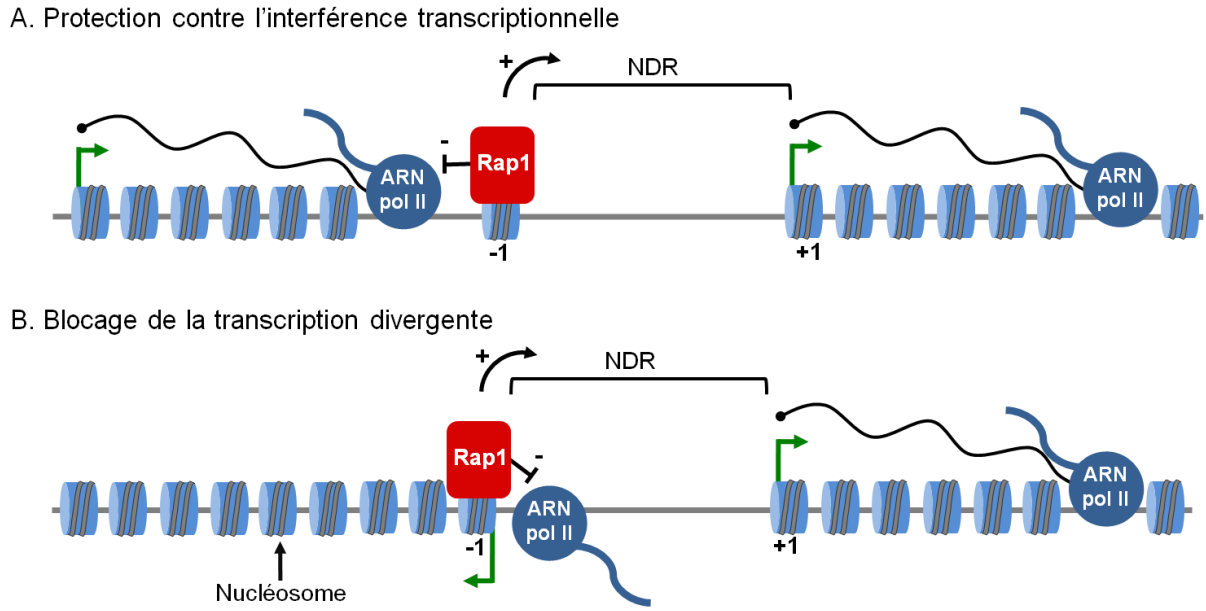

Figure 2.3.20. Modèle représentant la fonction de Rap1 dans l'établissement de NDR, dans la protection de la NDR contre la transcription envahissante et dans le blocage de la transcription divergente.
Rap1 favorise la formation de NDR au niveau des promoteurs qu'il cible (flèches noires) facilitant l'initiation de la transcription (flèches vertes : TSS).
Une fonction identique est assurée par Reb1 et semble aussi être assurée par Abf1.

II.3.3.4. D'autres road-blockers ?

Au laboratoire, nous nous sommes demandé si d'autres facteurs de transcription pourraient eux aussi bloquer l'élongation de la transcription. Pour répondre à cette question, j'ai testé la capacité de terminaison de quatre autres facteurs de transcription : Bas1, Tbf1, Fhl1 et Abf1, dont certaines caractéristiques sont similaires à celles de Reb1 et Rap1. J'ai inséré un site de fixation de chacun de ces quatre facteurs dans le système rapporteur utilisé pour tester la terminaison, mais je n'ai pas pu détecter de la terminaison en utilisant cette méthode. Il est possible que ceci soit du au fait que le promoteur Tet est un promoteur doté d'une fréquence d'initiation élevée, susceptible de nuire à la terminaison par des facteurs de road-block moins efficaces que Reb1. Il est aussi possible que la terminaison road-block

fonctionne de façon préférentielle dans des NDR et que les quatre facteurs testés soient moins efficaces pour induire la formation d'une NDR.

Nous avons analysé le rôle d'Abf1 dans son contexte naturel, dans l'espoir que les niveaux plus physiologiques de transcription dans le génome (par rapport au rapporteur) nous permettent de mettre en évidence des évènements de road-block. Une approche identique à celle entreprise pour l'analyse globale de la pause de l'ARN pol II en amont des sites de fixation de Rap1 (figure 2.3.15.) a été employée. Avec Tito Candelli, nous avons étudié la distribution de l'ARN pol II, révélée par NET-Seq (Churchman and Weissman, 2011), à la position des sites de fixation d'Abf1 (Harbison et al., 2004; MacIsaac et al., 2006). Cette analyse a permis de détecter l'accumulation d'ARN pol II une vingtaine de nucléotides en amont des sites de fixation d'Abf1 dans environ 50% des cas (figure 2.3.21.). Ce profil est similaire à ceux observés pour Rap1 et Reb1 (figure 2.3.15 et Colin et al., 2014), indiquant qu'Abf1 bloquerait la progression de l'ARN pol II par un mécanisme de road-block de la même manière que Rap1 et Reb1.

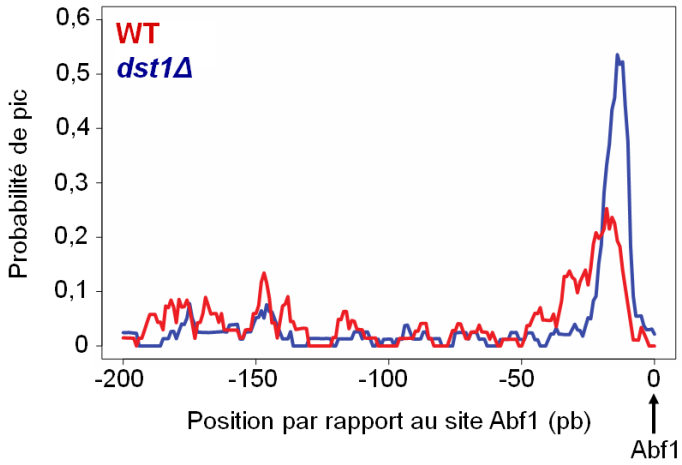

Figure 2.3.21. Analyse globale de la pause de l'ARN pol II en amont d'un métasite de fixation d'Abf1.
La fréquence de pic d'ARN pol II, révélée par NET-Seq (Churchman et Weissman, 2011), est représentée dans une fenêtre de 200 pb en amont des sites de fixation d'Abf1 répartis sur l'ensemble du génome (Harbison et al., 2004; MacIsaac et al., 2006). Cette analyse a été effectuée dans une souche sauvage (rouge) et en absence de TFIIS (*dst1*Δ, en bleu).

Cette hypothèse est corroborée par une étude suggérant qu'Abf1 protège le promoteur du gène *HIS7* contre l'interférence transcriptionnelle (Valerius et al., 2002). Les gènes *ARO4* et *HIS7* sont disposés en tandem. *ARO4* est transcrit en

amont et lorsque son promoteur est remplacé par un promoteur plus fort, la transcription d'*ARO4* est augmentée et, simultanément, la transcription de *HIS7* est réduite. Ces deux gènes sont séparés par 416 pb contenant notamment un site de fixation d'Abf1. Lorsque ce site est muté, conjointement à l'insertion d'un promoteur fort contrôlant l'expression du gène *ARO4*, on observe une diminution encore plus importante de la transcription d'*HIS7*. L'ampleur de la diminution de la transcription de *HIS7* est proportionnelle à la force du promoteur qui dirige l'expression d'*ARO4*, elle ne peut donc pas être uniquement attribuée à un défaut d'activation d'*HIS7* en absence d'Abf1 (Valerius et al., 2002). Même si ceci n'a pas été formellement démontré, il est vraisemblable qu'Abf1 protège le promoteur de *HIS7* contre les ARN pol II qui dépassent les signaux de terminaison de la transcription du gène *ARO4*.

De plus, comme Rap1 et Reb1, Abf1 semble également capable de bloquer la transcription divergente au niveau des NDR. En effet, il a été observé que l'inactivation d'Abf1 provoque l'apparition de transcrits associés à certains promoteurs (van Bakel et al., 2013), dont la plupart semblent issus d'une transcription divergente. Cette étude montre aussi des résultats identiques pour d'autres facteurs responsables de l'exclusion des nucléosomes comme Rap1, Tbf1 ou Rsc3, suggérant que plusieurs facteurs agissent sur la limitation de la transcription antisense, potentiellement par un mécanisme road-block.

Rap1, Reb1 et Abf1 ne bloquent pas uniquement la progression de la transcription. Ces trois facteurs semblent aussi bloquer la réplication par l'ADN pol δ. En effet, une étude récente a montré un fort enrichissement des extrémités 3' des fragments d'Okazaki une dizaine de nucléotides en amont des sites Rap1, Reb1 et Abf1 (Smith and Whitehouse, 2012). Ce profil, similaire à l'accumulation des extrémités 3' de transcrit en amont des sites de fixation de ces trois facteurs, suggère que Rap1, Reb1 et Abf1 bloquent de la même manière l'ADN pol δ et l'ARN pol II.

La progression des machineries de transcription et de réplication sur l'ADN peut perturber l'expression des gènes. Ces considérations suggèrent que les facteurs de transcription Rap1, Reb1 et Abf1 forment des barrières bloquant la progression des ARN et ADN polymérases pour limiter les perturbations qu'elles induisent.

II.3.4. Perspectives

Pour juger l'étendue de l'usage de la terminaison par Rap1 et Abf1, il faudrait évaluer la quantité de transcrits terminés par ces deux facteurs. Cette information pourrait être apportée par une étude transcriptomique à grande échelle permettant de comparer les populations de transcrits, leurs dimensions et la distribution des ARN pol II en présence et en absence de Rap1 ou d'Abf1. Puisque ce mécanisme de terminaison de secours pourrait être employé pour bloquer des évènements de transcription peu abondants, tel que des translectures de terminateur, l'étude transcriptomique devra avoir une grande sensibilité. Une analyse à haute résolution sera aussi nécessaire afin de pouvoir cartographier précisément la position de l'ARN pol II par rapport à la position d'éventuels sites de fixation de facteurs de road-block.
L'importance de la protection des NDR par Rap1 n'est pas formellement démontrée dans mes travaux. Pour montrer le rôle de Rap1 dans la protection contre l'interférence transcriptionnelle dans le génome de *S. cerevisiae* il faudrait montrer qu'une forme tronquée de Rap1, ne contenant pas le domaine C-ter notamment responsable de la formation des NDR (voir chapitre I.3.5.2.), est nécessaire et suffisante pour le maintien de la transcription de certains gènes grâce à sa fonction d'isolation contre la transcription provenant des régions environnantes. Une telle expérience est possible puisqu'il a été montré qu'une souche délétée du domaine C-ter de Rap1 est viable malgré d'importants défauts de croissance, indiquant que le domaine de liaison à l'ADN de Rap1 est suffisant pour la survie cellulaire. Il est possible que dans ces conditions de croissance difficiles, provoquées par l'absence des fonctions de régulation de l'initiation de la transcription et de la maintenance des télomères par Rap1, la survie cellulaire soit maintenue par la fonction de terminateur de Rap1. Ceci pourrait suggérer que la fonction de terminaison et d'isolation des unités de transcription représente une part importante de la fonction biologique de Rap1.
De nombreux facteurs peuvent se fixer à l'ADN, Rap1, Reb1 et Abf1 ne sont peut-être pas les seuls à pouvoir bloquer l'ARN pol II. Afin de savoir s'il existe d'autres facteurs de road-block il serait intéressant de comparer des données de la distribution de l'ARN pol II et la répartition des facteurs de liaison à l'ADN sur l'ensemble du génome. En détectant des sites d'accumulation de l'ARN pol II situés

quelques nucléotides en amont de site de liaison de protéines, on pourrait identifier l'ensemble des facteurs capables de bloquer l'élongation de la transcription.
La terminaison par un mécanisme de road-block n'est pas spécifique à *S. cerevisiae* et il serait intéressant d'étudier son impact chez d'autres organismes tel que l'Homme en particulier. Chez l'Homme, il existe déjà des indices suggérant qu'un tel mécanisme existe. Par exemple, le facteur de transcription MAZ, dont le mode d'action est mal connu, serait capable de bloquer l'interférence transcriptionnelle. Il existe au moins deux cas dans lesquels MAZ se lie entre deux gènes proches et protège le gène en aval contre l'interférence transcriptionnelle en favorisant la pause de l'ARN pol II et la terminaison de la transcription provenant du gène situé en amont (Ashfield et al., 1994; Yonaha and Proudfoot, 1999).
Il est probable que chez l'Homme, bien que son génome soit moins compact que celui de *S. cerevisiae*, l'importance de la ponctuation du génome et de la protection contre l'interférence transcriptionnelle par des facteurs liés à l'ADN soit répandu, surtout dans un contexte de transcription envahissante. Les facteurs de transcription ne seraient pas seulement de simples interrupteurs de la transcription des gènes mais ils pourraient aussi avoir un rôle important dans la maintenance de la stabilité du transcriptome.

