



HAL
open science

Etude génomique des mécanismes nucléaires de contrôle qualité et dégradation de l'ARN

Kevin Moreau

► **To cite this version:**

Kevin Moreau. Etude génomique des mécanismes nucléaires de contrôle qualité et dégradation de l'ARN. Sciences agricoles. Université d'Orléans, 2019. Français. NNT: 2019ORLE3012. tel-03101419

HAL Id: tel-03101419

<https://theses.hal.science/tel-03101419>

Submitted on 7 Jan 2021

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 D'ORLÉANS

ÉCOLE DOCTORALE SANTE, SCIENCES BIOLOGIQUES ET CHIMIE DU VIVANT

Centre de Biophysique Moléculaire, UPR4301 du CNRS

THÈSE présentée par :
Kévin MOREAU

soutenue le : **21 Mars 2019**

pour obtenir le grade de : **Docteur de l'université d'Orléans**

Discipline/ Spécialité : **Biologie Moléculaire et cellulaire, analyse bio-informatique**

**Etude génomique des mécanismes
nucléaires de contrôle qualité et dégradation
de l'ARN**

THÈSE dirigée par :

Dr. A. Rachid RAHMOUNI

Directeur de recherche, CNRS Orléans

Dr Yves BIGOT

Directeur de recherche, CNRS-INRA Nouzilly

RAPPORTEURS :

Dr. Benoit PALANCADE

Directeur de recherche, CNRS Paris

Dr. Mathieu ROUGEMAILLE

Chargé de recherche, CNRS Paris

JURY :

Pr. Stéphane MAURY

Professeur d'Université, Orléans

Dr. Benoit PALANCADE

Directeur de recherche, CNRS Paris

Dr. Mathieu ROUGEMAILLE

Chargé de recherche, CNRS Paris

Dr Gilles TRUAN

Directeur de recherche, INSA Toulouse

Dr Yves BIGOT

Directeur de recherche, CNRS-INRA Nouzilly

Dr. A. Rachid RAHMOUNI

Directeur de recherche, CNRS Orléans

Remerciements

Mes premiers remerciements nous ramènent un peu plus de 3 ans et demi dans le passé, au moment où je me suis assis dans le bureau du Dr Rachid Rahmouni. C'est pendant ce mois de mai 2015 que Rachid m'a présenté le sujet, qui, quelques mois plus tard deviendrait MON sujet de thèse. C'est donc tout naturellement que j'adresse mes plus sincères remerciements à mon mentor. Vous m'avez fait grandir scientifiquement et m'avez montré sans retenu ce qu'était la recherche publique. Nos longues discussions, qu'elles soient à propos de biologie ou non, resteront un souvenir impérissable.

N'oublions pas que cette thèse est une cotutelle Orléans-Tours, je remercie donc également mon co-directeur de thèse, le Dr Yves Bigot, qui ne manque pas une occasion de nous rappeler qu'un scientifique peut également faire preuve d'une bonne dose d'humour.

Je tiens également à remercier Benoît Piégu, qui a assuré ma formation initiale en bio-informatique, et qui m'a ouvert à un nouveau monde plein de possibilités. Merci également de m'avoir enseigné la « théorie du mur ».

Revenons dans les contrées Orléanaises, pour remercier le Dr Christine Mosrin-Huaman de m'avoir accompagné sur mes premières manipes et pour m'avoir montré ce panel de techniques de généticiens. Un grand merci également pour nos discussions sur l'avenir de la science et de l'humanité.

Des remerciements sont également de mise pour toute l'équipe « Mort Cellulaire Programmée » ! Sans mes stages de Master, je n'aurais pas pu vivre ces trois années. Donc un grand merci au Pr Alain Legrand. Mais aussi merci à Fabienne, Thierry, Stéphane, Martine. Un remerciement tout particulier pour Lucile qui est devenue bien plus qu'une enseignante ou même une collègue à mes yeux, mais bien une amie. Nos pauses cigarettes à refaire le monde me manqueront !

La personne qui suit dans ces remerciements n'est autre que ma collègue et amie Aurélia Le Dantec. Merci de m'avoir supporté durant ces 3 ans et demi. Merci pour ton implication dans ma thèse, je te dois une fière chandelle pour certaines des manipes qui sont exposées ici. Mais surtout merci pour ta joie, ta bonne humeur, tes bruitages (il est important

de le relever, tu emploies trop peu souvent le bruitage adapté à la situation que tu veux décrire...). Je ne pensais pas tisser des liens d'amitiés comme ceux que nous partageons au cours de ma thèse, et pourtant me voilà devenu témoin de ton mariage ! J'en profite pour également remercier Sieur Richard Depuydt, notre passion pour Dragon Ball ne se tarira jamais !

Merci à mon ami Loïck Ridou, les apéros post-labo étaient vraiment trop cools. J'espère que tu vas trouver ta voie pour la suite...

Plus qu'un merci est de mise, pour mon compère de toujours, Geoffrey Casas. Nous nous suivons désormais depuis 9 ans, et nous nous marrons toujours autant. J'espère que nous continuerons à nous soutenir longtemps après nos thèses.

Merci également à toutes les personnes que j'ai pu côtoyer au CBM. Christine, Nadia, Marie-Christine, et l'équipe administrative. Un énorme merci à Yannick Berteaux, tu m'as plusieurs fois sauvé la vie (de manière informatique souvent) ! Je n'oublie pas les thésards qui sont déjà partis (Amandine, Simon, Alexandra...) et ceux qui me suivent de près.

Je remercie également mes proches amis Bryan, Laurent, Thomas et Romain qui m'accompagnent déjà depuis longtemps et dont le soutien a été indéfectible au cours de ces trois dernières années.

Je clôture mes remerciements avec ceux envers ma famille. Un merci ne suffit pas à décrire la gratitude que j'éprouve envers mes parents. Vous avez toujours été là pour moi, vous avez toujours été un modèle de droiture et d'intégrité, de gentillesse et un véritable refuge dans les moments plus durs. Si j'en suis là aujourd'hui, c'est également parce que vous avez toujours cru en moi et m'avez toujours laissé faire mes choix. Mais plus important encore, vous m'avez toujours appuyé dans ces choix. Vous êtes les fondations de ce que je suis. Merci. Un merci et aussi des félicitations à ma petite sœur, qui à force de persévérance a trouvé sa voie, et sera, j'en suis sûr, excellente dans son futur travail (et aussi félicitation de supporter un grand frère comme moi !)

Je termine avec la femme qui partage ma vie, nous avons tous deux traversé ces trois ans, en poursuivant des thèses à 350 Km de distance. Et pourtant malgré cette distance, nous n'avons cessé d'être de plus en plus proches et aimant l'un pour l'autre. Merci Mélissa d'être la compagne la plus parfaite pour moi.

Je remercie sincèrement les membres du jury d'avoir accepté de prendre le temps d'expertiser mes travaux de thèse. C'est un honneur de soumettre ces derniers à vos yeux avisés. Un grand remerciement va aux docteurs Benoit Palancade et Mathieu Rougemaille pour leurs retours et corrections justes et précises qui m'ont permis d'améliorer ce manuscrit.

Table des matières

Table des matières	5
Abréviations	7
1. Introduction.....	9
1.1 De la biologie moléculaire à la génomique.	9
1.2 La biogenèse des acides ribonucléiques en particules ribonucléoprotéiques.....	12
1.2.1 RNA polymérase II et cofacteurs.....	14
1.2.1.1. RNA polymérase II : domaine C-terminal (CTD).....	15
1.2.1.2 RNA polymérase II : acteur principal de la transcription des mRNAs et la biogenèse des mRNPs	19
1.2.1.3 RNA polymérase II : transcription des longs ARNs non codant et des snRNAs	19
1.2.2 Empaquetage et Maturation des ARNm.....	21
1.2.2.1. Coiffe et épissage, initiation et élongation	22
1.2.2.2. Clivage et polyadénylation, terminaison.....	23
1.2.3 Empaquetage et export des ARNm : exemple du complexe THO	26
1.2.3.1 Composition et recrutement du complexe THO	27
1.2.3.2 Fonctions du complexe THO	28
1.2.3.3. Stabilité de la mRNP et compétence à l'export	30
1.2.4. Les ARNs non codants : Terminaison et Maturation	32
1.2.4.1.1. Le complexe Nrd1/Nab3/Sen1	35
1.2.4.2. Les longs ARNs non codants chez la levure.....	40
1.2.4.2.1. Rôles et hypothèses de fonctions des lncRNAs.....	40
1.2.4.2.2. Régulation particulière de ces lncRNAs.....	41
1.3. Contrôle qualité et dégradation	44
1.3.1. Machinerie canonique de dégradation : l'exosome	44

1.3.1.2. Structure et fonctionnement de l'exosome.....	46
1.3.1.2. Cofacteurs de l'exosome	47
1.3.1.2.1. Le complexe TRAMP	48
1.3.1.2.2. Autres cofacteurs de l'exosome	49
1.4. Disséquer les systèmes de contrôle qualité et dégradation nucléaires de l'ARN.....	51
1.4.1. L'hélicase translocase bactérienne Rho.....	51
1.4.1.1. Structure et fonction	51
1.4.1.2. Détournement du système bactérien pour amplifier la fenêtre d'observation d'évènements rares.....	54
1.5 Problématique de travail et projets de thèse.....	56
2. Résultat Partie 1 : Article n°1	59
2.1. Article n°1.	59
2.2. Discussion de l'article et perspectives liées	69
3. Résultat Partie 2 : Article n°2	71
3.1. Introduction à l'article	71
3.2. Manuscrit de l'article n°2	72
4. Résultats Partie 3 : Dissection des mécanismes menant à la détection des mRNPs aberrantes par le QC.....	97
4.1. Introduction et mise en contexte	97
4.2. Matériel et méthodes	101
4.3. Résultats	102
5. Travaux en cours	115
6. Discussion et Perspectives	117
Bibliographie	123

Abréviations

7mG	N7-methyl Guanosine	HTS	High Througput Screening
ADN	Acide DésoxyriboNucléique	lncRNAs	Long non coding RiboNucleic Acids
ADP	Adénosine DiPhosphate	mRNP	messenger RiboNucleoprotein Particule
ARN	Acide RiboNucléique	ncRNAs	Non coding RNAs
ARNm	Acide RiboNucléique messenger	NGS	Next Generation Sequencing
ARNr	Acide RiboNucléique ribosomal	NNS	Nrd1/Nab3/Sen1
ARNt	Acide RiboNucléique de transfert	PAR-CLIP	photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation
ATP	Adénosine TriPhosphate	PIC	Pre-Initiation Complex
BM	Biologie Moléculaire	poll ou II ou III	RNA polymerase I ou II ou III
BCM	BiCycloMycine	QC	Contrôle Qualité
CBC	Cap-Binding Complex	RNA	RiboNucleic Acid
CF	Cleavage Factor	snRNAs	small nuclear RNAs
ChIP	Chromatin ImmunoPrecipitation	snoRNAs	small nucleolar RNAs
CID	CTD-Interacting Domain	sn/snoRNPs	Small nucle(ol)ar RiboNucleoProtein
CPF	Cleavage/Polyadenylation Factor	SUTs	Stable Unannotated Transcripts
CTD	C-Terminal Domain	TRAMP	Trf4p-Air1/2p-Mtr4p Polyadenylation complex
CUTs	Cryptique Unstable Transcripts	TREX	TRanscription EXport complex

1. Introduction

1.1 De la biologie moléculaire à la génomique.

La Biologie, ou « la science du vivant », est un champ d'étude scientifique couvrant une multitude de domaines. Ce que nous considérons comme vivant étant extrêmement large, étendu et varié mais également difficile à définir clairement, conduit la biologie à couvrir des études allant de l'échelle moléculaire à celle de l'écosystème en passant par les échelles cellulaires, de l'organisme et de la population. La Biologie étant vaste, elle se doit d'être connectée avec d'autres Sciences et est intimement liée aux progrès technologiques. Ce dernier constat explique la rapide évolution des Sciences Biologiques à toutes les échelles.