III. Matériels et méthodes

III. 1. Souches de *S. cerevisiae* utilisées

Nom	Numéro	Genotype	Origine
WT	DLY671 (BMA64)	*ura3-1, ade2-1, his3-11, 5, trp1Δ, leu2-3,112, can1-100, MAT a*	(Baudin et al., 1993)
WT	W303	*ura3-1, ade2-1, his3-11, 5, trp1-1, leu2-3,112, can1-100, MAT a*	(Thomas and Rothstein, 1989)
cup1Δ	DLY841	*W303, cup1::KAN*	
e	DLY814	*BMA64, rrp6::KAN MAT alpha*	
rrp6Δ trf4Δ	DLY678	*BMA64, rrp6::URA3, trf4::KAN, MAT a*	
trf4Δ	DLY77	*BMA64, trf4::KAN, MAT a*	
P_{GAL1}*-NRD1 rrp6Δ*	DLY885	*W303, HIS::Pgal1-NRD1, rrp6::KAN, MAT a*	(Thiebaut et al., 2006)
rap1-2	DLY2207	*W303, rap1-2, MAT a*	(Kurtz and Shore, 1991)
rap1-2 rrp6Δ		*W303, rrp6::KAN, rap1-2, MAT a*	Cette etude.
rsp5-1	DLY2242	*W303, rsp5-1::HIS*	(Harreman et al., 2009)
WT Anchor Away	DLY2548	*W303, tor1-1, fpr1::NAT, RPL13A-2×FKBP12::TRP1, MAT alpha*	(Haruki et al., 2008)
RAP1-FRB Anchor Away		*W303, tor1-1, frp1::NAT, RPL13A-2XFKB12::TRP1 RAP1-FRB-RAP1::LEU, MAT alpha*	D. Shore

III. 2. Vecteurs utilisés

Nom	Description
pDL367	pCM190(*URA3*)-P_{TET}-*HSP104*-P_{GAL1}-*CUP1*
pDL430	pCM190(*TRP1*)-P_{TET}-CUT *IMD2*-P_{GAL1}-*LACZ*
pDL431	pCM190(*TRP1*)-P_{TET}-*HSP104*-X3-*HSP104*-P_{GAL1}-*LACZ*
pDL435	pCM190(*TRP1*)-P_{TET}-*HSP104*-X118-*HSP104*-P_{GAL1}-*LACZ*
pDL468	pCM190 (*URA3*)-P_{TET}-*HSP104-X118-HSP104*-P_{GAL1}-*CUP1*
pDL608	pETM-13-*RAP1*

III. 3. Techniques communes aux trois projets

Condition de culture

A moins que le contraire ne soit mentionné les levures ont été cultivées à 30°C dans un milieu riche (YPDA ou YPA GAL) ou minimal (CSM) dépourvu des acides aminés ou nucléotides appropriés et préparés comme décrit par Guthrie et Fink, 1991. Pour

les tests de croissance sur cuivre les concentrations de $CuSO_4$ ajoutées au milieu de culture (CSM-Ura GAL) sont indiquées sur les figures correspondantes.

PCR

Les différentes constructions ont été validées par PCR en utilisant la polymérase GoTaq (Promega) et les conditions d'amplification étaient : 0,5-1 µM d'amorces, 0,2 mM de dNTP, dénaturation à 94°C, appariement à 55°C, élongation à 72°C et 1 minute/kb dans le tampon réactionnel du fournisseur. Les matrices utilisées pour les sondes de nothern blot marquées en utilisant le fragment de Klenow ont également été amplifiées avec la polymérase GoTaq. Les PCR utilisées pour obtenir les constructions ont été réalisées avec une polymérase à haute fidélité (Phusion, Finzymes) en employant le tampon réactionnel du fournisseur, dénaturation à 98°C, appariement à 55°C, élongation à 72°C et 1 minute/kb.

Transformation de levure

Environ 25 à 50 unités de DO de levures en phase exponentielle de croissance ont été utilisées pour préparer des cellules compétentes pour la transformation. Les cellules ont été centrifugées, lavées deux fois dans 50 mL puis 1 mL d'eau, remises en suspension dans 500 µL de LiCl 100mM et incubées 15 minutes (min) à 30°C. Pour chaque transformation 50 µL de ces cellules ont été utilisées, auxquels ont été ajoutés l'ADN (de 100 à 500 ng de vecteur plasmidique et/ou au moins 200 ng de produit PCR) et 346 µL de solution de transformation (35% de PEG-4000, 100 mM de LiAc, 150 µg/mL d'ADN carrier). Les cellules ont été incubées 30 min à 30°C puis 20 min à 42°C et après centrifugation, remises en suspension dans de l'eau et étalées sur le milieu de culture approprié.

Composition des gels d'agarose et d'acrylamide utilisés

- Gel d'agarose pour northern blot : 1,2% d'agarose (poids/vol) dans un tampon de migration FA (20 mM de MOPS, 5 mM d'acétate de sodium, 1 mM d'EDTA, autoclavé) complété de 0,007 % de formaldéhyde et de 0,2 µg/mL de bromure d'éthidium.
- Gel d'acrylamide pour northern blot : 5% acrylamide/bis-acrylamide (19:1) dans 0,5X de TBE (1X : 89 mM de Tris, 89 mM d'acide borique, 2 mM d'EDTA pH=8), 7,5 M d'urée, 0,0004 % d'APS (poids/vol) et 0,003 % de TEMED (v/v).

- Gel d'acrylamide pour les retards sur gel : 5-8 % d'acrylamide/bis-acrylamide (37:5), dans 0,5X de TBE, 0,0004 % d'APS (poids/vol) et 0,003 % de TEMED (v/v).

III.4. Identification des séquences nécessaires à la terminaison du CUT du locus *IMD2*

Test d'efficacité de terminaison

PCR mutagénique du CUT du locus *IMD2*

Pour introduire des mutations aléatoires dans le CUT du locus *IMD2*, celui-ci a été amplifié par PCR mutagénique en utilisant les amorces 1932 et 1933 sur 0,5 µL de suspension de cellules chauffées à 100°C pendant 10 min. L'enzyme GoTaq a été utilisée (Promega) dans les conditions d'amplification suivantes : amorces 0,5 µg/µL, $MgCl_2$ 6,5 mM, $MnCl_2$ 0,5 mM, dCTP et dTTP 1 mM, dATP et dGTP 0,2 mM, dénaturation à 94°C, appariement à 55°C, élongation à 72°C. Pour éviter l'amplification de produits aspécifiques qui apparaissent lorsque le nombre de cycles d'amplification est trop important, le nombre de cycles optimal était de 20 à 25 cycles.

Test de terminaison

Les séquences mutées par PCR ont ensuite été insérées entre les promoteurs P_{TET} et P_{GAL1} du système rapporteur décrit dans la section résultat (chapitre II.1.2.1.) par recombinaison homologue chez *S. cerevisiae* en utilisant le protocole de transformation de levure décrit précédemment. Les vecteurs ainsi construits ont été extraits de la levure, amplifiés chez *E. coli* et ensuite insérés par transformation dans une souche sauvage et les clones ont été étalés sur un milieu CSM-TRP GAL. Le test d'expression de la β-galactosidase a été effectué par « overlay ».

Test d'expression de la β-galactosidase par overlay

Ce test consiste à recouvrir une boîte de Petri contenant les clones à tester par une fine couche de solution gélosée contenant du X-Gal. Composition de la solution d'overlay pour une boîte : 5 mL d'agarose 1 % à 50°C, 5 mL de tampon K/K_2PO_4 pH=7, 0,1% de SDS 10%, 0,08% de X-Gal remis en suspension dans du DMF. Coloration en quelques minutes, les clones blancs ou bleu pâle ont été isolés, striés et les inserts ont été séquencés.

Oligonucléotides utilisés dans ce projet

Nom	Séquence 5'-3'	Gène	Utilisation
DL292	TAGACACGCAAACACAAATAC	P*TET*	Amplification des CUT du locus *IMD2* sélectionnés dans le système rapporteur.
DL443	GAGGCACATCTGCGTTTCAGGAAC	P*GAL1*	
DL1932	AAATACACACACTAAATTACCGGATCAATTCGGG GGATCCGGTAAAAATTTCGGCTGGAAG	CUT *IMD2*	Clonage du CUT du locus *IMD2* par PCR mutagénique.
DL1933	AGAGAAGGTTTTTTTAGGCTAAGATAATGGGGCT CTTTACGCTAGCTTGGTAGGCTCTTGGTAAAG	CUT *IMD2*	

III. 5. Identification du site de fixation optimal du dimère Nrd1-Nab3 sur l'ARN

SELEX

Purification de l'hétérodimère Nrd1-Nab3

Le protocole de purification de l'hétérodimère Nrd1-Nab3 est présenté dans la publication Porrua et al., 2012.

Transcription *in vitro*

L'ADN de départ (oligonucléotide 1780) et les ADNc issus de chacun des cycles ont été amplifiés par PCR avec les amorces 1557 et 2447 (contenant un promoteur T7). Ces ADN servent de matrice à l'ARN pol T7 (New England Biolabs), la rétrotranscription est effectuée dans les conditions suivantes : tampon de transcription du fournisseur, 1 mM des 4 NTP, 100 unités de T7 RNA pol, 5 µL de matrice ADN produite par PCR, 20-60 µCi d' α-^{32}P-UTP, incubation 60 min à 37°C. A la fin de la transcription une unité de DNase1 a été ajoutée au mélange réaction qui a été incubé 15min à3 7°C. Pour purifier les ARN, les transcrits ont été ensuite migrés sur gel d'acrylamide 5 % puis le gel a été autoradiographiés, permettant ainsi de localiser la bande correspondant aux ARN. Cette bande a été découpée et les ARN ont été extraits du gel par élution sur une nuit dans un tampon adéquat (acétate de sodium pH=5,3 0,3 M, EDTA 1 mM, SDS 0,1 %). Après purification par purification du volume d'élution dans un volume égal de phénol:chlorophorm:isoamyl alcool, 25 :24 :1 (Sigma), précipitation des ARN et dissulotion du culot dans l'eau, la quantité d'ARN purifié a été estimée à l'aide d'un compteur à scintillation.

Liaison du dimère Nrd1-Nab3 aux ARN de séquence aléatoire

Condition de liaison : tampon Tris-HCl pH=7,5 50 mM, NaCl 150 mM, DTT 1 mM, Glycerol 5 %. Au premier cycle, 50 pmoles de Nrd1-Nab3 et 3,8 pmoles d'ARN sont ajoutés à la solution de liaison. Dans les cycles suivant, la quantité de protéine est diminuée pour augmenter la stringence de la sélection. Le ratio de la quantité de Nrd1-Nab3 par rapport à la quantité d'ARN est indiqué dans le tableau ci-dessous. Le mélange de liaison est incubé 20 min à température ambiante puis un tampon de charge contenant 0,1 % de xylène cyanol, 0,1 % de bleu de bromophénol et 5 % de glycérol est ajouté avant migration sur gel.

Cycle	1	2	3	4	5	6	7	8	9	10
Ratio Nrd1-Nab3/ARN	13	13	10	7,5	5.00	2,5	1,75	1.00	0,5	0,5

Retard sur gel

La migration des complexes ARN-protéines a été effectuée dans un gel d'acrylamide 5 % (composition décrite précédemment), à 4°C.

Elution des ARN sélectionnés

Les gels retard ont été autoradiographiés, permettant ainsi de localiser la bande correspondant aux complexes ARN-protéines. Cette bande a été découpée et les ARN ont été extraits de la bande de gel découpée en utilisant un système d'électro-élution Elutrap (Whatman). Les ARN ont été élués dans un tampon TBE 0,5X (pendant une nuit à 100V à température ambiante. Le volume d'élution a été prélevé et le tampon d'élution remplacé pour effectuer une deuxième étape d'élution pendant 1h à 130V. Après purification du volume d'élution dans un volume égal de phénol:chlorophorm:isoamyl alcool, 25 :24 :1 (Sigma), précipitation des ARN et dissolution du culot dans l'eau, la moitié des ARN extraits ont été rétrotranscrits.