La biologie moléculaire, la « sous-partie » de la biologie qui nous intéresse ici, ne déroge pas à la règle. Son évolution depuis les années 1930 a été fulgurante et pavée de découvertes majeures qui ont, à jamais, changée la façon dont l'humanité perçoit sa propre existence et celle des espèces qui l'entourent. La biologie moléculaire, qui est liée à la génétique et à la biochimie, vise à comprendre les mécanismes fondamentaux du fonctionnement de la cellule à l'échelle moléculaire. Le terme de « biologie moléculaire » sera utilisé pour la première fois par Warren Weaver en 1938, directeur des Sciences Naturelles pour la fondation Rockefeller à cette époque.

Pour situer rapidement l'importance de la biologie moléculaire, voici quelques découvertes qui ont été faites à travers cette discipline : l'ADN comme support de l'hérédité par Oswald Avery en 1944 (Avery et al., 1944) ; mise à jour de la structure en double hélice de l'ADN par J.D Watson, F. Crick, R. Franklin & M. Wilkins en 1953 (Watson and Crick, 1953); formulation du dogme central de la biologie moléculaire par F. Crick en 1958; démonstration que la réplication de l'ADN est semi conservative par M.S. Meselson et F. W. Stahl, également en 1958 (Meselson and Stahl, 1958); Découverte de l'ARNm, molécule intermédiaire entre l'ADN et la protéine (Jacob and Monod, 1961); Code génétique résolu en 1965; séquençage de courte séquence d'ARN par F. Sanger en 1968 (Brownlee et al., 1968). Dix ans plus tard le premier génome entier est séquencé (celui du virus bactérien phiX174, en 1977). En 1983, K. Mullis invente la PCR. En 1990, 3 chercheurs découvrent les ARNs non codants interférants. Ceci n'est qu'un échantillon de la richesse des découvertes en biologie moléculaire (BM) au XX^e siècle. Ce siècle nous a donc donné les clés pour comprendre comment étudier

l'expression des gènes, les fonctions des ARNs, celles des protéines, et la régulation des procédés qui permettent de passer de l'information génétique (portée par l'ADN) aux fonctions cellulaires.

Cette discipline nous a permis de mieux appréhender comment la vie fonctionnait, quelles étaient ses règles, et elle continue de nous apporter des réponses. L'évolution des techniques de séquençages, en lien direct avec les découvertes de la BM, ont transportés le monde de la biologie dans une toute nouvelle dimension. Là où par le passé nous étudions la régulation d'un gène, d'un transcrit, la structure d'un ARN ou juste sa séquence, nous pouvons désormais, en un temps remarquablement court, déterminer l'ensemble des séquences qui composent un génome complexe, ou encore élucider l'ensemble des structures ARNs d'une cellule.

Ces derniers progrès dans les méthodes et technologies de séquençages (NGS), qui figurent en bonne place dans la boîte à outil des biologistes, se sont faits en parallèle des pas de géants réalisés dans le domaine de l'informatique. Les liens entre biologie et informatique sont de plus en plus étroits, car les masses d'informations génomiques, transcriptomiques, épigénomiques, etc générées par les séquenceurs nécessitent de puissantes machines pour les traiter et en extraire les réponses aux questions biologiques que se pose la communauté. L'informatique dans la biologie, ou bio-informatique, avait déjà une place de choix pour la détermination de structures de protéines, ou l'alignement local ou global de séquence (blast, MUSCLE, etc) mais l'avènement des NGS a fait de la bio-informatique couplée aux données de séquençage un outil de choix pour l'étude de la biologie moléculaire à une échelle globale.

Les résultats exposés dans ce manuscrit de thèse relatent des faits de biologie moléculaire fondamentale qui régissent des phénomènes qui se déroulent de manière co-transcriptionnelle et qui ont été observés par des méthodes de séquençage à haut débit et analysés avec des outils bio-informatiques. Toutes les expérimentations et informations décrites ici sont issues de la levure *Saccharomyces cerevisiae* (*S. cerevisiae*), sauf si précisé. L'utilisation de la levure comme modèle d'étude possède plusieurs avantages. Tout d'abord, cet eucaryote peut être utilisé et étudié à l'état haploïde ou diploïde (formule chromosomique 1n ou 2n) et possède peu de gènes codants (~5000-6000), et donc l'étude d'un gène peut se faire en observation directe lorsque la levure est haploïde (pas de biais due à la présence d'un second allèle). De plus, les techniques de génétique sont facilement applicables à la levure et

sa culture en laboratoire est aisée. Il est également possible d'exploiter la faculté de la levure à réaliser la recombinaison homologue dès qu'on lui présente un fragment d'ADN qui possède des séquences homologues à des régions de son génome. Ceci va permettre, par exemple, de marquer une protéine de manière endogène en apportant la séquence nucléotidique d'un marqueur flanquée de séquences homologues à la zone d'insertion. Un dernier avantage est que le génome de la levure, très étudié, est séquencé et annoté, permettant des études à l'échelle de ce dernier sans avoir besoin d'assembler des séquences *de novo*.

1.2 La biogenèse des acides ribonucléiques en particules ribonucléoprotéiques.

La transcription est le procédé biologique qui permet de copier une séquence d'ADN en une séquence d'ARN. Ce procédé est conservé chez les procaryotes et les eucaryotes. Il existe, dans le noyau des cellules eucaryotes, trois différents complexes protéiques qui gèrent la transcription, les ARN polymérases. L'ARN polymérase I (polI) gère la transcription des ARNs ribosomiques (ARNr), hormis l'ARN 5S. L'ARN polymérase III (polIII) s'occupe de la transcription des ARNs de transferts (ARNt), de l'ARNr 5S et d'autres petits ARNs. **L'ARN polymérase II (polII), l'un des acteurs au cœur de mes travaux de thèse, transcrit plusieurs classes d'ARNs différents.** Tout d'abord il y a les ARNs messagers (ARNm), porteurs de l'information nécessaire à la production de protéines, par le biais de leur traduction une fois exportés dans le cytoplasme. La polII transcrit aussi les petits ARNs nuclé(ol)aires (small nucle(ol)ar RNA ou sn/snoRNAs) qui portent plusieurs fonctions dont un rôle majeur dans l'épissage des ARNm (snRNA) et également un rôle dans la maturation des ARNs ribosomiaux (snoRNAs), d'autres petits ARNs (comme les micro-ARNs). Enfin, plus récemment découverts, polII transcrit également les ARNs non-codant appelés, longs ARNs non codants (lncRNAs).

PolII se compose de 12 sous-unités chez la levure et chez l'homme. Elle a été cristallisée pour la première fois en 2001 (Cramer et al., 2001). La figure 1 présente une vision récente de la structure de ce complexe enzymatique (Hahn and Young, 2011).

La transcription peut être séquentiellement divisée en trois phases : l'initiation, l'élongation et la terminaison. À elle seule, polII n'est pas capable de se lier aux séquences promotrices des unités transcriptionnelles. La présence de nombreux cofacteurs sur le promoteur est requise pour que polII soit recrutée à son tour et commencer la transcription de l'ARN. L'élongation est l'étape centrale de la transcription. Elle correspond à l'ajout successif de ribonucléotides au brin d'ARN. Pour que cette étape puisse se faire, polII va subir des modifications post-traductionnelles (phosphorylations en majorité) qui sont assurées par d'autres cofacteurs. Enfin, la terminaison de la transcription s'effectue lorsque polII transcrit un signal de polyadénylation qui va être reconnu par différentes protéines pour le clivage et la polyadénylation du brin d'ARN.

Tout au long des différentes étapes de la transcription, l'ARN naissant va être empaqueté, mûri, modifié par différents acteurs protéiques et/ou ribonucléiques. Ce procédé aboutit à ce que nous appelons une particule ribonucléoprotéique (RNP). Ces étapes se déroulent de manière co-transcriptionnelle et constituent ce qui est appelé la « biogenèse » de la mRNP (« m » dans le cas d'un ARN messager).

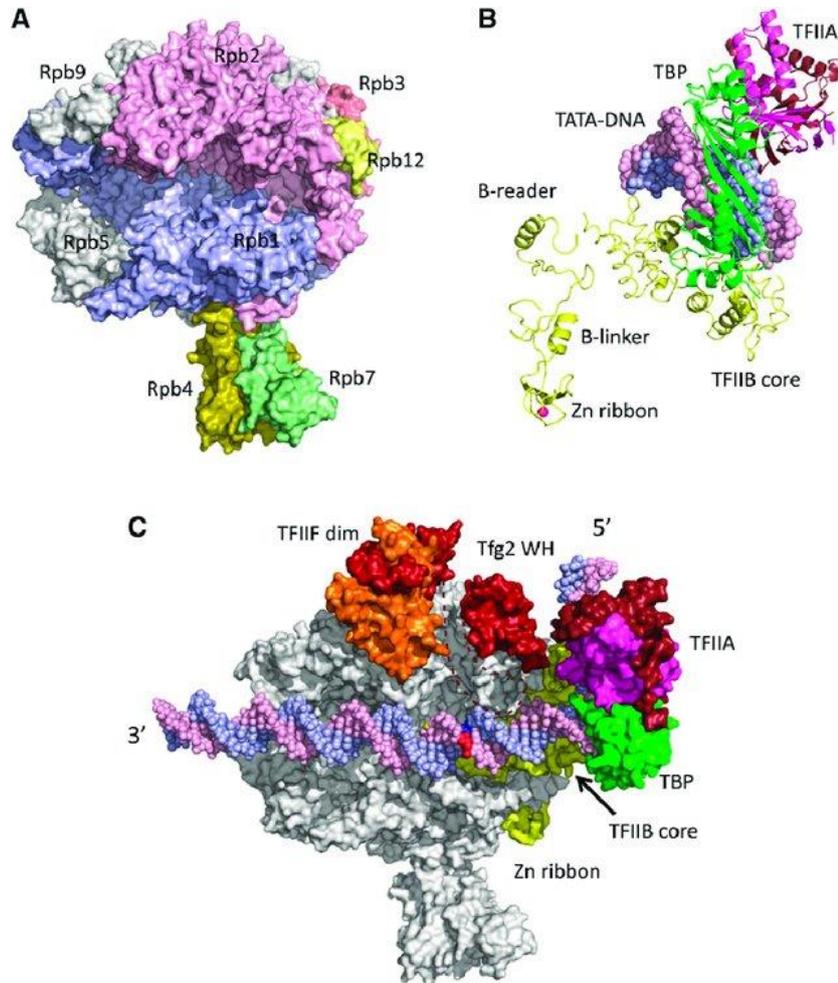


Figure 1 Structure de l'ARN polymérase II (Hahn and Young, 2011).
A) polII seule. B) Modèle de liaison à l'ADN de cofacteurs de polII au niveau de boîte TATA lors de la formation du PIC (complexe de pré-initiation). C) Modèle du PIC de la levure

1.2.1 RNA polymérase II et cofacteurs

PoIII assure une bonne partie du travail mais ne possède pas intrinsèquement toutes les fonctions nécessaires au bon déroulement de la transcription et de la biogenèse de la mRNP. Il existe donc, différents cofacteurs jouant des rôles clés à différentes étapes de la transcription dépendante de polII, ces derniers peuvent avoir un rôle dans le procédé de transcription ou bien dans les procédés de biogenèse du RNP.

Avant l'initiation de la transcription, des facteurs généraux d'initiations vont se lier à des séquences spécifiques d'ADN situées dans le promoteur et qui sont appelées « boîtes » (TATA, CAAT, GC, etc.). Ces facteurs appelés facteurs de transcription multifonctionnels ou généraux II (TF_{II} ; X pouvant être A, B, C, D, E, F, G, H) vont avoir différents rôles, comme la reconnaissance des « boîtes », l'ouverture de la double hélice d'ADN grâce à une activité hélicase, ou le transport de polII sur le site d'initiation. Certains autres cofacteurs se fixent à longue distance du site de transcription, sur des séquences appelé activateur (« enhancer »), qui vont activer à distance la transcription (par interaction avec le promoteur).

Lors de l'élongation, certaines protéines sont recrutées spécifiquement sur polII grâce à un mécanisme décrit dans la section suivante, afin de permettre sa progression le long de l'ADN ou à l'inverse lui faire marquer une pause.

Enfin, la terminaison est assurée par un complexe multi-protéique indépendant, qui va se fixer sur une séquence spécifique de l'ARN pour provoquer le clivage au site de terminaison du messager, puis polyadényler l'extrémité 3' de l'ARN.

Les phénomènes qui intéressent mon travail de thèse, vont avoir lieu tout au long des différentes étapes de la transcription. De fait, les différentes étapes de ce procédé biologique ainsi que les différentes voies aboutissant à la biogenèse d'une particule ribonucléoprotéique (RNP) fonctionnelle, seront abordés tout au long de ce manuscrit.

Ces différentes étapes de transcriptions conduisent à la formation d'un ARNm possédant une organisation qui lui est propre. En effet, cet ARNm code pour une protéine mais seule une partie de sa séquence est dédiée au codage de l'information qui définira la séquence de la protéine. A chaque extrémité de la molécule ribonucléique se trouve une séquence non codante appelé 5'- et 3'-UTR (« UnTranslated Region »). L'ARNm possède également une coiffe (décrite plus tard) en 5', ainsi qu'une queue polyA en 3' (Figure 2).



Figure 2 Schéma général d'organisation d'un ARN messager. Le 'C' à l'extrémité 5' symbolise la coiffe, les régions en noir représentent les UTRs, la région verte symbolise la séquence codante de la molécule messagère et le segment noir à l'extrémité 3' schématise la queue polyA.

1.2.1.1. RNA polymérase II : domaine C-terminal (CTD)

RPB1, la plus grosse sous-unité de l'ARN polymérase II présente une particularité nécessaire au bon fonctionnement du complexe enzymatique polII. En effet, le domaine C-terminal (CTD) de RPB1, appelé CTD de polII, est un domaine non structuré composé de répétitions (plus ou moins longues selon l'espèce : 26 répétitions pour la levure, 52 chez l'humain) d'un hepta-peptide (heptades). Les sept résidus d'acides aminés répétés en tandem dans le CTD sont Tyr¹-Ser²-Pro³-Thr⁴-Ser⁵-Pro⁶-Ser⁷. Il semblerait que les heptades fonctionnent en duos car lorsqu'une alanine est insérée entre toutes les heptades chez la levure, la cellule n'est pas viable, alors qu'une insertion toutes les deux heptades n'est pas létale (Stiller and Cook, 2004).

Le CTD de polII peut être vu comme une plateforme de recrutement spatio-temporel de différents acteurs i) de la maturation, de l'empaquetage et du contrôle qualité de l'ARN, ii) de recrutement de certains facteurs de modification de la chromatine et iii) de facteurs de transcription généraux.

Les différents résidus du CTD peuvent subir des modifications post-traductionnelles temporelles, principalement des phosphorylations sur la Ser², Thr⁴, Ser⁵ et Ser⁷, et dans une moindre mesure des glycosylations sur la Thr⁴ et les Ser ou encore une isomérisation des Pro (Figure3).

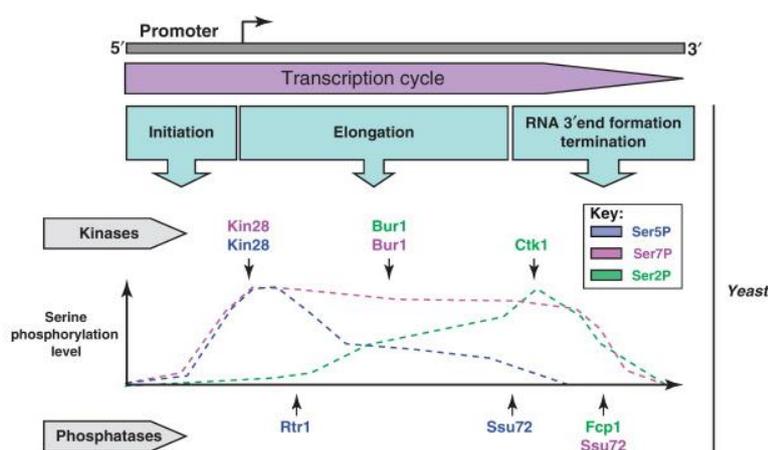


Figure 3 Phosphorylations des différents résidus du CTD de polII en fonction des étapes de la transcription (Egloff et al., 2012). Schéma de l'implication des différentes kinases et phosphatases de levure, impliquées dans le jeu de modifications des sérines du CTD. Ces modifications sont les marqueurs d'étapes de la transcription dans lesquelles se trouve polII ainsi que les acteurs du recrutement des différents facteurs de transcription, de maturation de l'ARN et d'empaquetage.

Ces phosphorylations sont essentielles à certaines étapes de la transcription et surtout vont permettre, pendant des étapes clés, le recrutement des différents facteurs protéiques cités précédemment. Ces facteurs vont alors pouvoir se fixer sur l'ARN en cours de transcription pour le maturer et/ou l'empaqueter (Harlen et al., 2016; Milligan et al., 2016) (Figure 4 & 5).

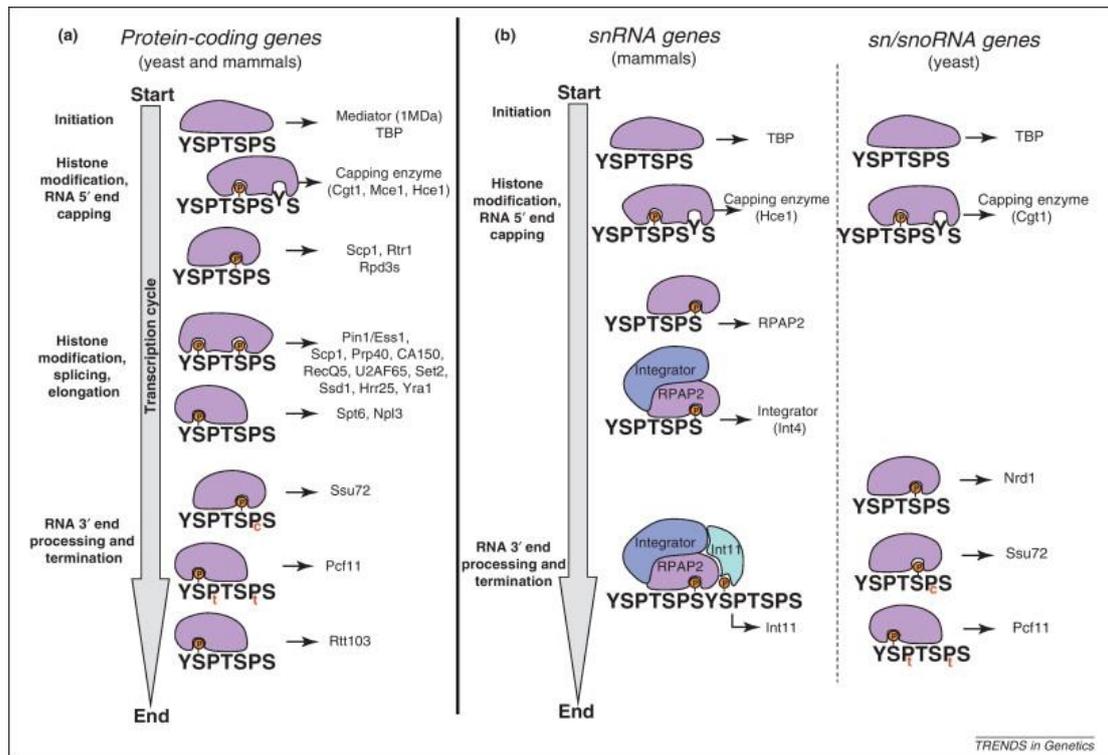


Figure 4 Rôle du CTD dans le recrutement de protéines sur les ARNs transcrits par polIII.