Rétrotranscription

La rétrotranscription a été effectuée en utilisant la moitié des ARN extraits, l'amorce 1557 et l'enzyme M-MLV RT (Invitrogen) dans les conditions décrites par le fournisseur. Les ADNc ont ensuite été amplifiés par PCR avec l'enzyme GoTaq en utilisant les amorces 1557 et 2447. Pour l'étape de séquençage, une fraction des

ADNc initiaux et issus des cycles 7 et 10 ont été amplifiés avec les amorces 2063, 2060 (cycle 0), 2061 (cycle 7) et 2062 (cycle 10).

Nombre de séquences obtenues après séquençage à l'issue du SELEX

	Nombre de séquences	
Cycle	Nombre total	Séquences uniques
0	1 231 771	1 230 517
7	1 027 469	1 026 657
10	696 834	656 736

Analyse statistique de l'éspacement entre les sites Nrd1 et Nab3 dans les séquences issues du SELEX

L'analyse statistique de l'espacement entre les sites Nrd1 et Nab3 a été effectuée en utilisant un algorithme permettant de calculer un « score » qui représente l'abondance d'un certain motif dans les séquences issues du SELEX par rapport aux séquences initiales. La méthode utilisée a été décrite par Van Helden et al., 1998, elle est basée sur la comparaison de la fréquences du motif recherché dans les populations de séquences initiales et issues du SELEX. Plus un motif est enrichi à l'issu du SELEX, plus son score est élevé.

Oligonucléotides utilisés dans ce projet

Nom	**Séquence 5'-3'**	**Utilisation**
DL1780	TTTCCTGGATTCCGGCTGTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAGCAGAATTCGGTAAGCGT	Séquence aléatoire de 40 nucléotides
DL1557	ACGCTTACCGAATTCTGCTT	Amplification et rétro-transcription des séquences aléatoires
DL2447	AGCATAGTAATACGACTCACTATAGGGTTTCCTGGATTCCGGCTGTC	Amplification des séquences aléatoires. Contient un promoteur T7.
DL2060	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAATGGATTCCGGCTGTC	Amplification des séquences sélectionnées pour le séquençage.
DL2061	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTGGATTCCGGCTGTC	
DL2062	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTGGATTCCGGCTGTC	
DL2063	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTACCGAATTCTGCTT	

III. 6. Terminaison road-block par Rap1

Sélection de terminateurs à partir de séquences aléatoires

Cette expérience a été menée comme indiqué dans une publication antérieure du laboratoire (Porrua et al., 2012).

Extraction d'ARN

Les cellules ont été cultivées jusqu'en phase exponentielle de croissance puis récoltées (environ 20 unités de DO) et centrifugées. Les cellules ont été remises en suspension dans 400 µL de tampon TES (Tris-HCl pH=7,5 10 mM, EDTA pH=8 5 mM, SDS 1 %) et 400 µL de solution de phénol saturée avec un tampon citrate pH 4,3 (Sigma, P4682) puis incubées pendant 45min à 65°C sous agitation. Le mélange a été centrifugé pendant 10 min à 16000 g à 4°C et le surnageant a été récupéré. L'extraction par phénol a été répétée une deuxième fois suivie d'une purification au chloroforme des échantillons. Les ARN ont été précipités (éthanol 80%, LiCl 60 mM), les culots ont été séchés puis dissous/remis en solution dans de l'eau.

Northern Blot

Pour chaque analyse, 10 µg d'ARN totaux ont été utilisés. Les ARN ont été dénaturés à 85°C dans un tampon de charge (xylène cyanol 0,1 %, bleu de bromophénol 0,1 %, formamide 100%) et ont été migrés soit sur gel d'agarose 1,2 % soit sur gel d'acrylamide 5 % dont les compositions sont décrites précédemment. Après migration, les ARN ont été transférés sur une membrane de nylon chargée positivement (Hybond N+, GE Healthcare), soit par capillarité dans le cas d'un gel d'agarose en utilisant un tampon SSC 20X (20X : NaCl 3 M, trisodium acétate 300 mM, pH=7 ajusté avec du HCl), soit par transfert semi-sec dans le cas d'un gel d'acrylamide. Après transfert les ARN ont été pontés à la membrane par exposition aux UV (254nm).

Les sondes de northern blot ont été préparées soit, 1) en phosphorylant l'extrémité 5' d'un oligonucléotide avec du γ-^{32}P-ATP en utilisant la T4 polynucléotide kinase (NEB, M0201S), soit 2) en marquant un produit PCR avec du α-^{32}P-dCTP en utilisant le fragment de Klenow du kit Amersham Megaprime DNA Labelling System (GE Healthcare) en suivant le protocole des fournisseurs.

Les membranes hybridées avec un oligonucléotide marqué ont été pré-hybridées pendant 1 heure à 42°C dans un tampon d'hybridation (Ultrahyb-oligo, Ambion). Les membranes ont été incubées avec l'oligonucléotide marqué pendant une nuit, à 42°C et lavées deux fois pendant 30min à 42°C avec une solution SSC 0,5X, SDS 0,1 %.
Les membranes hybridées avec un produit PCR ont été pré-hybridées 1 heure à 42°C dans un tampon d'hybridation (UltraHyb, Ambion). La sonde a été dénaturée 5 min à 95°C et ajoutée au tampon d'hybridation. Les membranes ont été incubées pendant une nuit à 42°C et lavées deux fois avec une solution SSC 2X, SDS 0,1 %, pendant 5 min à 42°C et deux fois avec une solution SSC 0,1X, SDS 0,1 % pendant 15 min à 42°C.
La révélation des northern blot a été effectuée en utilisant un Storm 820 phosphorimager (Amersham).

Sélection d'ARN polyadénylés

La sélection des fractions d'ARN polyadénylés et non polyadénylés a été éffectuée en utilisant des billes magnétiques couplées à des oligo dT^{25} (Dynabeads, Invitrogen). 30 µg d'ARN totaux ont été utilisés pour chaque sélection. Les ARN ont été remis en solution dans un volume final de 100 µL d'eau, dénaturés pendant 5 min à 70°C puis incubés 5 min dans la glace. 50 µL de suspension de billes ont été lavés deux fois dans 50 µL de tampon de liaison (Tris-Cl pH=7,5 20 mM, LiCl 1 M, EDTA 2 mM) puis les 50 µL de suspension de billes ont été ajoutés à la solutiond'ARN. L'incubation a été réalisée pendant 20 min à 30°C sous agitation puis le surnageant contenant la fraction d'ARN non polyadénylés a été prélevé et précipité avec 3-4 volumes d'éthanol 95% (la quantité de sel du tampon de liaison est suffisante). Les billes contenant la fraction d'ARN polyadénylés ont été lavées deux fois avec 200 µL de tampon de lavage (Tris-Cl pH=8 10 mM, LiCl 0,15 M, EDTA 1 mM) et les ARN ont été élués deux fois par une incubation à 85°C pendant 10 min avec 10 µL d'eau . Les échantillons ont été déshydratés puis remis en solution dans 10 µL de tampon de charge (xylène cyanol 0,1 %, bleu de bromophénol 0,1 %, formamide 100%). Des echantillons de 10-15 µg d'ARN ont été déposés sur gel pour les analyses par northern blot.

RNAseH

10 µg d'ARN totaux ont été digérés par de la RNase H (Invitrogen) en utilisant 2 unités d'enzymes et 40 pmol d'oligo par échantillon à 37°C pendant 45 min dans les conditions réactionnelles préconisées par le fabricant/ selon les recommandations du fabricant. (tampon réactionnel : Tris pH=8 50 mM, KCl 75 mM, $MgCl_2$ 3 mM, DTT 10 mM).

Après digestion, l'enzyme a été éliminée par une extraction des ARN avec du phénol/chloroforme 5/1 pH=4,7 (Sigma) et les ARN ont été précipités comme décrit dans la section « extraction d'ARN ».

Purification de la protéine rRap1

La protéine Rap1 recombinante (6xHis-TEV-Rap1) a été purifiée chez *E. coli* (souche BL21/DE3) à partir d'un vecteur d'expression pETM-13 fourni par M-H LeDu (Matot et al., 2012). Les bactéries contenant ce plasmide ont été pré-cultivées à 37°C dans un milieu LB supplémenté en chloramphénicol et kanamycine puis lorsque la culture a atteint une DO d'environ 0,7 elle a été transférée à 20°C. Lorsque la DO=1 a été atteinte, l'expression de Rap1 a été induite par l'ajoutd'isopropyl β-D-1-thiogalactopyranoside (IPTG) (concentration finale 0,5 mM) et laissée une nuit à 20°C. La culture a été arrêtée à DO=4 et les cellules ont été centrifugées à 10krpm pendant 15 min. Le culot a été remis en suspension dans 12 mL de tampon de lyse (Na_2HPO_4 pH=7,5 50 mM, NaCl 1 M, imidazole pH=7,5 20 mM, AEBSF 0,2 mM, Glycérol 5 %) et lysées par sonication (sonicateur vibracell 75042, 500W, Bioblock scientific) avec une sonde d'1 cm (S&M 220) et en utilisant 11 pulses de 20 sec à 20-25% d'amplitude entrecoupés de 40 sec de pause dans la glace. L'extrait a été centrifugé à 13krpm à 4°C. Rap1 a été purifiée en utilisant un FPLC AKTA Purifier (GE Healthcare) en deux étapes. 1) l'extrait a été passé sur une colonne His-Trap (GE Healthcare) dans laquelle les protéines ont été éluées dans un gradient d'imidazole de 50 mM à 300 mM créé en mélangeant de deux tampons(tampon A : Tris-HCl pH 7,5 50 mM, NaCl 500mM, glycerol 5% ; tanpom B : Tris-HCl pH 7,5 50 mM, NaCl 500mM, glycerol 5%, imidazole 500mM) . 2) Les fractions contenant la plus grande concentration de Rap1 (estimée par SDS-PAGE et coloration au bleu de Coomassie (EZblue kit, Sigma)) ont ensuite été purifiées par filtration sur gel en utilisant une colonne Superdex 200 (GE Healthcare) et un tampon de filtration (Na_2HPO_4 pH=7,5 50 mM, NaCl 1 M, Glycerol 10 %, imidazole 50 mM). Les fractions

les plus concentrées en Rap1 (estimation par SDS-PAGE et coloration au bleu de Coomassie) ont été conservées et la concentration de Rap1 a été dosée par la méthode de Bradford (Quick Start Bradford Proatein Assay, Bio-Rad). Du glycérol a été ajouté aux échantillons pour atteindre une concentration finale de 20 % et ils ont été congelés dans de l'azote liquide et conservés à -80°C.

Retard sur gel

La liaison des complexes Rap1-oligonucléotides a été réalisée dans les conditions suivantes : Na_2HPO_4/NaH_2PO_4 pH=8 10 mM, NaCl 100 mM, Glycerol 5 %, les concentrations en protéine et en sonde sont indiquées sur la figure correspondante. La migration des complexes Rap1/acides nucléiques a été effectuée dans un gel d'acrylamide 5 % (composition décrite précédemment), à 4°C. Les séquences des sondes oligonucléotidiques utilisées sont indiquées dans le tableau final (oligos DL3001, DL3002 et DL3001 ARN).

Déplétion nucléaire de Rap1 par la technique Anchor away

Le système anchor away est décrit dans Haruki et al., 2008. La souche contenant la protéine Rap1-FRB a été fournie par D. Shore. L'étiquette FRB est intégrée dans la séquence d'acide aminée de Rap1 à la position 135. Pour purifier des ARN à partir de cette souche, ces cellules ont été cultivées dans 100 mL de milieu CSM-Trp jusqu'à atteindre une DO=0,2-0,3 puis séparées en deux cultures de 50 mL. Rap1 a été déplété du noyau par l'ajout de rapamycine (1 µg/mL final) dans l'une des deux cultures et après 2 h à 30°C, les cellules ont été récoltées par centrifugation et les ARN extraits comme décrit précédemment.

Rétrotranscription et PCR quantitative

Pour analyser la quantité de transcrits par qPCR, les ARN ont été rétrotranscrits en utilisant l'enzyme M-MLV RT (Invitrogen) dans les conditions décrites par le fournisseur. Les amorces utilisées pour la rétrotranscription de chacun des transcrits analysés correspondent aux amorces antisens utilisées pour la qPCR.

Les qPCR ont été effectuées avec un LightCycler 480 (Roche), en suivant le Kit LightCycler SYBR Green I Master (Roche) et le protocole du fournisseur (dénaturation à 95°C, hybridation à 52°C, élongation à 72°C).

Analyse de la pause transcriptionnelle au niveau de métasites Rap1/Abf1

La pause de l'ARN pol II provoquée par Rap1 et Abf1 a été analysée en utilisant des données de positionnement au nucléotide près de la polymérase à l'échelle du génome obtenues par PAR-CLIP (Schaughency et al., 2014) ou NET-Seq dans des souches sauvage ou *dst1Δ* (Churchman and Weissman, 2011). Respectivement 67 et 151 sites liés par Rap1 et d'Abf1 ont été définis en utilisant les données de la littérature (Harbison et al., 2004; MacIsaac et al., 2006). Dans le cas de la figure 2.3.21., le motif de fixation de Rap1 était défini par la séquence ACACCCANAC retrouvée 271 fois dans le génome.

Pour analyser la pause de l'ARN pol II à chaque site de fixation de ces facteurs, les pics locaux de l'ARN pol II ont été recherchés dans une fenêtre de 200 nt en amont des sites. Un pic était défini par une valeur de « read » supérieur à la moyenne plus deux fois l'écart type calculés à partir de tous les résultats de séquençage non nuls dans une fenêtre de 200 nt. Toutes les fenêtres qui contenaient moins de trois reads non nuls ont été exclues de cette analyse, engendrant un total de 151 sites Abf1 pris en compte pour la figure 2.3.21 (Analyse globale de la pause de l'ARN pol II en amont des sites de fixation d'Abf1), ainsi que 96 (WT) et 98 (*dst1Δ*) sites ACACCCANAC pris en compte dans les souches WT et *dst1,* respectivement, pour la figure 2.3.16 et 68 sites Rap1 pris en compte pour la figure 2.3.15. Afin de rendre ces analyses comparables, le nombre de pics détectés dans la fenêtre de 200 nt a été divisé par le nombre total de « read » pour chaque position, ce qui donne la fréquence de pic.

Oligonucléotides utilisés dans ce projet

Nom	Séquence 5'-3'	Gène	Utilisation
DL190	TTGAGCCAACGTCAAAATCGTTAGAGCCCTTTCTGTAAATT GCGTTTGGTCGTTCAT	*HSP104*	Sonde PNK
DL2627	ATTCAAAAGCGAACACCGAATTGACCATGAGGAGACGGTC TGGTTTAT	*U4*	Sonde PNK
DL377	ATGTTCCCAGGTATTGCCGA	*ACT1*	Sonde Klenow (PCR)
DL378	ACACTTGTGGTGAACGATAG	*ACT1*	
DL1702	TTCGGATCATCAACATCCACAATTACAACCTATACATATTCT AGCTGCCTTCATTGAAACG	*HSP104*	Clonage
DL1666	GAATCTTTCTTCGAAATCACCTTTGTATTTAGCACCTGCGG TTAATGCGG	*HSP104*	Clonage
DL2099	ATTACAACCTATACATATTCTAGCTGCCTTCATTGAAACG**ACCCAGAC A**		Clonage du site Rap1 court (gras)
DL2100	TCGAAATCACCTTTGTATTTAGCACCTGCGGTTAATGCGG**TGTCTGGG T**CGTTT		
DL2128	ACAACCTATACATATTCTAGCTGCCTTCATTGAAACG**GTAAACCCAGA CA**		Clonage du site Rap1 étendu (gras)
DL2129	CACCTTTGTATTTAGCACCTGCGGTTAATGCGG**TGTCTGGGTTTAC**CG TTT		
DL2052	CTTCGAAATCACCTTTGTATTTAGCACCTGCGGTTAATGCGGCCAGAA ATTGGTTACAGCTCGTGTG		Clonage de la séquence X118 ΔBS
DL3001	ACACACGAGCTGTAAACCCAGACACCAATTTCTGGCCGCA		ADN utilisé pour le test d'affinité de Rap1
DL3002	TGCGGCCAGAAATTGGTGTCTGGGTTTACAGCTCGTGTGT		Brin d'ADN complémentaire utilisé pour le test d'affinité de Rap1
DL3001 ARN	ACACACGAGCUGUAAACCCAGACACCAAUUUCUGGCCGCA		ARN utilisé pour le test d'affinité de Rap1
DL2997	ATACATATTCTAGCTGCCTTCATTGAAACGCTGCAACACCC AGACCTTGCTGCAGTGATATCATCCACCCATACCCCTCGA CCGCATTAACCGCAGGTGC	*uHYP2*	Clonage des deux sites Rap1 du locus *uHYP2*
DL2998	ATACATATTCTAGCTGCCTTCATTGAAACGCTGCAACCCAG ACCCATTGCTGCAGTGATATCATCCACCCATACCCCTCGA CCGCATTAACCGCAGGTGC	*uHYP2*	Clonage des deux sites Rap1 du locus *uHYP2*. Site A muté.
DL2999	ATACATATTCTAGCTGCCTTCATTGAAACGCTGCAACACCC AGACCTTGCTGCAGTGATATCATCCCCATACCCACCTCGA CCGCATTAACCGCAGGTGC	*uHYP2*	Clonage des deux sites Rap1 du locus *uHYP2*. Site B muté.

DL3000	ATACATATTCTAGCTGCCTTCATTGAAACGCTGCAACCCAG ACCCATTGCTGCAGTGATATCATCCCCATACCCACCTCGA CCGCATTAACCGCAGGTGC	*uHYP2*	Clonage des deux sites Rap1 du locus *uHYP2*. Site A et B mutés.
DL3249	AACGGGAACGGCGACTTG	*uHYP2*	qPCR
DL3248	AGCGTCCAGCTACAGCGT	*uHYP2*	
DL3199	CCTCGATGTATTCCGTAG	*uHYP2*	
DL3198	TGTCGCCTCACACGGACC	*uHYP2*	
DL3150	ATACATATTCTAGCTGCCTTCATTGAAACGAGAAGCAATCA TGCC	*dSTF1*	Clonage du site Rap1 du locus *dSTF1*
DL3151	CTTTGTATTTAGCACCTGCGGTTAATGCGGTGGTATTACAT TGGA	*dSTF1*	
DL3243	TATAGTATATACGTCACG	*dSTF1*	qPCR
DL3208	CTAACTGACGGTGTTCCT	*dSTF1*	
DL3244	TCTTACACAAAACATCCG	*dSTF1*	
DL3245	GCGTTTCATATAATCCTTC	*dSTF1*	
DL1360	ATCCCCCGAATTGATCCGG	*dSTF1*	

Séquences étudiées dans ce projet

Nom	**Séquences**
X118	AAACGTTAGCGGATACCGGTCACCGCATTCGTCGCGTAACAATCTGCGCACGC TCAACACACGAGCTGT**AAACCCAGACA**CCAATTTCTGGCCGCATTAA
X108	AAACGCTCGCGAGTTATCATACATTCAAGATTATTACAGTTCGGCATGATTCTAT GAGGGATACCGGTCACCGCATTCGTAAAATAAACCCGAACCAGTGATGCCCGA CGAGA**ACACCCAGACA**TCACTACAGGCCCTCCCGCATTAA
X84	AAACGAAAAGACTAATATGGTGTGTTTGGTGGGTTGGCCATTGGAAATCAGAAC CGCGCGATACCGGTCACCGCATTCGCTTAAACTTCAAA**ACACCCACACA**CCAG CATAGGAAAATTCCGGTGACTTTACCCTGACACCGCATTAA
X3	AAACGCGCAGGTATTATAGCGATACCGGTCACCGCATTCGCAAGGACTATAGTT CATTTA**GTTACCCGG**GTCGGGGTGCCTTCCGACCAGGCCTAACGGTCCGCATT AA
X9	AAACGTAAAATCGACTTGACCAGCTTAATATCTATACTGTTCGACTTATAAATCAT TCTTGCACTGATACCGGTCACCGCATTCGTAGTAGTTAAACTTTTTACCGAAAGA TAAGTCAATAAGGTCTATTAATCTTTTCCAGCCCGCATTAA

Rouge : site de fixation de Rap1

Bleu : site de fixation de Reb1

IV. Références bibliographiques

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572–576.

Allen, N.P., Huang, L., Burlingame, A., and Rexach, M. (2001). Proteomic analysis of nucleoporin interacting proteins. J. Biol. Chem. *276*, 29268–29274.

Allison, D.S., and Hall, B.D. (1985). Effects of alterations in the 3' flanking sequence on in vivo and in vitro expression of the yeast SUP4-o tRNATyr gene. EMBO J. *4*, 2657–2664.

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999). The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases. Genes Dev. *13*, 2148–2158.

Amrani, N., Minet, M., Wyers, F., Dufour, M.E., Aggerbeck, L.P., and Lacroute, F. (1997). PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I. Mol. Cell. Biol. *17*, 1102–1109.

Anderson, J.R., Mukherjee, D., Muthukumaraswamy, K., Moraes, K.C.M., Wilusz, C.J., and Wilusz, J. (2006). Sequence-specific RNA binding mediated by the RNase PH domain of components of the exosome. RNA N. Y. N *12*, 1810–1816.

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. Mol. Cell. Biol. *13*, 2730–2741.

Arcus, V.L., Bäckbro, K., Roos, A., Daniel, E.L., and Baker, E.N. (2004). Distant structural homology leads to the functional characterization of an archaeal PIN domain as an exonuclease. J. Biol. Chem. *279*, 16471–16478.

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

Arigo, J.T., Carroll, K.L., Ames, J.M., and Corden, J.L. (2006b). Regulation of yeast NRD1 expression by premature transcription termination. Mol. Cell *21*, 641–651.

Ashfield, R., Patel, A.J., Bossone, S.A., Brown, H., Campbell, R.D., Marcu, K.B., and Proudfoot, N.J. (1994). MAZ-dependent termination between closely spaced human complement genes. EMBO J. *13*, 5656–5667.

Auld, K.L., Brown, C.R., Casolari, J.M., Komili, S., and Silver, P.A. (2006). Genomic association of the proteasome demonstrates overlapping gene regulatory activity with transcription factor substrates. Mol. Cell *21*, 861–871.

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

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., et al. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol. Cell *32*, 878–887.

Van Bakel, H., Tsui, K., Gebbia, M., Mnaimneh, S., Hughes, T.R., and Nislow, C. (2013). A Compendium of Nucleosome and Transcript Profiles Reveals Determinants of Chromatin Architecture and Transcription. PLoS Genet. *9*.

Barabino, S.M., Hübner, 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 Dev. *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. EMBO J. *19*, 3778–3787.

Barillà, 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. Proc. Natl. Acad. Sci. U. S. A. *98*, 445–450.

Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. Cell *116*, 699–709.

Belitsky, B.R., and Sonenshein, A.L. (2011). Roadblock repression of transcription by Bacillus subtilis CodY. J. Mol. Biol. *411*, 729–743.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science *301*, 1090–1093.

Bernstein, J., Patterson, D.N., Wilson, G.M., and Toth, E.A. (2008). Characterization of the essential activities of Saccharomyces cerevisiae Mtr4p, a 3'->5' helicase partner of the nuclear exosome. J. Biol. Chem. *283*, 4930–4942.

Berretta, J., Pinskaya, M., and Morillon, A. (2008). A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. Genes Dev. *22*, 615–626.

Berthiau, A.-S., Yankulov, K., Bah, A., Revardel, E., Luciano, P., Wellinger, R.J., Géli, V., and Gilson, E. (2006). Subtelomeric proteins negatively regulate telomere elongation in budding yeast. EMBO J. *25*, 846–856.

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.

De Boer, C.G., and Hughes, T.R. (2012). YeTFaSCo: a database of evaluated yeast transcription factor sequence specificities. Nucleic Acids Res. *40*, D169–D179.

Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009). The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. Cell *139*, 547–559.

Braglia, P., Percudani, R., and Dieci, G. (2005). Sequence context effects on oligo(dT) termination signal recognition by Saccharomyces cerevisiae RNA polymerase III. J. Biol. Chem. *280*, 19551–19562.