Protein/complex	Role in RNA biogenesis	Phospho-CTD bound	References
TFIIE	Preinitiation complex	Hypophosphorylated CTD	[63, 77]
TFIIF	Preinitiation complex	Hypophosphorylated CTD	[77]
TBP	Preinitiation complex (TFIID)	Hypophosphorylated CTD	[78]
Mediator Complex	Transcription activation/repression	Hypophosphorylated CTD	[48, 79]
Ceg1	Capping	Ser5-P	[80–85]
Abd1	Capping	PCTD	[83]
Set1	Histone methylation	Ser5-P	[86]
Rpd3C(Rco1)	Histone deacetylation	Ser2-P + Ser5-P	[87, 88]
Spt6	Histone chaperone	Ser2-P	[89]
Nrd1	Transcription termination/processing	Ser5-P	[90]
Sen1	Transcription termination/processing	Unknown	[91]
Asr1	Pol II ubiquitylation (Rpb4/7 Ejection)	Ser5-P	[92]
Ess1	Proline isomerase	Ser2-P	[93, 94]
Set2	Histone methylation	Ser2-P + Ser5-P	[95, 96]
Prp40	Splicing	PCTD	[97]
Npl3	Promotes elongation/prevents polyadenylation	Ser2-P	[98]
Pcf11	Cleavage/polyadenylation (CF1A)	Ser2-P	[99, 100]
Rna14	Cleavage/polyadenylation (CF1A)	PCTD	[101]
Rna15	Cleavage/polyadenylation (CF1A)	PCTD	[101]
Ydh1	Cleavage/polyadenylation (CPF)	PCTD	[102]
Yhh1	Cleavage/polyadenylation (CPF)	PCTD	[103]
Pta1	Cleavage/polyadenylation (CPF)	Ser5-P	[104]
Rtt103	5'-3' Exonuclease (Rat1)	Ser2-P	[105]
Sus1	mRNA export	Ser5-P	[106]
Yra1	mRNA export	Hyperphosphorylated CTD	[107]
Rsp5	Pol II ubiquitylation (DNA damage response)	Ser2-P	[108, 109]
Hrr25	DNA damage repair	PCTD	[24, 110]

Figure 5 Tableau récapitulatif des protéines connues pour se lier au CTD de polII chez *S. cerevisiae* (Zhang et al., 2011). Le tableau présente le complexe ou la protéine d'intérêt, le rôle dans la biogénèse et la marque de phosphorylation du CTD qui permet le recrutement. PCTD signifie un mélange de marque de phosphorylations (générées *in vitro*).

L'initiation et le début de l'élongation sont marqués par l'état phosphorylé des serines-5 des heptades du CTD, qui vont être ensuite progressivement déphosphorylés pour laisser place à une phosphorylation des thréonine-4 et finalement une phosphorylation de plus en plus marquée des serines-2 lorsque polII avance vers l'extrémité 3' jusqu'à un maximum de phosphorylation de ces résidus en fin de transcription.

Ce jeu de modifications post-traductionnelles du CTD a été appelé code du CTD (Buratowski, 2003; Cassart et al., 2012).

Très récemment, une implication encore plus importante du CTD dans la formation du complexe de pré-initiation de la transcription a été montrée. Dans ce modèle, la nature non structurée du CTD conduit plusieurs polII à interagir entre-elles par l'intermédiaire de ce domaine, induisant une séparation de phase liquide/liquide facilitant l'accès de ces dernières à l'ADN (Boehning et al., 2018). En revanche, ces interactions ne permettent pas d'initier

correctement la transcription, et c'est la phosphorylation des serines-5 des CTDs qui va conduire, par répulsion électrostatique, à désassembler les polymérase et permettre ainsi d'initier la transcription (Figure 6).

Le concept de séparation de phase est basé sur l'observation que de nombreux compartiments au sein des cellules sont dépourvus de membranes. Ce concept repose sur les propriétés physico-chimiques d'un corps/organelle comprenant une haute concentration de certaines composantes (e.g. protéines, ARN, etc). La différence de concentration de ces composantes entre cette espace supersaturé et l'extérieur vont spontanément créer une séparation physique (Boeynaems et al., 2018).

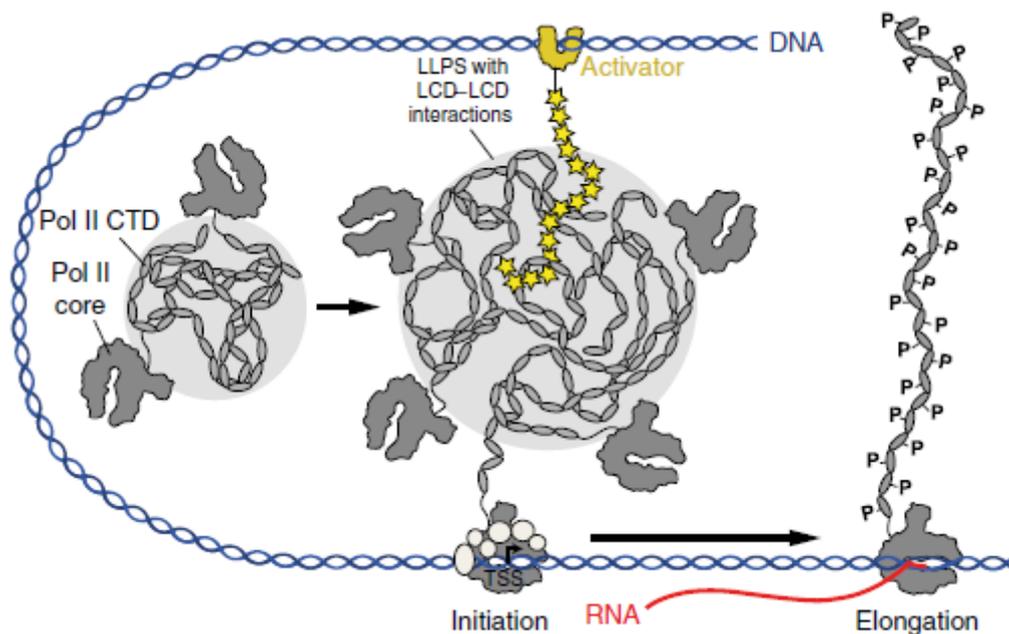


Figure 6 Modèle proposé par Boehning et al. (Boehning et al., 2018) pour l'initiation de la transcription à travers un nouveau rôle du CTD.

1.2.1.2 RNA polymerase II : acteur principal de la transcription des mRNAs et la biogenèse des mRNPs

L'ARN polymérase II a été décrite, en premier lieu, comme l'enzyme responsable de la synthèse de l'ARN messager. Nous avons également vu, que le complexe polIII est un des piliers de la biogenèse des mRNPs en servant de plateforme de recrutement aux acteurs de la maturation, de l'empaquetage et de l'export. Nous allons, dans la suite de ce manuscrit, insister sur chaque étape de la transcription des ARNm et sur les événements de maturations et d'empaquetages qui contribuent à la biogenèse des mRNPs. Nous allons également faire le parallèle avec la synthèse des ARNs non codants pour laquelle il existe certaines différences.

1.2.1.3 RNA polymerase II : transcription des longs ARNs non codant et des snRNAs

L'apparition des techniques de séquençages sur puces (microarray, *e.g* agilent), couplés à certaines expériences de génétiques (déplétion de la machinerie de dégradation des ARNs) ont permis de mettre au jour un tout nouveau paysage transcriptionnel (Figure 7). À partir de ce moment, la communauté scientifique a établi que la transcription n'était pas limitée aux ARNs codants, ribosomiques, de transferts, et autres petits ARNs bien décrits auparavant. Avec le développement de la nouvelle génération de séquenceurs (type illumina : Miseq, NEXTseq, etc) ces observations ont pu être confirmées. Ainsi, le concept de région non transcrite en ARN, qui correspondait aux quelque 80% d'ADN que l'on pensait jusque-là être non transcrit et sans réel intérêt, s'est vu radicalement changé. En effet, une large part des génomes sont transcrits, et cette transcription pervasive d'ARNs non codant et pour la plupart long (> 200 nt), a ouvert un tout nouveau champ d'étude.

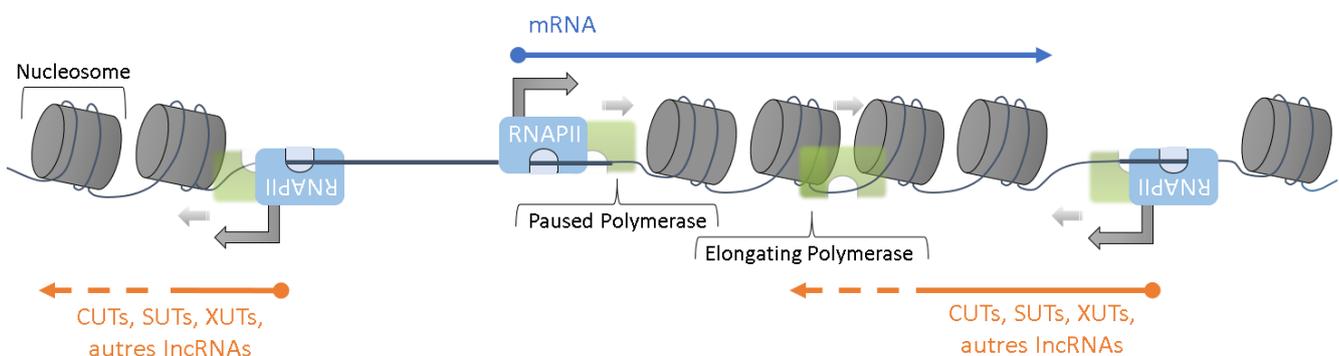


Figure 7 Schéma illustrant certaines origines du nouveau paysage transcriptionnel. Les IncRNAs peuvent être issus d'événements de transcriptions initiés par des promoteurs bidirectionnels de gènes codants, ou encore de région intergénique, etc.

Le concept de transcription pervasive mis en évidence par Wyers *et al.*, en 2005 repose sur la capacité de polII à initier des événements de transcription à partir de régions d'ADN anciennement définies comme transcriptionnellement inactives. Les ARNs issus de cette transcription, naissant de régions distinctes des ARNm, sont nommés ARNs non codant.

Les longs ARNs non codants (lncRNAs) sont synthétisés par polIII et peuvent avoir différentes origines génomiques. Les lncRNAs peuvent provenir des mêmes promoteurs que les ARNm mais sont transcrits à partir du brin d'ADN inverse complémentaire (Xu *et al.*, 2009). Ils peuvent aussi trouver leur origine dans les régions introniques des gènes, dans les séquences inter-géniques, dans les régions activatrices, et être transcrits en anti-sens d'autres classes d'ARNs. Ces diverses origines de transcriptions, ne permettent pas de leur attribuer un rôle précis par similarité de production avec une autre classe d'ARN. Un autre fait surprenant accompagne ces lncRNAs : ces transcrits possèdent de nombreuses similarités avec les ARNm. Ils peuvent être « coiffés » à leur extrémité 5', sont polyadénylés à leur extrémité 3', et certains d'entre eux sont également épissés.

Malgré leurs diverses origines dans le génome et leur forte ressemblance avec les ARNm, certains lncRNAs ont une durée de vie extrêmement courte. Mise à part certains d'entre eux qui semblent s'exprimer et être maintenus pour jouer un rôle à certaines étapes clés du développement embryonnaire (par exemple (XiST) chez les mammifères), la plupart des lncRNAs sont dégradés juste après leur production. Ce dernier fait explique que la communauté scientifique n'ait pas pu les observer avant l'apparition de technologies de séquençage, couplé à des outils de génétiques classiques, permettant d'observer ce genre d'évènements.

1.2.2 Empaquetage et Maturation des ARNm

La transcription par polIII comprend les ARNm et une multitude d'ARNs non codants plus ou moins long. Nous allons donc maintenant décrire comment se déroule la biogenèse des RNPs pour les molécules messagères qui sont les mieux décrites, puis celle des ARN non codants.