Braglia, P., Heindl, K., Schleiffer, A., Martinez, J., and Proudfoot, N.J. (2010). Role of the RNA/DNA kinase Grc3 in transcription termination by RNA polymerase I. EMBO Rep. *11*, 758–764.

Bratkovič, T., and Rogelj, B. (2014). The many faces of small nucleolar RNAs. Biochim. Biophys. Acta *1839*, 438–443.

Bregman, D.B., Halaban, R., van Gool, A.J., Henning, K.A., Friedberg, E.C., and Warren, S.L. (1996). UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells. Proc. Natl. Acad. Sci. U. S. A. *93*, 11586–11590.

Briggs, M.W., Burkard, K.T., and Butler, J.S. (1998). Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. J. Biol. Chem. *273*, 13255–13263.

Briggs, S.D., Xiao, T., Sun, Z.-W., Caldwell, J.A., Shabanowitz, J., Hunt, D.F., Allis, C.D., and Strahl, B.D. (2002). Gene silencing: Trans-histone regulatory pathway in chromatin. Nature *418*, 498–498.

Buck, S.W., and Shore, D. (1995). Action of a RAP1 carboxy-terminal silencing domain reveals an underlying competition between HMR and telomeres in yeast. Genes Dev. *9*, 370–384.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549–561.

Cairns, B.R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B., and Kornberg, R.D. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell *87*, 1249–1260.

Calvo, O., and Manley, J.L. (2005). The transcriptional coactivator PC4/Sub1 has multiple functions in RNA polymerase II transcription. EMBO J. *24*, 1009–1020.

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007). Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell *131*, 706–717.

Carey, M., Li, B., and Workman, J.L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. Mol. Cell *24*, 481–487.

Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D., and Corden, J.L. (2004). Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. Mol. Cell. Biol. *24*, 6241–6252.

Carroll, K.L., Ghirlando, R., Ames, J.M., and Corden, J.L. (2007). Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. RNA *13*, 361–373.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M.P., et al. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581–592.

Castelnuovo, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F., and Zenklusen, D. (2013). Bimodal expression of PHO84 is modulated by early termination of antisense transcription. Nat. Struct. Mol. Biol. *20*, 851–858.

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.

Chen, J., and Moore, C. (1992). Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. Mol. Cell. Biol. *12*, 3470–3481.

Chen, S., and Hyman, L.E. (1998). A specific RNA-protein interaction at yeast polyadenylation efficiency elements. Nucleic Acids Res. *26*, 4965–4974.

Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski- Wills, J.-C., Nilges, M., et al. (2010). Architecture of the RNA polymerase II–TFIIF complex revealed by cross- linking and mass spectrometry. EMBO J. *29*, 717–726.

Cheung, A.C.M., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249–253.

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.

Chintakayala, K., Singh, S.S., Rossiter, A.E., Shahapure, R., Dame, R.T., and Grainger, D.C. (2013). E. coli Fis protein insulates the cbpA gene from uncontrolled transcription. PLoS Genet. *9*, e1003152.

Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: The exosome. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1829*, 552–560.

Choi, S.-K., and Saier, M.H. (2005). Regulation of sigL expression by the catabolite control protein CcpA involves a roadblock mechanism in Bacillus subtilis: potential connection between carbon and nitrogen metabolism. J. Bacteriol. *187*, 6856–6861.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

Clapier, C.R., and Cairns, B.R. (2009). The Biology of Chromatin Remodeling Complexes. Annu. Rev. Biochem. *78*, 273–304.

Colin, J., Libri, D., and Porrua, O. (2011). Cryptic Transcription and Early Termination in the Control of Gene Expression. Genet. Res. Int. *2011*, e653494.

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. Mol. Cell *56*, 667–680.

Connelly, S., and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. *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.

Corden, J.L. (2013). The RNA Polymerase II C-terminal Domain: Tethering Transcription to Transcript and Template. Chem. Rev. *113*, 8423–8455.

Cormack, B.P., and Struhl, K. (1992). The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. Cell *69*, 685–696.

Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., et al. (2008). Structure of eukaryotic RNA polymerases. Annu. Rev. Biophys. *37*, 337–352.

Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet. *7*, e1002329.

Cristodero, M., Böttcher, B., Diepholz, M., Scheffzek, K., and Clayton, C. (2008). The Leishmania tarentolae exosome: purification and structural analysis by electron microscopy. Mol. Biochem. Parasitol. *159*, 24–29.

De la Cruz, J., Kressler, D., Tollervey, D., and Linder, P. (1998). Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae. EMBO J. *17*, 1128–1140.

Daniel, J.A., Torok, M.S., Sun, Z.-W., Schieltz, D., Allis, C.D., Yates, J.R., and Grant, P.A. (2004). Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. J. Biol. Chem. *279*, 1867–1871.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. Proc. Natl. Acad. Sci. U. S. A. *103*, 5320–5325.

Davis, C.A., and Ares, M. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. *103*, 3262–3267.

Deminoff, S.J., and Santangelo, G.M. (2001). Rap1p requires Gcr1p and Gcr2p homodimers to activate ribosomal protein and glycolytic genes, respectively. Genetics *158*, 133–143.

Deuschle, U., Gentz, R., and Bujard, H. (1986). lac Repressor blocks transcribing RNA polymerase and terminates transcription. Proc. Natl. Acad. Sci. U. S. A. *83*, 4134–4137.

Dheur, S., Vo, L.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. EMBO J. *22*, 2831–2840.

Dichtl, B., and Keller, W. (2001). Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. EMBO J. *20*, 3197–3209.

Dichtl, B., Blank, D., Sadowski, M., Hübner, W., Weiser, S., and Keller, W. (2002a). Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. EMBO J. *21*, 4125–4135.

Dichtl, B., Blank, D., Ohnacker, M., Friedlein, A., Roeder, D., Langen, H., and Keller, W. (2002b). A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. Mol. Cell *10*, 1139–1150.

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 N. Y. N *10*, 965–977.

Van Dijk, E.L., Chen, C.L., d' Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114–117.

Dikstein, R. (2011). The unexpected traits associated with core promoter elements. Transcription *2*, 201–206.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. (2009). RNAi in budding yeast. Science *326*, 544–550.

Drinnenberg, I.A., Fink, G.R., and Bartel, D.P. (2011). Compatibility with Killer Explains the Rise of RNAi-Deficient Fungi. Science *333*, 1592–1592.

Dunn, E.F., Hammell, C.M., Hodge, C.A., and Cole, C.N. (2005). Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. Genes Dev. *19*, 90–103.

Egloff, S., Dienstbier, M., and Murphy, S. (2012). Updating the RNA polymerase CTD code: adding gene-specific layers. Trends Genet. *28*, 333–341.

Eichner, J., Chen, H.-T., Warfield, L., and Hahn, S. (2010). Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. EMBO J. *29*, 706–716.

Emre, N.C.T., Ingvarsdottir, K., Wyce, A., Wood, A., Krogan, N.J., Henry, K.W., Li, K., Marmorstein, R., Greenblatt, J.F., Shilatifard, A., et al. (2005). Maintenance of Low Histone Ubiquitylation by Ubp10 Correlates with Telomere-Proximal Sir2 Association and Gene Silencing. Mol. Cell *17*, 585–594.

Evers, R., and Grummt, I. (1995). Molecular coevolution of mammalian ribosomal gene terminator sequences and the transcription termination factor TTF-I. Proc. Natl. Acad. Sci. U. S. A. *92*, 5827–5831.

Ezeokonkwo, C., Ghazy, M.A., Zhelkovsky, A., Yeh, P.-C., and Moore, C. (2012). Novel interactions at the essential N-terminus of poly(A) polymerase that could regulate poly(A) addition in Saccharomyces cerevisiae. FEBS Lett. *586*, 1173–1178.

Falk, S., Weir, J.R., Hentschel, J., Reichelt, P., Bonneau, F., and Conti, E. (2014). The Molecular Architecture of the TRAMP Complex Reveals the Organization and Interplay of Its Two Catalytic Activities. Mol. Cell *55*, 856–867.

Feldmann, E.A., and Galletto, R. (2014). The DNA-Binding Domain of Yeast Rap1 Interacts with dsDNA in Multiple Binding Modes. Biochemistry (Mosc.).

Fernández-Tornero, C., Böttcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W.H., Sentenac, A., Müller, C.W., and Schoehn, G. (2007). Insights into Transcription Initiation and Termination from the Electron Microscopy Structure of Yeast RNA Polymerase III. Mol. Cell *25*, 813–823.

Flynn, R.A., Almada, A.E., Zamudio, J.R., and Sharp, P.A. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proc. Natl. Acad. Sci. U. S. A. *108*, 10460–10465.

Fourel, G., Miyake, T., Defossez, P.-A., Li, R., and Gilson, E. (2002). General regulatory factors (GRFs) as genome partitioners. J. Biol. Chem. *277*, 41736–41743.

Freeman, K., Gwadz, M., and Shore, D. (1995). Molecular and genetic analysis of the toxic effect of RAP1 overexpression in yeast. Genetics *141*, 1253–1262.

Ganapathi, M., Palumbo, M.J., Ansari, S.A., He, Q., Tsui, K., Nislow, C., and Morse, R.H. (2011). Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. Nucleic Acids Res. *39*, 2032–2044.

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. EMBO J. *22*, 1588–1598.

Garbett, K.A., Tripathi, M.K., Cencki, B., Layer, J.H., and Weil, P.A. (2007). Yeast TFIID Serves as a Coactivator for Rap1p by Direct Protein-Protein Interaction. Mol. Cell. Biol. *27*, 297–311.

Gavin, A.-C., Bösche, 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., 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. Mol. Cell. Biol. *29*, 2296–2307.

Ghazy, M.A., Gordon, J.M.B., 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 Res. *40*, 1214–1225.

Gillette, T.G., Gonzalez, F., Delahodde, A., Johnston, S.A., and Kodadek, T. (2004). Physical and functional association of RNA polymerase II and the proteasome. Proc. Natl. Acad. Sci. U. S. A. *101*, 5904–5909.

Gilson, E., Roberge, M., Giraldo, R., Rhodes, D., and Gasser, S.M. (1993). Distortion of the DNA double helix by RAP1 at silencers and multiple telomeric binding sites. J. Mol. Biol. *231*, 293–310.

Goh, W.S., Orlov, Y., Li, J., and Clarke, N.D. (2010). Blurring of High-Resolution Data Shows that the Effect of Intrinsic Nucleosome Occupancy on Transcription Factor Binding is Mostly Regional, Not Local. PLoS Comput Biol *6*, e1000649.

Gong, X.Q., Zhang, C., Feig, M., and Burton, Z.F. (2005). Dynamic Error Correction and Regulation of Downstream Bubble Opening by Human RNA Polymerase II. Mol. Cell *18*, 461–470.

Goodrich, J.A., and Kugel, J.F. (2006). Non-coding-RNA regulators of RNA polymerase II transcription. Nat. Rev. Mol. Cell Biol. *7*, 612–616.

Gordân, R., Hartemink, A.J., and Bulyk, M.L. (2009). Distinguishing direct versus indirect transcription factor–DNA interactions. Genome Res. *19*, 2090–2100.

Gordon, J.M.B., 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 (Mosc.) *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 Res. *27*, 888–894.

Grant, P.A., Duggan, L., Côté, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., et al. (1997). Yeast Gcn5 functions in

two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev. *11*, 1640–1650.

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. J. Biol. Chem. *277*, 7752–7760.

Greenblatt, J., Nodwell, J.R., and Mason, S.W. (1993). Transcriptional antitermination. Nature *364*, 401–406.

Gross, S., and Moore, C. (2001a). Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. Proc. Natl. Acad. Sci. U. S. A. *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. Mol. Cell. Biol. *21*, 8045–8055.

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

Gu, W., and Reines, D. (1995a). Identification of a decay in transcription potential that results in elongation factor dependence of RNA polymerase II. J. Biol. Chem. *270*, 11238–11244.

Gu, W., and Reines, D. (1995b). Variation in the Size of Nascent RNA Cleavage Products as a Function of Transcript Length and Elongation Competence. J. Biol. Chem. *270*, 30441–30447.

Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nat. Struct. Mol. Biol. *15*, 786–794.

Gudipati, R.K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L.M., Jacquier, A., and Libri, D. (2012). Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells. Mol. Cell *48*, 409–421.

Guo, Z., and Sherman, F. (1996). 3'-end-forming signals of yeast mRNA. Trends Biochem. Sci. *21*, 477–481.

C. Guthrie and G. Fink, Eds. (1991). Guide to Yeast Genetics and Molecular Biology. Volume 194 of Methods in Enzymology, Academic Press.

Hage, A.E., Koper, M., Kufel, J., and Tollervey, D. (2008). Efficient termination of transcription by RNA polymerase I requires the 5′ exonuclease Rat1 in yeast. Genes Dev. *22*, 1069–1081.

Hainer, S.J., Pruneski, J.A., Mitchell, R.D., Monteverde, R.M., and Martens, J.A. (2011). Intergenic transcription causes repression by directing nucleosome assembly. Genes Dev. *25*, 29–40.

Hamill, S., Wolin, S.L., and Reinisch, K.M. (2010). Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. Proc. Natl. Acad. Sci. U. S. A. *107*, 15045–15050.

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. Mol. Cell. Biol. *22*, 6441–6457.

Hao, N., Krishna, S., Ahlgren-Berg, A., Cutts, E.E., Shearwin, K.E., and Dodd, I.B. (2014). Road rules for traffic on DNA-systematic analysis of transcriptional roadblocking in vivo. Nucleic Acids Res. *42*, 8861–8872.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.-B., Reynolds, D.B., Yoo, J., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99–104.

Hardy, C.F., Balderes, D., and Shore, D. (1992a). Dissection of a carboxy-terminal region of the yeast regulatory protein RAP1 with effects on both transcriptional activation and silencing. Mol. Cell. Biol. *12*, 1209–1217.

Hardy, C.F., Sussel, L., and Shore, D. (1992b). A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. Genes Dev. *6*, 801–814.

Harreman, M., Taschner, M., Sigurdsson, S., Anindya, R., Reid, J., Somesh, B., Kong, S.E., Banks, C.A.S., Conaway, R.C., Conaway, J.W., et al. (2009). Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. Proc. Natl. Acad. Sci. *106*, 20705–20710.

Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell *137*, 445–458.

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol. Cell *31*, 925–932.

He, B., and Zalkin, H. (1992). Repression of Escherichia coli purB is by a transcriptional roadblock mechanism. J. Bacteriol. *174*, 7121–7127.

He, X., and Moore, C. (2005). Regulation of yeast mRNA 3' end processing by phosphorylation. Mol. Cell *19*, 619–629.

He, X., Khan, A.U., Cheng, H., Pappas, D.L., Hampsey, M., and Moore, C.L. (2003). Functional interactions between the transcription and mRNA 3′ end processing machineries mediated by Ssu72 and Sub1. Genes Dev. *17*, 1030–1042.

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. EMBO J. *21*, 1800–1810.

Helmling, S., Zhelkovsky, A., and Moore, C.L. (2001). Fip1 regulates the activity of Poly(A) polymerase through multiple interactions. Mol. Cell. Biol. *21*, 2026–2037.

Henry, K.W., Wyce, A., Lo, W.-S., Duggan, L.J., Emre, N.C.T., Kao, C.-F., Pillus, L., Shilatifard, A., Osley, M.A., and Berger, S.L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev. *17*, 2648–2663.

Henry, M., Borland, C.Z., Bossie, M., and Silver, P.A. (1996). Potential RNA binding proteins in Saccharomyces cerevisiae identified as suppressors of temperature-sensitive mutations in NPL3. Genetics *142*, 103–115.

Heride, C., Urbé, S., and Clague, M.J. (2014). Ubiquitin code assembly and disassembly. Curr. Biol. CB *24*, R215–R220.

Hirata, A., Klein, B.J., and Murakami, K.S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. Nature *451*, 851–854.

Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S., and Stefl, R. (2011). Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. J. Biol. Chem. *286*, 3645–3657.

Holstege, F.C., Fiedler, U., and Timmers, H.T. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. EMBO J. *16*, 7468–7480.

Holub, P., Lalakova, J., Cerna, H., Pasulka, J., Sarazova, M., Hrazdilova, K., Arce, M.S., Hobor, F., Stefl, R., and Vanacova, S. (2012). Air2p is critical for the assembly and RNA-binding of the TRAMP complex and the KOW domain of Mtr4p is crucial for exosome activation. Nucleic Acids Res. *40*, 5679–5693.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and Bradbury, E.M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 “tail” to DNA. J. Biol. Chem. *268*, 305–314.

Hoof, A. van, Lennertz, P., and Parker, R. (2000). Yeast Exosome Mutants Accumulate 3′-Extended Polyadenylated Forms of U4 Small Nuclear RNA and Small Nucleolar RNAs. Mol. Cell. Biol. *20*, 441–452.

Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C., and Tollervey, D. (2006). Microarray detection of novel nuclear RNA substrates for the exosome. Yeast Chichester Engl. *23*, 439–454.

Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., and Vogelauer, M. (2008). A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Mol. Cell *32*, 685–695.

Hsin, J.-P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. *26*, 2119–2137.

Huisinga, K.L., and Pugh, B.F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Mol. Cell *13*, 573–585.

Idrissi, F.-Z., Garcia-Reyero, N., Fernandez-Larrea, J.B., and Piña, B. (2001). Alternative Mechanisms of Transcriptional Activation by Rap1p. J. Biol. Chem. *276*, 26090–26098.

Iyer, V., and Struhl, K. (1995). Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO J. *14*, 2570–2579.

Izban, M.G., and Luse, D.S. (1992a). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. Genes Dev. *6*, 1342–1356.

Izban, M.G., and Luse, D.S. (1992b). Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. J. Biol. Chem. *267*, 13647–13655.

Jansen, A., and Verstrepen, K.J. (2011). Nucleosome Positioning in Saccharomyces cerevisiae. Microbiol. Mol. Biol. Rev. *75*, 301–320.

Jenny, A., Minvielle-Sebastia, L., Preker, P.J., and Keller, W. (1996). Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. Science *274*, 1514–1517.

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. Mol. Cell *52*, 473–484.

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.

Jiang, C., and Pugh, B.F. (2009). A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. Genome Biol. *10*, R109.

Ju, Q.D., Morrow, B.E., and Warner, J.R. (1990). REB1, a yeast DNA-binding protein with many targets, is essential for growth and bears some resemblance to the oncogene myb. Mol. Cell. Biol. *10*, 5226–5234.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096–1099.

Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2010). Nucleosome sequence preferences influence in vivo nucleosome organization. Nat. Struct. Mol. Biol. *17*, 918–920; author reply 920–922.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484–1488.

Kawauchi, J., Mischo, H., Braglia, P., Rondon, A., and Proudfoot, N.J. (2008). Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. Genes Dev. *22*, 1082–1092.

Kelly, S.M., Pabit, S.A., Kitchen, C.M., Guo, P., Marfatia, K.A., Murphy, T.J., Corbett, A.H., and Berland, K.M. (2007). Recognition of polyadenosine RNA by zinc finger proteins. Proc. Natl. Acad. Sci. U. S. A. *104*, 12306–12311.

Kelly, S.M., Leung, S.W., Apponi, L.H., Bramley, A.M., Tran, E.J., Chekanova, J.A., Wente, S.R., and Corbett, A.H. (2010). Recognition of polyadenosine RNA by the zinc finger domain of nuclear poly(A) RNA-binding protein 2 (Nab2) is required for correct mRNA 3'-end formation. J. Biol. Chem. *285*, 26022–26032.

Keogh, M.-C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., et al. (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593–605.

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. J. Biol. Chem. *271*, 27167–27175.

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 Dev. *11*, 2545–2556.

Khaperskyy, D.A., Ammerman, M.L., Majovski, R.C., and Ponticelli, A.S. (2008). Functions of Saccharomyces cerevisiae TFIIF during Transcription Start Site Utilization. Mol. Cell. Biol. *28*, 3757–3766.

Kim, T., and Buratowski, S. (2009). Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. Cell *137*, 259–272.

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. EMBO J. *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.

Koerber, R.T., Rhee, H.S., Jiang, C., and Pugh, B.F. (2009). Interaction of Transcriptional Regulators with Specific Nucleosomes across the Saccharomyces Genome. Mol. Cell *35*, 889–902.

König, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996). The Crystal Structure of the DNA-Binding Domain of Yeast RAP1 in Complex with Telomeric DNA. Cell *85*, 125–136.

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004). Ssu72 Is an RNA Polymerase II CTD Phosphatase. Mol. Cell *14*, 387–394.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., et al. (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol. Cell. Biol. *23*, 4207–4218.

Krogan, N.J., Lam, M.H.Y., Fillingham, J., Keogh, M.-C., Gebbia, M., Li, J., Datta, N., Cagney, G., Buratowski, S., Emili, A., et al. (2004). Proteasome involvement in the repair of DNA double-strand breaks. Mol. Cell *16*, 1027–1034.

Kuehner, J.N., and Brow, D.A. (2008). Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. Mol. Cell *31*, 201–211.

Kuhn, A., Bartsch, I., and Grummt, I. (1990). Specific interaction of the murine transcription termination factor TTF I with class-I RNA polymerases. Nature *344*, 559–562.

Kurtz, S., and Shore, D. (1991). RAP1 protein activates and silences transcription of mating-type genes in yeast. Genes Dev. *5*, 616–628.

Kvint, K., Uhler, J.P., Taschner, M.J., Sigurdsson, S., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2008). Reversal of RNA Polymerase II Ubiquitylation by the Ubiquitin Protease Ubp3. Mol. Cell *30*, 498–506.

Kwak, H., and Lis, J.T. (2013). Control of Transcriptional Elongation. Annu. Rev. Genet. *47*, 483–508.

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 Res. *31*, 3936–3945.

Kyrion, G., Boakye, K.A., and Lustig, A.J. (1992). C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in Saccharomyces cerevisiae. Mol. Cell. Biol. *12*, 5159–5173.

Kyrion, G., Liu, K., Liu, C., and Lustig, A.J. (1993). RAP1 and telomere structure regulate telomere position effects in Saccharomyces cerevisiae. Genes Dev. *7*, 1146–1159.

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

Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. (2006). A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. EMBO J. *25*, 118–128.

Lang, W.H., and Reeder, R.H. (1993). The REB1 site is an essential component of a terminator for RNA polymerase I in Saccharomyces cerevisiae. Mol. Cell. Biol. *13*, 649–658.

Lang, W.H., Morrow, B.E., Ju, Q., Warner, J.R., and Reeder, R.H. (1994). A model for transcription termination by RNA polymerase I. Cell *79*, 527–534.

Lardenois, A., Liu, Y., Walther, T., Chalmel, F., Evrard, B., Granovskaia, M., Chu, A., Davis, R.W., Steinmetz, L.M., and Primig, M. (2011). Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. Proc. Natl. Acad. Sci. U. S. A. *108*, 1058–1063.

Lascaris, R.F., Mager, W.H., and Planta, R.J. (1999). DNA-binding requirements of the yeast protein Rap1p as selected in silico from ribosomal protein gene promoter sequences. Bioinforma. Oxf. Engl. *15*, 267–277.

Lebreton, A., Tomecki, R., Dziembowski, A., and Séraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature *456*, 993–996.

Lee, T.I., and Young, R.A. (2000). Transcription of Eukaryotic Protein-Coding Genes. Annu. Rev. Genet. *34*, 77–137.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73–84.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat. Genet. *39*, 1235–1244.

Li, B., Howe, L., Anderson, S., Yates, J.R., and Workman, J.L. (2003). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. J. Biol. Chem. *278*, 8897–8903.

Li, D., Dong, Y., Jiang, Y., Jiang, H., Cai, J., and Wang, W. (2010). A de novo originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. Cell Res. *20*, 408–420.

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. Mol. Cell *9*, 1101–1111.

Lieb, J.D., Liu, X., Botstein, D., and Brown, P.O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat. Genet. *28*, 327–334.

Lingner, J., Radtke, I., Wahle, E., and Keller, W. (1991). Purification and characterization of poly(A) polymerase from Saccharomyces cerevisiae. J. Biol. Chem. *266*, 8741–8746.

Liu, C.L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S.L., Friedman, N., and Rando, O.J. (2005). Single-nucleosome mapping of histone modifications in S. cerevisiae. PLoS Biol. *3*, e328.

Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell *127*, 1223–1237.

Logan, J., Falck-Pedersen, E., Darnell, J.E., and Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. Proc. Natl. Acad. Sci. U. S. A. *84*, 8306–8310.