La biogenèse des mRNPs se déroule de manière co-transcriptionnelle lors de l'étape de synthèse (Bentley, 2014). En effet, comme introduit précédemment, un grand nombre de facteurs protéiques interviennent dans la maturation et l'empaquetage de l'ARN de manière co-transcriptionnelle.

Ces derniers sont recrutés, pour la plupart, par les acides aminés modifiés du CTD de polIII. Suite à une immuno-précipitation de la polymérase II avec des anticorps dirigés contre chaque résidu de l'heptade modifiée (Harlen et al., 2016), des expériences de spectrométrie de masse ont permis de mettre à jour les différentes protéines qui se lient à polIII pendant le cycle de transcription (Figure 8). Ces résultats ont permis d'élucider certains événements majeurs de maturation de l'ARN et surtout d'identifier les acteurs de ces événements.

Il existe de nombreuses étapes qui conduisent à la formation d'une mRNP mature pendant la transcription de l'ARN. Les sous-chapitres suivants décrivent un aperçu de ces étapes.

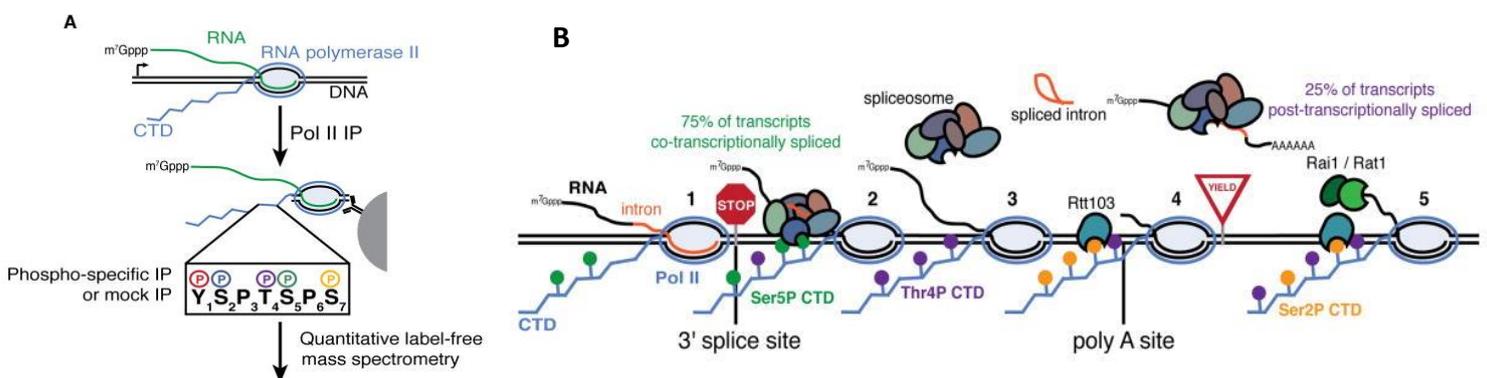


Figure 8 Schéma de la technique utilisée par Harlen et al., 2016 (Harlen et al., 2016) pour déterminer les interactants du CTD et résumé des résultats. A) Technique d'immuno-précipitation spécifique de phospho-résidus du CTD de polIII. B) Facteurs de maturations de l'ARN interagissant avec des résidus phosphorylés particuliers de polIII à des étapes clés de la biogenèse de la mRNP.

1.2.2.1. Coiffe et épissage, initiation et élongation

Lorsqu'un événement de transcription doit être initié, la polymérase II dont le CTD n'a subi aucune phosphorylation est recrutée au promoteur par les facteurs généraux de transcription (TF_{II}) ainsi que par les facteurs de transcription spécifiques du gène à transcrire. Ce super complexe est appelé complexe de pré-initiation (PIC) (Murakami et al., 2013; Zhang et al., 2012). L'initiation de la transcription débute lorsque le CTD se fait phosphoryler au niveau des Ser5 par la kinase Kin28 (qui fait partie intégrante du facteur de transcription général TFI_{II}H). Cette modification vers l'état phosphorylé de la sérine5 permet à la polymérase d'échapper au promoteur et ainsi marque le début de la transcription à proprement parlé.

Du point de vue de la biogenèse de la mRNP, cette modification de la Ser5, en plus de permettre à la transcription de s'initier, va également permettre le recrutement des protéines d'ajout de la coiffe en 5' de l'ARN (Cho et al., 1997). Ce complexe permet de protéger l'ARN naissant de la dégradation par des exonucléases 5'-3', en ajoutant à l'extrémité 5' dudit ARN, une guanosine méthylée en position N7 (7mG), appelé coiffe (Suh et al., 2010). Le procédé par lequel passe cet ajout de base modifiée chez les eucaryotes est découpé en 3 étapes : I/ hydrolyse du triphosphate en 5' de l'ARN en diphosphate par l'ARN triphosphatase Cet1p, II/ liaison d'une GMP (guanosine monophosphate) au 5' diphosphate en liaison 5'-5' par la guanylyltransférase Ceg1p, III/ cette guanosine est méthylée par la méthyltransférase Abd1p. Les trois enzymes précitées sont celles de la levure. Ces trois étapes conduisent à la production d'une coiffe fonctionnelle schématisée par l'expression "7mGpppN". Un complexe de fixation à la coiffe (CBC pour *Cap Binding Complex*) se lie alors à cette 7mG lorsque polII marque une pause dans l'élongation de l'ARN dans une fenêtre d'environ 100 pb après le site d'initiation de la transcription (Churchman and Weissman, 2011). Le CBC est une pierre angulaire de l'export de la mRNP vers le cytoplasme. Une fois exporté dans le cytoplasme, c'est le facteur de traduction eIF4F qui va se fixer à la coiffe, pour assurer la protection contre la dégradation 5'-3' et permettre l'initiation de la traduction.

L'étape suivante consiste au désassemblage du PIC par les facteurs généraux d'élongation. L'entrée en phase d'élongation, est marquée par la phosphorylation progressive de la serine 2 du CTD. Les premières phosphorylations de la Ser2 sont effectuées par la kinase Bur1 qui est recrutée via les phospho-Ser5 du CTD. Les événements suivants de phosphorylation de la Ser2 sont assurés par la kinase Ctk1 et vont permettre l'entrée en phase

d'élongation. Cette augmentation du taux de Ser2 phosphorylées sur le CTD s'accompagne d'une diminution du taux de phospho-Ser5, les marques de phosphorylation de cette sérine étant retiré par la phosphatase Rtr1. Ce nouveau phospho-résidu majoritaire qu'est la Ser2, va permettre le recrutement d'autres facteurs de maturation de l'ARN.

Un des premiers complexes de maturation de l'ARN à être recruté est le complexe d'épissage (spliceosome). Ce dernier, composé d'environ une centaine de protéines et dont l'activité dépend de la présence de 5 snRNAs, voit certains des sous-complexes qui le constitue se lier au CTD. En plus d'être lié à l'état de phosphorylation du CTD, l'épissage est directement dépendant de la présence du CBC. De surcroît, la machinerie d'épissage possède également une fonction de régulation du point de vue du recrutement des machineries de clivage et polyadénylation à l'extrémité 3' de l'ARN néo-synthétisé (Herzel et al., 2017).

1.2.2.2. Clivage et polyadénylation, terminaison

La terminaison de la transcription, chez la levure, est un procédé intensément étudié pour plusieurs raisons : I/ les découvertes récentes sur l'existence des lncRNAs amènent à se demander comment ces derniers sont terminés pour être dégradés peu après leurs productions et quelles sont leurs différences avec les mRNAs, II/ cette étape va définir la forme finale du transcriptome, à savoir, la taille des unités transcriptionnelles, et III/ va ainsi pouvoir prévenir des évènements de transcriptions sens et anti-sens au niveau d'une même région génomique qui pourrait être une source de perturbation pour la cellule, et conduire à l'accumulation de produits ribonucléiques non fonctionnels.

Il existe plusieurs voies de terminaison pour les transcrits polIII en fonction de la nature de l'ARN transcrit. La première voie concerne les ARNm. Elle passe par les complexes de clivage et polyadénylation, CPF & CF. La seconde voie termine les ARNnc comme les sn/snoRNAs et prépare les lncRNAs à la dégradation. Ce deuxième mécanisme est assuré par le complexe Nrd1-Nab3-Sen1 (NNS).

La terminaison de la transcription des gènes codants des protéines est principalement dépendante de trois complexes protéiques : CPF, CFIA et CFIB (complexes conservés au travers de l'évolution chez les eucaryotes). Plusieurs protéines du complexe CPF (telles que Rna15, Cft1, Cft2, Yth1, Mpe1 et Hrp1) ont la capacité de se fixer sur des signaux de terminaisons

présents sur le brin d'ARN, dans sa région 3'UTR. Le recrutement des complexes précédemment cités passe tout d'abord par le recrutement de Pcf11 par le CTD de polII lorsque les Ser2 sont phosphorylées, puis la fixation aux motifs de terminaisons d'autres protéines du CPF (Lunde et al., 2011). L'ARN est ensuite clivé par la sous-unité Ysh1 au niveau de son site de polyadénylation (pA), puis la poly(A) polymérase Pap1 vient ajouter des adénosines au groupement hydroxyle libre de l'extrémité 3'. Une fois la queue poly(A) synthétisée, une ou plusieurs protéines de liaison à la poly(A), poly(A)-binding protein viennent s'y fixer et vont affecter la taille de cette dernière, assurer la protection de l'ARN contre les exonucléase 3'-5', et participer à la capacité d'export de la molécule ribonucléoprotéique. Cependant, l'identité des protéines de liaison à la poly(A) reste controversée. Pab1 et Nab2 ont été proposés comme protéines pouvant se lier à la queue poly(A) très tôt lors de sa genèse, et pourraient potentiellement escorter l'ARN jusqu'au cytoplasme (Dunn et al., 2005; Hector et al., 2002). Les ARNs terminés par la voie CPF-CF sont exportés rapidement vers le cytoplasme. Leur demi-vie est généralement déterminée par le système de turnover cytoplasmique. L'export vers le cytoplasme est décrit comme étant rapide et non limité (Mouaikel et al., 2013). Ce dernier fait a été suggéré comme permettant de ne pas surcharger la machinerie de dégradation des ARNs nucléaires, lui laissant alors le champs libre pour dégrader et mûrer ses cibles (Gudipati et al., 2012).

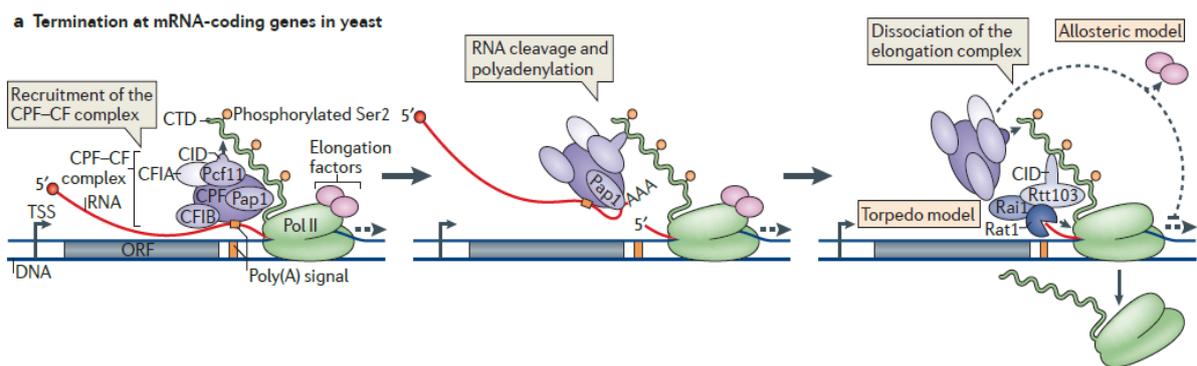


Figure 9 Modèle de terminaison de la transcription via les complexes CPF-CF chez la levure (adaptée de Porrua et al., 2015). Illustre en 3 étapes le recrutement du complexe de clivage, son action, puis la terminaison de la transcription via le modèle torpedo.

Ces étapes de terminaison de la transcription et de maturation de l'extrémité 3' du transcrit sont intimement liées, et il est compliqué de définir où se situe la limite entre les complexes impliqués dans la terminaison et ceux impliqués dans la maturation, sachant que nombre d'entre eux jouent un rôle dans les deux procédés. L'étape limitante de la terminaison est la dissociation de polII de la matrice ADN. Deux modèles de terminaison de la transcription des gènes codants des protéines ont été proposés. Le premier modèle appelé « modèle allostérique » propose que lorsque la transcription passe par le site de polyadénylation, un changement conformationnel du complexe d'élongation aurait lieu et provoquerait la dissociation de polII de l'ADN. Le second modèle, appelé modèle « torpedo », implique l'exonucléase 5'-3' Rat1, qui une fois le clivage de l'ARNm effectué pourrait se fixer sur l'extrémité 5', non protégée, de l'ARN sortant de polII. Rat1 dégraderait alors l'ARN jusqu'à rattraper la polymérase et provoquer sa séparation de l'ADN matrice (Mischo and Proudfoot, 2013) . Ces deux modèles, bien que séduisants, n'ont pas pu être confirmés avec certitude et le modèle « torpedo » s'est montré insuffisant à une dissociation efficace de polII de la matrice ADN (Schaughency et al., 2014). En effet, une étude publiée en 2006 montre que la seule inactivation de Rat1 dans une souche mutante de levure thermosensible à 37°C (*rat1-1*), ne permet pas d'observer l'apparition d'un ARN naissant en 3' du gène terminé. En revanche, cette même publication propose un modèle dans lequel Rat1 serait à la fois impliqué dans le procédé de terminaison et celui de maturation du 3' des ARNs (Luo et al., 2006). Ce modèle est supporté par le fait que la mutation de Rat1 dans cette étude, inhibe le recrutement des facteurs de terminaisons Pcf11 et Rna15. Une fois de plus, le lien entre le clivage de l'ARNm transcrit et la terminaison de la transcription apparaît comme étant très étroit.

Comme présenté précédemment, la transcription dépendante de polII n'est pas limitée aux ARN codants pour des protéines, une multitude d'ARNs non codants sont également le produit de la transcription par cette polymérase. Nous nous focaliserons sur les snoRNAs et les lncRNAs (plus spécifiquement les CUTs) plus tard dans ce manuscrit.

La terminaison de ces unités de transcription non codantes est assurée par le complexe NNS, composé des protéines Nrd1, Nab3, deux protéines de liaisons à l'ARN et Sen1, une RNA & DNA hélicase. Le complexe NNS ainsi que la terminaison et maturation de ces transcrits sera décrit en détail plus tard dans ce manuscrit.

Pour l'heure, un dernier évènement de biogenèse des mRNPs se doit d'être abordé de par son importance capitale dans la « vie » post-transcriptionnelle de ces derniers : l'empaquetage et la compétence à l'export.

1.2.3 Empaquetage et export des ARNm : exemple du complexe THO

Une série d'évènements de maturation de l'ARN naissant ont lieu au cours de la transcription. Ces évènements de maturations, comme l'ajout de la coiffe en 5' (décrit précédemment) ou encore l'épissage des introns, sont accompagnés par la fixation de diverses protéines d'empaquetage le long du brin d'acide ribonucléique. Ces protéines d'empaquetage vont assurer, entre autre, l'intégrité de l'ARN ainsi que sa capacité à être exporté dans le cytoplasme pour y être traduit en protéine. La molécule d'ARN est assemblée avec les protéines pour donner la particule ribonucléoprotéique nommée mRNP, s'il s'agit d'un ARN messager, snRNP si l'assemblage concerne un sn/snoRNA, etc.

En plus des machineries de transcription et de maturation (complexes de « capping », machinerie d'épissage, les complexes protéiques de terminaison et poly(A)), les mRNPs vont être liés par une multitude de protéines de liaison à l'ARN (RBPs). Ces dernières peuvent elles-mêmes recruter d'autres facteurs protéiques par interaction protéine-protéine (Mitchell and Parker, 2014). Une fois le clivage et la polyadénylation effectués, la mRNP doit être suffisamment compactée pour permettre son passage à travers les pores nucléaires vers le cytoplasme (Aguilera, 2005; Mor et al., 2010; Rougemaille et al., 2008; Saguez and Jensen, 2009).

Chez la levure, un des complexes les plus importants dans le procédé d'empaquetage pour l'export est le complexe THO. Ce dernier se fixe sur l'ARNm pendant la phase d'élongation de la transcription et va pouvoir servir de plateforme de recrutement à une variété de protéines dont des facteurs d'exports pendant la biogenèse pour former le complexe TREX (Strasser et al., 2002).

1.2.3.1 Composition et recrutement du complexe THO

Le complexe THO est composé de 4 sous-unités interagissant fortement entre-elles (Chavez et al., 2000; Piruat and Aguilera, 1998) (Figure 10). Ces sous-unités sont Tho2, Hpr1, Mft1 et Thp2. La stabilité du complexe repose sur les 4 sous-unités, car la déplétion de l'une entraîne une baisse du niveau protéique des 3 autres (Chavez et al., 2000; Huertas et al., 2006). Pour former le complexe TREX, le complexe THO va interagir avec Sub2, Yra1 et Tex1 (cette dernière pouvant être considérée comme partie intégrante du complexe THO, bien que non indispensable à son intégrité). C'est ce dernier super-complexe qui possède la fonction d'export de la mRNP du noyau vers le cytoplasme (Strasser et al., 2002).

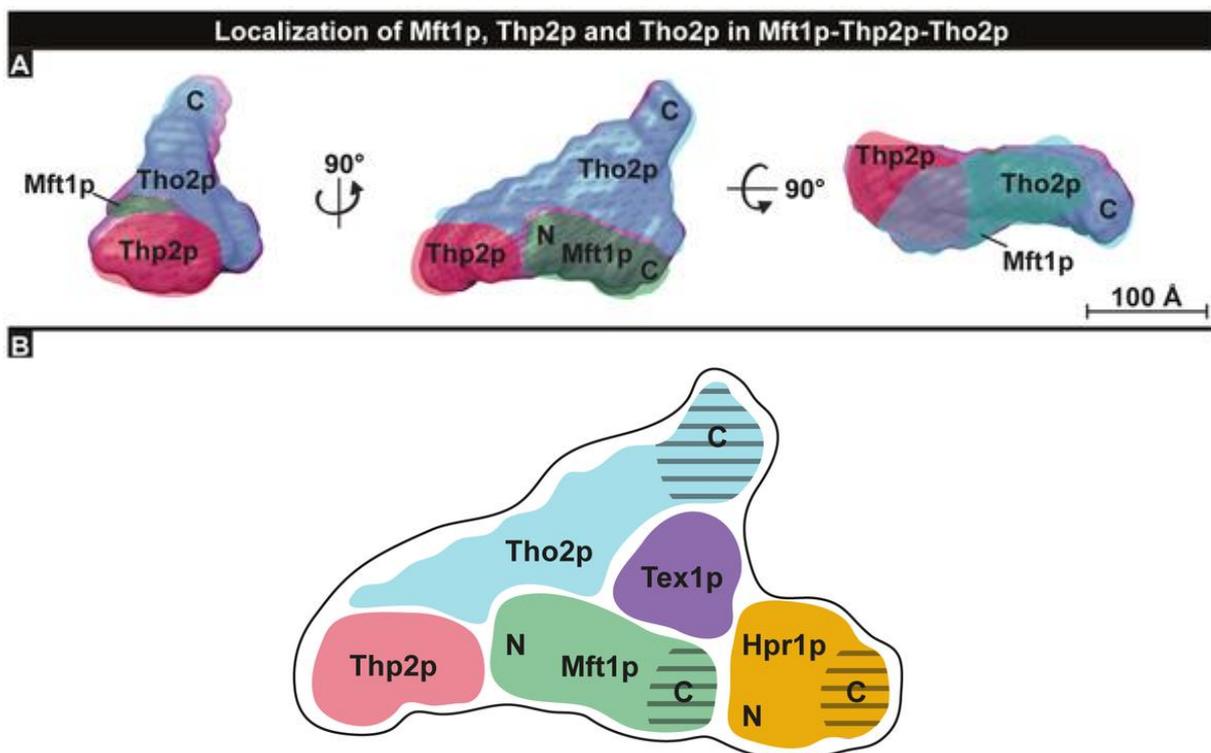


Figure 10 **Modélisation et schéma du complexe THO.** A) Modèle proposant les localisations et orientation des protéines au sein du complexe THO. B) Modèle schématique de l'arrangement des sous-unités du complexe THO issu de résultats de microscopie électronique et de SAXS. Les zones hachurées représentent les zones prédites comme étant flexible. (Poulsen et al., 2014)

Le complexe THO est recruté, de manière co-transcriptionnelle sur le transcrit en cheminant d'abord par le CTD de la polymérase lorsque le domaine est phosphorylé essentiellement sur la Ser2. Le recrutement repose également sur la capacité du domaine C-terminal de la sous-unité Tho2 à se lier aux acides nucléiques en général (ADN et ARN). Divers travaux ont montré que la quantité de complexe THO augmente de l'extrémité 5' à celle en 3' d'un gène, à l'image de l'augmentation de la phosphorylation des Ser2 du CTD qui augmente lors de l'élongation le long du gène jusqu'à atteindre un pic en 3'. Il y'a donc un lien entre le CTD de la polymérase et le complexe THO (Luna et al., 2012; Mayer et al., 2010; Meinel et al., 2013). Cette caractéristique d'augmentation de la présence du complexe vers l'extrémité 3' le distingue des autres complexes recrutés lors de la phase d'élongation de la transcription et plus généralement des protéines recrutées par la phospho-Ser2 du CTD (Meinel et al., 2013). L'intensité croissante de la présence du complexe THO/TREX en allant vers l'extrémité 3' semble être une condition nécessaire au bon empaquetage de la mRNP, et donc à sa capacité d'être exportée dans le cytoplasme. La présence du complexe semble également être un prérequis à une bonne polyadénylation des ARNm ainsi qu'au couplage de la terminaison, du processing de l'extrémité 3' et de l'export par le pore nucléaire (Bonnet et al., 2015; Rougemaille et al., 2008; Saguez et al., 2008).