Long, M., Betrán, E., Thornton, K., and Wang, W. (2003). The origin of new genes: glimpses from the young and old. Nat. Rev. Genet. *4*, 865–875.

Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. (2008). Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. Mol. Cell *29*, 717–728.

Loya, T.J., O'Rourke, T.W., and Reines, D. (2012). A genetic screen for terminator function in yeast identifies a role for a new functional domain in termination factor Nab3. Nucleic Acids Res. *40*, 7476–7491.

Loya, T.J., O'Rourke, T.W., and Reines, D. (2013). Yeast Nab3 protein contains a self-assembly domain found in human heterogeneous nuclear ribonucleoprotein-C (hnRNP-C) that is necessary for transcription termination. J. Biol. Chem. *288*, 2111–2117.

Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z., and Lingner, J. (2008). The Rat1p 5′ to 3′ Exonuclease Degrades Telomeric Repeat-Containing RNA and Promotes Telomere Elongation in Saccharomyces cerevisiae. Mol. Cell *32*, 465–477.

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 Dev. *20*, 954–965.

Luse, D.S. (2013). Promoter clearance by RNA polymerase II. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1829*, 63–68.

MacIsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC Bioinformatics *7*, 113.

Malik, S., Bagla, S., Chaurasia, P., Duan, Z., and Bhaumik, S.R. (2008). Elongating RNA polymerase II is disassembled through specific degradation of its largest but not other subunits in response to DNA damage in vivo. J. Biol. Chem. *283*, 6897–6905.

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.

Mangus, D.A., Smith, M.M., McSweeney, J.M., and Jacobson, A. (2004). Identification of factors regulating poly(A) tail synthesis and maturation. Mol. Cell. Biol. *24*, 4196–4206.

Marcand, S., Gilson, E., and Shore, D. (1997). A protein-counting mechanism for telomere length regulation in yeast. Science *275*, 986–990.

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.

Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature *429*, 571–574.

Martin, F.H., and Tinoco, I. (1980). DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. Nucleic Acids Res. *8*, 2295–2299.

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. EMBO J. *15*, 2593–2603.

Mas, G., de Nadal, E., Dechant, R., Rodríguez de la Concepción, M.L., Logie, C., Jimeno-González, S., Chávez, S., Ammerer, G., and Posas, F. (2009). Recruitment of a chromatin remodelling complex by the Hog1 MAP kinase to stress genes. EMBO J. *28*, 326–336.

Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. *15*, 108–121.

Matot, B., Le Bihan, Y.-V., Lescasse, R., Perez, J., Miron, S., David, G., Castaing, B., Weber, P., Raynal, B., Zinn-Justin, S., et al. (2012). The orientation of the C-terminal domain of the Saccharomyces cerevisiae Rap1 protein is determined by its binding to DNA. Nucleic Acids Res. *40*, 3197–3207.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. *18*, 1073–1083.

Max, T., Søgaard, M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal Repeat Domain of RNA Polymerase II Facilitates Dissociation of Its Complex with Mediator. J. Biol. Chem. *282*, 14113–14120.

Merkl, P., Perez-Fernandez, J., Pilsl, M., Reiter, A., Williams, L., Gerber, J., Böhm, M., Deutzmann, R., Griesenbeck, J., Milkereit, P., et al. (2014). Binding of the Termination Factor Nsi1 to Its Cognate DNA Site Is Sufficient To Terminate RNA Polymerase I Transcription In Vitro and To Induce Termination In Vivo. Mol. Cell. Biol. *34*, 3817–3827.

Midtgaard, S.F., Assenholt, J., Jonstrup, A.T., Van, L.B., Jensen, T.H., and Brodersen, D.E. (2006). Structure of the nuclear exosome component Rrp6p reveals an interplay between the active site and the HRDC domain. Proc. Natl. Acad. Sci. *103*, 11898–11903.

Minvielle-Sebastia, L., Winsor, B., Bonneaud, N., and Lacroute, F. (1991). Mutations in the yeast RNA14 and RNA15 genes result in an abnormal mRNA decay rate;

sequence analysis reveals an RNA-binding domain in the RNA15 protein. Mol. Cell. Biol. *11*, 3075.

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. EMBO J. *17*, 7454–7468.

Mischo, H.E., and Proudfoot, N.J. (2013). Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. Biochim. Biophys. Acta *1829*, 174–185.

Mizuno, T., Kishimoto, T., Shinzato, T., Haw, R., Chambers, A., Wood, J., Sinclair, D., and Uemura, H. (2004). Role of the N-terminal region of Rap1p in the transcriptional activation of glycolytic genes in Saccharomyces cerevisiae. Yeast Chichester Engl. *21*, 851–866.

Mizzen, C.A., Yang, X.J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., et al. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell *87*, 1261–1270.

Moretti, P., Freeman, K., Coodly, L., and Shore, D. (1994). Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. Genes Dev. *8*, 2257–2269.

Morrow, B.E., Johnson, S.P., and Warner, J.R. (1989). Proteins that bind to the yeast rDNA enhancer. J. Biol. Chem. *264*, 9061–9068.

Murray, S.C., Serra Barros, A., Brown, D.A., Dudek, P., Ayling, J., and Mellor, J. (2012). A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. Nucleic Acids Res. *40*, 2432–2444.

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. J. Biol. Chem. *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. Mol. Cell *29*, 577–587.

Neil, H., Malabat, C., d' Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038–1042.

Németh, A., Perez-Fernandez, J., Merkl, P., Hamperl, S., Gerber, J., Griesenbeck, J., and Tschochner, H. (2013). RNA polymerase I termination: Where is the end? Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1829*, 306–317.

Ng, H.H., Xu, R.-M., Zhang, Y., and Struhl, K. (2002). Ubiquitination of Histone H2B by Rad6 Is Required for Efficient Dot1-mediated Methylation of Histone H3 Lysine 79. J. Biol. Chem. *277*, 34655–34657.

Noble, C.G. (2004). Rna14-Rna15 assembly mediates the RNA-binding capability of Saccharomyces cerevisiae cleavage factor IA. Nucleic Acids Res. *32*, 3364–3375.

Norton, V.G., Imai, B.S., Yau, P., and Bradbury, E.M. (1989). Histone acetylation reduces nucleosome core particle linking number change. Cell *57*, 449–457.

Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R., et al. (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nat. Struct. Mol. Biol. *20*, 923–928.

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. EMBO J. *19*, 37–47.

Orioli, A., Pascali, C., Pagano, A., Teichmann, M., and Dieci, G. (2012). RNA polymerase III transcription control elements: Themes and variations. Gene *493*, 185–194.

Pal, M., Ponticelli, A.S., and Luse, D.S. (2005). The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. Mol. Cell *19*, 101–110.

Pappas, D.L., and Hampsey, M. (2000). Functional Interaction between Ssu72 and the Rpb2 Subunit of RNA Polymerase II in Saccharomyces cerevisiae. Mol. Cell. Biol. *20*, 8343–8351.

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell *125*, 703–717.

Pérez-Cañadillas, J.M. (2006). Grabbing the message: structural basis of mRNA 3′UTR recognition by Hrp1. EMBO J. *25*, 3167–3178.

Phillips, S., and Butler, J.S. (2003). Contribution of domain structure to the RNA 3′ end processing and degradation functions of the nuclear exosome subunit Rrp6p. RNA *9*, 1098–1107.

Piña, B., Fernández-Larrea, J., García-Reyero, N., and Idrissi, F.-Z. (2003). The different (sur)faces of Rap1p. Mol. Genet. Genomics MGG *268*, 791–798.

Pinskaya, M., Gourvennec, S., and Morillon, A. (2009). H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. EMBO J. *28*, 1697–1707.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517–527.

Porrua, O., and Libri, D. (2013a). RNA quality control in the nucleus: the Angels' share of RNA. Biochim. Biophys. Acta *1829*, 604–611.

Porrua, O., and Libri, D. (2013b). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nat. Struct. Mol. Biol. *20*, 884–891.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nat. Rev. Mol. Cell Biol. *16*, 190–202.

Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R.K., Stefl, R., and Libri, D. (2012). In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. EMBO J. *31*, 3935–3948.

Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription regulation. Crit. Rev. Biochem. Mol. Biol. *48*, 575–608.

Preker, P.J., and Keller, W. (1998). The HAT helix, a repetitive motif implicated in RNA processing. Trends Biochem. Sci. *23*, 15–16.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851–1854.

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.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell *123*, 233–248.

Ratner, J.N., Balasubramanian, B., Corden, J., Warren, S.L., and Bregman, D.B. (1998). Ultraviolet radiation-induced ubiquitination and proteasomal degradation of the large subunit of RNA polymerase II. Implications for transcription-coupled DNA repair. J. Biol. Chem. *273*, 5184–5189.

Reid, J.L., Iyer, V.R., Brown, P.O., and Struhl, K. (2000). Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. Mol. Cell *6*, 1297–1307.

Reinberg, D., and Sims, R.J. (2006). de FACTo nucleosome dynamics. J. Biol. Chem. *281*, 23297–23301.

Reines, D. (1992). Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. J. Biol. Chem. *267*, 3795–3800.

Reiter, A., Hamperl, S., Seitz, H., Merkl, P., Perez-Fernandez, J., Williams, L., Gerber, J., Németh, A., Léger, I., Gadal, O., et al. (2012). The Reb1-homologue Ydr026c/Nsi1 is required for efficient RNA polymerase I termination in yeast. EMBO J. *31*, 3480–3493.

Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H., and Price, D.H. (2001). A Highly Purified RNA Polymerase II Elongation Control System. J. Biol. Chem. *276*, 42601–42609.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408–1419.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.

Robzyk, K., Recht, J., and Osley, M.A. (2000). Rad6-Dependent Ubiquitination of Histone H2B in Yeast. Science *287*, 501–504.

Rusché, L.N., Kirchmaier, A.L., and Rine, J. (2002). Ordered Nucleation and Spreading of Silenced Chromatin inSaccharomyces cerevisiae. Mol. Biol. Cell *13*, 2207–2222.

Russo, P., Li, W.Z., Guo, Z., and Sherman, F. (1993). Signals that produce 3' termini in CYC1 mRNA of the yeast Saccharomyces cerevisiae. Mol. Cell. Biol. *13*, 7836–7849.

Sachs, A.B., and Deardorff, J.A. (1992). Translation initiation requires the PAB-dependent poly(A) ribonuclease in yeast. Cell *70*, 961–973.

Sachs, A.B., Bond, M.W., and Kornberg, R.D. (1986). A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression. Cell *45*, 827–835.

Sachs, A.B., Davis, R.W., and Kornberg, R.D. (1987). A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. Mol. Cell. Biol. *7*, 3268–3276.

Sadowski, M., Dichtl, B., Hübner, W., and Keller, W. (2003). Independent functions of yeast Pcf11p in pre-mRNA 3′ end processing and in transcription termination. EMBO J. *22*, 2167–2177.

Sadowski, M., Suryadinata, R., Tan, A.R., Roesley, S.N.A., and Sarcevic, B. (2012). Protein monoubiquitination and polyubiquitination generate structural diversity to control distinct biological processes. IUBMB Life *64*, 136–142.

Saeki, H., and Svejstrup, J.Q. (2009). Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. Mol. Cell *35*, 191–205.

Schaeffer, D., Reis, F.P., Johnson, S.J., Arraiano, C.M., and van Hoof, A. (2012). The CR3 motif of Rrp44p is important for interaction with the core exosome and exosome function. Nucleic Acids Res. *40*, 9298–9307.

Schaft, D., Roguev, A., Kotovic, K.M., Shevchenko, A., Sarov, M., Shevchenko, A., Neugebauer, K.M., and Stewart, A.F. (2003). The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. Nucleic Acids Res. *31*, 2475–2482.

Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-wide mapping of yeast RNA polymerase II termination. PLoS Genet. *10*, e1004632.

Schawalder, S.B., Kabani, M., Howald, I., Choudhury, U., Werner, M., and Shore, D. (2004). Growth-regulated recruitment of the essential yeast ribosomal protein gene activator Ifh1. Nature *432*, 1058–1061.

Schmidt, K., Xu, Z., Mathews, D.H., and Butler, J.S. (2012). Air proteins control differential TRAMP substrate specificity for nuclear RNA surveillance. RNA *18*, 1934–1945.

Schuch, B., Feigenbutz, M., Makino, D.L., Falk, S., Basquin, C., Mitchell, P., and Conti, E. (2014). The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. EMBO J. *33*, 2829–2846.