1.2.3.2 Fonctions du complexe THO

Une des premières fonctions du complexe THO à avoir été démontrée est sa capacité à empêcher la formation de R-loop, hybride ADN:ARN formé à l'arrière de polII pendant la phase d'élongation (Gomez-Gonzalez et al., 2011) (Figure 11). La déplétion d'une des sous-unités de ce complexe se traduit par un défaut d'élongation à cause de la formation de R-loops, et par un défaut de production d'ARNm compétents pour l'export (Jimeno et al., 2002).

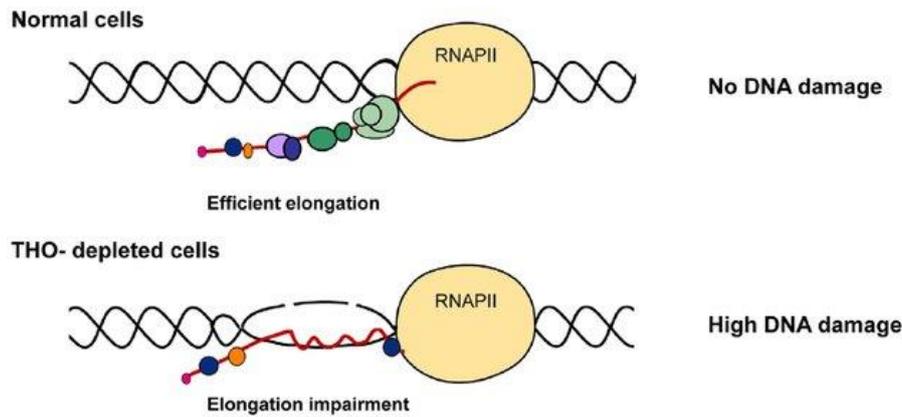


Figure 11 Schéma de la formation de R-loop en absence de certaines sous-unités du complexe THO et dommages à l'ADN associés. (Adaptée de Dominguez-Sanchez et al., 2011)

Certaines études ont démontré (Rougemaille et al., 2008) qu'en condition de choc thermique dans une souche mutante pour le complexe THO, une accumulation d'agrégats composés d'acides nucléiques et de protéines est observée, certaines protéines sont impliquées dans le clivage et la polyadénylation de l'extrémité 3' (Eshleman et al., 2016). Ce phénomène montre que le complexe THO possède également une fonction dans le contrôle des étapes de maturation de l'ARNm.

Une introduction plus approfondie et dédiée au complexe THO sera présentée plus loin dans ce manuscrit (dans la partie résultat).

Ce tour d'horizon des modifications apportées à l'ARNm en court de transcription nous a permis de bien établir les bases de ce qu'est la biogenèse des mRNPs. Les chapitres qui suivent vont dépeindre certaines étapes de la transcription et de la maturation des ARNs non codants auxquels nous nous sommes également intéressés au cours de ma thèse.

1.2.3.3. Stabilité de la mRNP et compétence à l'export

L'ensemble des procédés de maturation et d'emballage, décrit jusqu'à présent, et aboutissant à la formation d'une mRNP mature sont indispensables à sa stabilité et nécessaire à son export dans le cytoplasme. Certains facteurs qui empaquète l'ARNm seront aussi essentiels pour la traduction. Une mRNP compétente à l'export sera donc définie comme le produit transcriptionnel d'un gène codant, mûri par un ensemble de facteurs intervenant de manière coordonnée au cours de la transcription, et fixé par de nombreuses protéines (ou complexe protéique) indispensables à son export et à son intégrité. De plus cette définition peut être complétée par le fait qu'une mRNP est compétente à l'export si les protéines de contrôle qualité (décrites plus tard dans ce manuscrit) ne détectent aucune malformation.

L'assemblage cotranscriptionnel d'un ARNm en mRNP contribue à la stabilité de l'acide nucléique en le protégeant de la dégradation par les nucléases. En effet, l'ajout de la coiffe à l'extrémité 5' de l'ARN et la fixation du complexe CBC à cette dernière va protéger l'ARN des exonucléase 5'-3' (e.g. Rat1, Xrn1). De même l'ajout d'une queue polyA à l'extrémité 3' de ce même ARN va permettre la fixation de complexe de liaison à cette extrémité modifiée et ainsi protéger de la dégradation par des exonucléase 3'-5' (e.g. Rrp6, Dis3).

Le tableau suivant présente certains facteurs de biogénèse des mRNPs et d'export jouant un rôle à des étapes clés de la maturation et de la préparation à la capacité d'export de ces mRNPs. Ces facteurs ont été étudiés par Baejen et al., en 2014 via la technique de PAR-CLIP. Cette étude a montré comment ces facteurs reconnaissent le pre-mRNA *in vivo*. Ils sont donc des acteurs primordiaux de la formation d'une mRNP conforme, stable et compétente à l'export dans le cytoplasme pour la traduction.

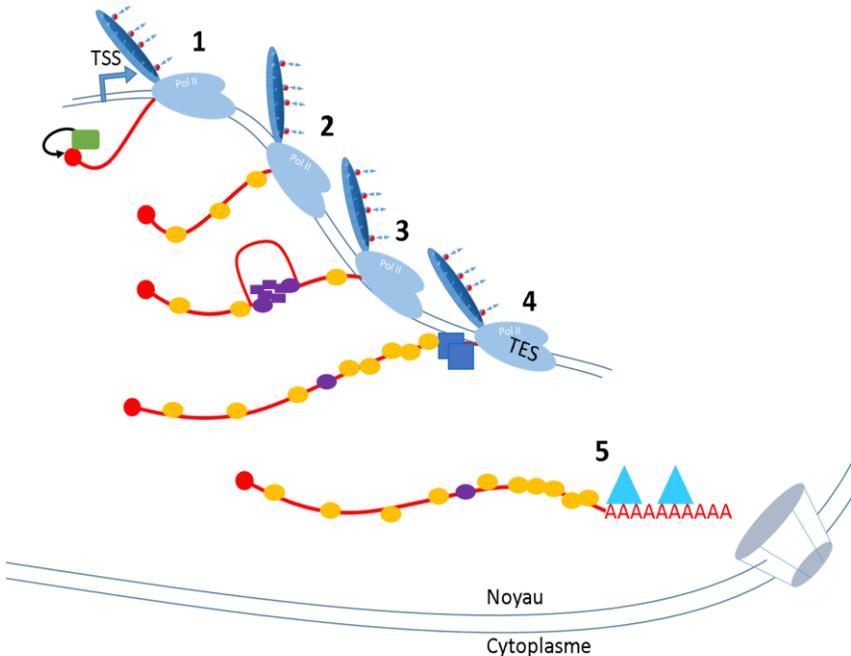
Un schéma général de la biogénèse de l'ARNm reprenant les grandes étapes précédemment décrite est proposé Figure 13.

Table 1. mRNP Biogenesis Factors Analyzed here by PAR-CLIP

Biogenesis Event	Factor/ Subunit	Complex	RNA-Binding Domain ^a	PAR-CLIP Crosslink Sites	False Discovery Rate (%)
Capping	Cbc2	CBC	RRM	98,034	0.178
Splicing	Luc7	U1 snRNP	ZF	93,261	1.035
	Mud1	U1 snRNP	RRM	99,384	1.918
	Nam8	U1 snRNP	RRM	151,813	1.675
	Snp1	U1 snRNP	RRM	25,493	0.447
	Ist3	U2 snRNP	RRM	66,003	3.184
	Mud2	BBP-U2AF65	RRM	801,430	1.769
	Msl5	BBP-U2AF65	ZN	476,370	1.961
3' processing	Rna15	CFIA	RRM	582,756	3.463
	Mpe1	CPF	ZF	122,500	2.262
	Yth1	CPF (PFI)	ZF	59,049	3.432
	Cft2	CPF (CFII)	-	189,866	1.723
	Pab1	-	RRM	233,513	2.052
	Pub1	-	RRM	371,902	1.332
	Export	Hpr1	THO/TREX	-	249,887
Tho2		THO/TREX	-	400,965	1.064
Sub2		TREX	-	228,620	1.085
Mex67		TREX	-	288,579	1.010
Yra1		Export adaptor	RRM	400,156	0.681
Nab2		Export adaptor	ZF	283,606	2.413
Npl3		Export adaptor	RRM	770,240	1.282
Hrb1		SR-like	RRM	395,402	0.976
Gbp2		SR-like	RRM	65,692	0.182

^aRRM, RNA recognition motif; ZF, zinc finger domain.

Figure 12 Tableau récapitulatif des facteurs de biogénèse des mRNP analysés par la technique de PAR-CLIP par Baejen et al., 2014. Le tableau indique l'étape de biogénèse, le facteur étudié, le complexe dont il fait partie, le type de domaine de liaison à l'ARN qu'il possède, le nombre de sites total de liaison (par la méthode utilisée) ainsi que la statistique de taux de « faux positif ».



Étapes	Facteurs (liste non exhaustive)	Fonctions	État de phosphorylation des sérines 2 et 5 du CTD
1	Cet1, Ceg1, Abd1, CBC	Ajout de la coiffe (extrémité 5'); Fixation du complexe de liaison à la coiffe	Phospho-Ser5 Phospho-Ser2
2	Complexe THO (Tho2, Hpr1, Thp2, Mft1)/TREX (THO + Yra1, Sub2, Tex1, Mex67)	Assemblage de la mRNP	
3	Complexe d'épissage ~100 protéines (dont le complexe Prp19, Mud2, etc)	Épissage	
4	Complexe CPF (dont Cft1, Cft2, Ydh1, Mpe1, Ysh1, etc); complexe CFIA (dont Rna14, RNA15, Clp1, Pcf11); CFIB (dont Hrp1); Pap1; Pab1, Nab2, etc	Clivage et polyadénylation, Fixation à la queue polyA	
5	Mex67, Mtr2, Npl3, Nab2, Gbp2, Hrb1, complexe TREX-2, Prp19, Mud2, protéines du pore nucléaire	Export Noyau -> Cytoplasme	

Figure 13 Schéma général de la biogénèse des mRNP & tableau récapitulatif des facteurs impliqués. Le schéma reprend les grandes étapes de la biogénèse des mRNP, le code couleur est conservé entre le schéma et le tableau. La liste des facteurs de maturation et d'emballage est à titre indicatif et n'est pas exhaustive. La colonne 1 sert de repère par rapport au schéma, la seconde colonne cite certains facteurs clés de la biogénèse, la troisième colonne indique la fonction de ces facteurs, la dernière colonne indique grossièrement le taux de phosphorylation des résidus d'acide aminés Ser5 et Ser2 du domaine CTD de polII.