Schweikhard, V., Meng, C., Murakami, K., Kaplan, C.D., Kornberg, R.D., and Block, S.M. (2014). Transcription factors TFIIF and TFIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms. Proc. Natl. Acad. Sci. *111*, 6642–6647.

Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr. Opin. Struct. Biol. *19*, 65–71.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.

Selth, L.A., Sigurdsson, S., and Svejstrup, J.Q. (2010). Transcript Elongation by RNA Polymerase II. Annu. Rev. Biochem. *79*, 271–293.

Shi, Y.G., and Tsukada, Y. (2013). The Discovery of Histone Demethylases. Cold Spring Harb. Perspect. Biol. *5*, a017947.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

Shore, D., and Nasmyth, K. (1987). Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell *51*, 721–732.

Shukla, A., and Bhaumik, S.R. (2007). H2B-K123 ubiquitination stimulates RNAPII elongation independent of H3-K4 methylation. Biochem. Biophys. Res. Commun. *359*, 214–220.

Siddiqui, N., Mangus, D.A., Chang, T.-C., Palermino, J.-M., Shyu, A.-B., and Gehring, K. (2007). Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein. J. Biol. Chem. *282*, 25067–25075.

Sigurdsson, S., Dirac-Svejstrup, A.B., and Svejstrup, J.Q. (2010). Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. Mol. Cell *38*, 202–210.

Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. Curr. Opin. Cell Biol. *21*, 344–351.

Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A., and Arndt, K.M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. EMBO J. *22*, 1846–1856.

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. Mol. Cell *10*, 1429–1439.

Smith, D.J., and Whitehouse, I. (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature *483*, 434–438.

Smolle, M., Venkatesh, S., Gogol, M.M., Li, H., Zhang, Y., Florens, L., Washburn, M.P., and Workman, J.L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. Nat. Struct. Mol. Biol. *19*, 884–892.

Somesh, B.P., Reid, J., Liu, W.-F., Søgaard, T.M.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2005). Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. Cell *121*, 913–923.

Somesh, B.P., Sigurdsson, S., Saeki, H., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2007). Communication between distant sites in RNA polymerase II through ubiquitylation factors and the polymerase CTD. Cell *129*, 57–68.

Stead, J.A., Costello, J.L., Livingstone, M.J., and Mitchell, P. (2007). The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein. Nucleic Acids Res. *35*, 5556–5567.

Steinmetz, E.J., and Brow, D.A. (1996). Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. Mol. Cell. Biol. *16*, 6993–7003.

Steinmetz, E.J., and Brow, D.A. (1998). Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. Proc. Natl. Acad. Sci. U. S. A. *95*, 6699–6704.

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. Mol. Cell. Biol. *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. Mol. Cell *24*, 735–746.

Stirnimann, C.U., Petsalaki, E., Russell, R.B., and Müller, C.W. (2010). WD40 proteins propel cellular networks. Trends Biochem. Sci. *35*, 565–574.

Sun, Z.-W., and Allis, C.D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature *418*, 104–108.

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. Mol. Cell. Biol. *16*, 1557–1566.

Suntharalingam, M., Alcázar-Román, 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. J. Biol. Chem. *279*, 35384–35391.

Sussel, L., and Shore, D. (1991). Separation of transcriptional activation and silencing functions of the RAP1-encoded repressor/activator protein 1: isolation of viable mutants affecting both silencing and telomere length. Proc. Natl. Acad. Sci. U. S. A. *88*, 7749–7753.

Svejstrup, J.Q. (2007). Contending with transcriptional arrest during RNAPII transcript elongation. Trends Biochem. Sci. *32*, 165–171.

Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R.R., Grimmond, S.M., Schroder, K., et al. (2009). Tiny RNAs associated with transcription start sites in animals. Nat. Genet. *41*, 572–578.

Tanny, J.C., Erdjument-Bromage, H., Tempst, P., and Allis, C.D. (2007). Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. Genes Dev. *21*, 835–847.

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

Tarun, S.Z., Wells, S.E., Deardorff, J.A., and Sachs, A.B. (1997). Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. Proc. Natl. Acad. Sci. U. S. A. *94*, 9046–9051.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol. Cell *23*, 853–864.

Thiebaut, M., Colin, J., Neil, H., Jacquier, A., Séraphin, B., Lacroute, F., and Libri, D. (2008). Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in S. cerevisiae. Mol. Cell *31*, 671–682.

Thompson, D.M., and Parker, R. (2007). Cytoplasmic Decay of Intergenic Transcripts in Saccharomyces cerevisiae. Mol. Cell. Biol. *27*, 92–101.

Tillo, D., and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics *10*, 442.

Tisseur, M., Kwapisz, M., and Morillon, A. (2011). Pervasive transcription – Lessons from yeast. Biochimie *93*, 1889–1896.

Tran, E.J., Zhou, Y., Corbett, A.H., and Wente, S.R. (2007). The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. Mol. Cell *28*, 850–859.

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. Mol. Cell *55*, 467–481.

Uhler, J.P., Hertel, C., and Svejstrup, J.Q. (2007). A role for noncoding transcription in activation of the yeast PHO5 gene. Proc. Natl. Acad. Sci. *104*, 8011–8016.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26–46.

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 Res. *32*, 2441–2452.

Valentini, S.R., Weiss, V.H., and Silver, P.A. (1999). Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation. RNA N. Y. N *5*, 272–280.

Valerius, O., Brendel, C., Düvel, K., and Braus, G.H. (2002). Multiple factors prevent transcriptional interference at the yeast ARO4-HIS7 locus. J. Biol. Chem. *277*, 21440–21445.

Vanácová, 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 Biol. *3*, e189.

Vasiljeva, L., and Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. Mol. Cell *21*, 239–248.

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. Nat. Struct. Mol. Biol. *15*, 795–804.

Verma, R., Oania, R., Fang, R., Smith, G.T., and Deshaies, R.J. (2011). Cdc48/p97 mediates UV-dependent turnover of RNA Pol II. Mol. Cell *41*, 82–92.

Vignais, M.L., Woudt, L.P., Wassenaar, G.M., Mager, W.H., Sentenac, A., and Planta, R.J. (1987). Specific binding of TUF factor to upstream activation sites of yeast ribosomal protein genes. EMBO J. *6*, 1451–1457.

Viphakone, N., Voisinet-Hakil, F., and Minvielle-Sebastia, L. (2008). Molecular dissection of mRNA poly(A) tail length control in yeast. Nucleic Acids Res. *36*, 2418–2433.

Wasmuth, E.V., Januszyk, K., and Lima, C.D. (2014). Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. Nature *511*, 435–439.

Whitehouse, I., and Tsukiyama, T. (2006). Antagonistic forces that position nucleosomes in vivo. Nat. Struct. Mol. Biol. *13*, 633–640.

Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature *450*, 1031–1035.

Williams, T.L., Levy, D.L., Maki-Yonekura, S., Yonekura, K., and Blackburn, E.H. (2010). Characterization of the yeast telomere nucleoprotein core: Rap1 binds independently to each recognition site. J. Biol. Chem. *285*, 35814–35824.

Wilson, M.D., Harreman, M., and Svejstrup, J.Q. (2013a). Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. Biochim. Biophys. Acta *1829*, 151–157.

Wilson, M.D., Harreman, M., Taschner, M., Reid, J., Walker, J., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2013b). Proteasome-Mediated Processing of Def1, a Critical Step in the Cellular Response to Transcription Stress. Cell *154*, 983–995.

Wilson, S.M., Datar, K.V., Paddy, M.R., Swedlow, J.R., and Swanson, M.S. (1994). Characterization of nuclear polyadenylated RNA-binding proteins in Saccharomyces cerevisiae. J. Cell Biol. *127*, 1173–1184.

Wlotzka, W., Kudla, G., Granneman, S., and Tollervey, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J. *30*, 1790–1803.

Wolf, J., Valkov, E., Allen, M.D., Meineke, B., Gordiyenko, Y., McLaughlin, S.H., Olsen, T.M., Robinson, C.V., Bycroft, M., Stewart, M., et al. (2014). Structural basis for Pan3 binding to Pan2 and its function in mRNA recruitment and deadenylation. EMBO J. *33*, 1514–1526.

Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J.F., et al. (2003). Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol. Cell *11*, 267–274.

Wotton, D., and Shore, D. (1997). A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in Saccharomyces cerevisiae. Genes Dev. *11*, 748–760.

Wright, D.E., Wang, C.-Y., and Kao, C.-F. (2012). Histone ubiquitylation and chromatin dynamics. Front. Biosci. Landmark Ed. *17*, 1051–1078.

Wu, J., Delneri, D., and O'Keefe, R.T. (2012). Non-coding RNAs in *Saccharomyces cerevisiae* : what is the function? Biochem. Soc. Trans. *40*, 907–911.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005a). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725–737.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev. *17*, 654–663.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.

Yadon, A.N., Mark, D.V. de, Basom, R., Delrow, J., Whitehouse, I., and Tsukiyama, T. (2010). Chromatin Remodeling around Nucleosome-Free Regions Leads to Repression of Noncoding RNA Transcription. Mol. Cell. Biol. *30*, 5110–5122.

Yang, C., and Ponticelli, A.S. (2012). Evidence that RNA polymerase II and not TFIIB is responsible for the difference in transcription initiation patterns between Saccharomyces cerevisiae and Schizosaccharomyces pombe. Nucleic Acids Res. *40*, 6495–6507.

Yarragudi, A., Miyake, T., Li, R., and Morse, R.H. (2004). Comparison of ABF1 and RAP1 in Chromatin Opening and Transactivator Potentiation in the Budding Yeast Saccharomyces cerevisiae. Mol. Cell. Biol. *24*, 9152–9164.

Yarrington, R.M., Richardson, S.M., Lisa Huang, C.R., and Boeke, J.D. (2012). Novel transcript truncating function of Rap1p revealed by synthetic codon-optimized Ty1 retrotransposon. Genetics *190*, 523–535.

Yonaha, M., and Proudfoot, N.J. (1999). Specific Transcriptional Pausing Activates Polyadenylation in a Coupled In Vitro System. Mol. Cell *3*, 593–600.

Yu, L., and Morse, R.H. (1999). Chromatin Opening and Transactivator Potentiation by RAP1 in Saccharomyces cerevisiae. Mol. Cell. Biol. *19*, 5279–5288.

Yu, Q., Qiu, R., Foland, T.B., Griesen, D., Galloway, C.S., Chiu, Y.-H., Sandmeier, J., Broach, J.R., and Bi, X. (2003). Rap1p and other transcriptional regulators can function in defining distinct domains of gene expression. Nucleic Acids Res. *31*, 1224–1233.

Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626–630.

Zawel, L., Kumar, K.P., and Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. Genes Dev. *9*, 1479–1490.

Zeng, X., Deminoff, S.J., and Santangelo, G.M. (1997). Specialized Rap1p/Gcr1p transcriptional activation through Gcr1p DNA contacts requires Gcr2p, as does hyperphosphorylation of Gcr1p. Genetics *147*, 493–505.

Zhang, Z., and Dietrich, F.S. (2005). Mapping of transcription start sites in Saccharomyces cerevisiae using 5′ SAGE. Nucleic Acids Res. *33*, 2838–2851.

Zhang, C., Zobeck, K.L., and Burton, Z.F. (2005a). Human RNA polymerase II elongation in slow motion: role of the TFIIF RAP74 alpha1 helix in nucleoside triphosphate-driven translocation. Mol. Cell. Biol. *25*, 3583–3595.

Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodríguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. (2012). Ssu72 Phosphatase-dependent Erasure of Phospho-Ser7 Marks on the RNA Polymerase II C-terminal Domain Is Essential for Viability and Transcription Termination. J. Biol. Chem. *287*, 8541–8551.

Zhang, Z., Fu, J., and Gilmour, D.S. (2005b). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. Genes Dev. *19*, 1572–1580.

Zhao, J., Kessler, M.M., and Moore, C.L. (1997). Cleavage Factor II of Saccharomyces cerevisiaeContains Homologues to Subunits of the Mammalian Cleavage/ Polyadenylation Specificity Factor and Exhibits Sequence-specific, ATP-dependent Interaction with Precursor RNA. J. Biol. Chem. *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. Mol. Cell. Biol. *18*, 5942.

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 N. Y. N *12*, 435–445.

Zuo, Y., and Deutscher, M.P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. Nucleic Acids Res. *29*, 1017–1026.