1.2.4. Les ARNs non codants : Terminaison et Maturation

Bien que la transcription des ARNs non codants (courts ou longs) dépende de polII, les caractéristiques de leur biogenèse sont bien différentes de celle des ARNm. Les complexes protéiques qui participent à leur terminaison et leur maturation sont distincts de ceux contribuant à la biogenèse des ARN codants. Ces ARNs non codants sont différents des ARNs messagers en plusieurs points : I/ leurs fonctions, II/ leurs localisations cellulaires, III/ leurs durées de vie (pour les lncRNAs), IV/ leurs origines génomiques, V/ leurs structures et les événements de maturation qu'ils subissent. Ces différences fondamentales entre ARN messagers et ARN non codants font de ces derniers des « cibles » de maturations par des complexes protéiques différents de ceux dédiés aux ARNm. Afin de décrire et d'expliquer clairement la diversité des voies de terminaison des ARNs non codants, cette section va se découper en deux parties, une sur les sn/snoRNAs et une sur les longs ARNs non codants avec pour exemple les CUTs.

Classe d'ARNnc	Nombre de gènes
ARNt	299
ARNr	1
snRNA	6
snoRNA	77
CUTs	1496
SUTs	847
NUTs	~1500
SRATs	853
XUTs	1658

Figure 14 Tableau récapitulatif du nombre d'ARNs non codants par classe chez la levure *S.cerevisiae*.

1.2.4.1. Les snoRNAs chez la levure

Les snoRNAs, au nombre de 77 chez la levure, sont de petits ARNs non codants qui sont subdivisés en deux classes sur la base de motifs conservés : les snoRNAs à boîte C/D et les snoRNAs à boîte H/ACA (Figure 15). Les deux classes de snoRNAs sont impliquées dans la maturation et la modification des ARNs pré-ribosomiaux. Elles se distinguent également par leurs structures secondaires distinctes, par leurs capacités à se lier à différentes protéines et surtout par leurs capacités à guider des réactions de modifications de l'ARN différentes : 2'-O-

ribose méthylation (C/D) et pseudouridylation (H/ACA) (Kiss et al., 2002). La majorité des snoRNAs possèdent une région complémentaire aux ARNs ribosomiaux (Decatur and Fournier, 2002). Les modifications apportées par les snoRNAs sur les ARNr vont se retrouver en cluster dans des régions fonctionnelles importantes au sein du ribosome. Certains snoRNAs des deux classes vont également avoir une implication dans le clivage des pré-ARNr (Tollervey and Hurt, 1990). Récemment, les percées dans le domaine du séquençage ont permis de cartographier les bases modifiées de l'ARNr et de décrypter les zones d'interactions inter ARN. Ces informations ont ainsi soulevé l'hypothèse que certaines modifications sur les bases des ARNs messagers pouvaient être guidées par les snoRNAs chez l'homme (Sharma et al., 2016).

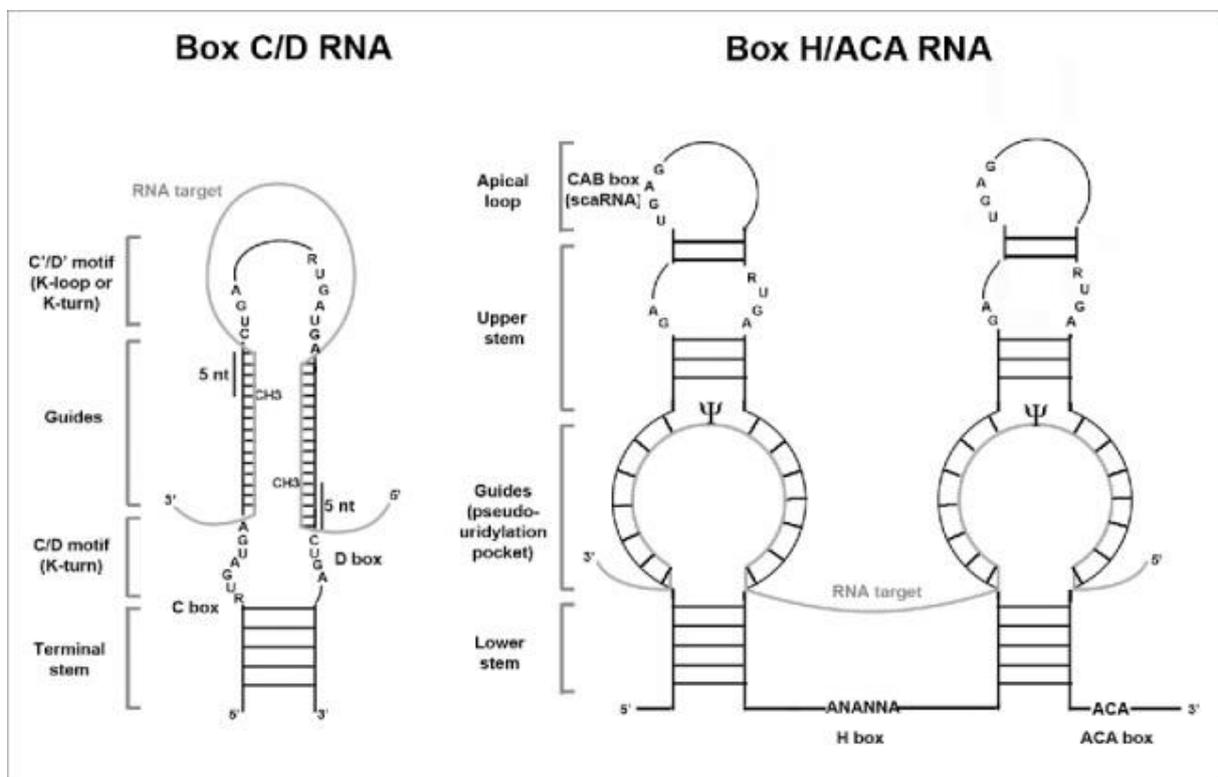


Figure 15 Différences entre les snoRNAs à boîte C/D et H/ACA. Schématise la structure des deux différentes boîtes retrouvés dans les deux classes de snoRNAs, en précisant le rôle de chaque composante des structures (Massenet et al., 2017).

Les snoRNAs peuvent fixer différentes protéines en fonction de leurs classes, et former des snoRNPs fonctionnelles. Le recrutement co-transcriptionnel des protéines composant le snoRNP est nécessaire à la fonction. Par ailleurs, ce recrutement est essentiel pour protéger les extrémités du snoRNA de la dégradation par des exonucléases et il est aussi important pour la stabilité finale du snoRNP mature (Samarsky et al., 1998). Pour former un snoRNP mature, l'extrémité 3' doit être maturée par l'ARN exosome aidé de différents adaptateurs comme par exemple la protéine Rrp47 (Costello et al., 2011) qui sera décrite ultérieurement.

La maturation de l'extrémité 3' est une étape clé dans la synthèse/production d'un snoRNP fonctionnel. Cette étape de maturation est intrinsèquement liée à la terminaison de la transcription de cette unité transcriptionnelle. Chez la levure, le système de terminaison de la transcription des snoRNAs passe par le recrutement du complexe NNS composé de 3 protéines : Nrd1, Nab3 et la dernière étant potentiellement Sen1. Suite à la terminaison de la transcription, la maturation à proprement parler sera assurée par l'exosome et sa sous-unité catalytique nucléaire Rrp6 guidé par le complexe cofacteur de l'exosome, le complexe TRAMP.

1.2.4.1.1. Le complexe Nrd1/Nab3/Sen1

Nous allons donc nous intéresser au complexe qui assure la terminaison des snoRNAs, le complexe NNS. Cette section est la parfaite transition pour introduire les lncRNAs, puisque c'est ce même complexe qui assure leurs terminaisons.

Le complexe NNS est composé de ces 3 protéines identifiées par différentes études (Figure 16). Au début des années 1990, l'étude des hnRNPs (*heterogenous nuclear ribonucleoprotein particles*), des particules nucléaires composées de protéines et d'ARNs, a été menée par diverses approches génétiques (Wilson et al., 1994).

C'est dans le cadre de ces études prospectives que Nab3 a été découverte. En effet, en utilisant la levure pour identifier biochimiquement des protéines nucléaires se liant à l'ARN polyadénylé, Nab3 fut l'une des nombreuses protéines isolées du noyau à répondre aux critères de recherche. Très rapidement le domaine RRM (motif de reconnaissance de l'ARN) de Nab3 a été identifié (Lunde et al., 2011). Accolé à ce domaine, la protéine présente un domaine N-terminal de faible complexité riche en acides aspartique et glutamique. À l'extrémité carboxyterminale de la protéine se trouve une région riche en proline et glutamine (Figure 16). Ce dernier domaine est vital pour la levure (Carroll et al., 2007).



Figure 16 Répartition des domaines des protéines qui composent le complexe NNS. Schéma descriptif des domaines retrouvés dans les trois protéines du complexe NNS. En vert, les « motifs de reconnaissance de l'ARN », en orange les domaines riches en Proline et Glutamine, en jaune le domaine de Nab3 riche en Aspartate et Glutamate, en rose le domaine de Nrd1 responsable de l'interaction avec le domaine CTD de polIII, en bleu foncé les domaines d'interactions mutuels de Nab3 et Nrd1 et en bleu clair le domaine hélicase ATP dépendant de Sen1 (Arndt et al.,).

Un autre domaine important de Nab3 est celui qui lui permet d'interagir avec une autre protéine hnRNP, Nrd1 (Carroll et al., 2007). Différents tests ont permis de montrer que Nab3 et Nrd1 forment un hétérodimère stable et également que le RRM de Nab3 est important pour

la viabilité cellulaire (Conrad et al., 2000). La structure du RRM de Nab3 a posé les bases moléculaires permettant de suggérer le motif ARN de fixation de la protéine (Hobor et al., 2011; Porrua et al., 2012). Ce motif ARN a été initialement décrit comme UCUU (Carroll et al., 2007), mais des études récentes utilisant la technique de pontage des complexes ARN-protéine aux UV (PAR-CLIP) ont étendues ce motif (Webb et al., 2014b).

L'hypothèse de la dynamique du recrutement de Nab3 et Nrd1 pour la terminaison serait que les deux protéines ont une affinité moyenne pour leurs séquences cibles respectives, mais que plus le transcrit naissant est long, plus la possibilité que plusieurs dimères Nab3/Nrd1 se fixent et se stabilisent sur l'ARN est forte (Carroll et al., 2007; Loya et al., 2013; Webb et al., 2014a). Nrd1, pouvant se lier à polII, le réseau d'interaction entre Nab3, Nrd1, polII et le transcrit naissant forment un ensemble possédant une fonction biologique (Conrad et al., 2000). Plusieurs protéines Nab3 recombinantes purifiées peuvent, *in vitro*, se polymériser par leur domaine C-terminal riche en prolines et en glutamines, et de ce fait former une transition de phase (ou hydrogel) (Loya et al., 2013; O'Rourke et al., 2015), qui peut potentiellement être un aspect important du rôle de terminaison de Nab3 *in vivo* (Loya et al., 2017; Loya and Reines, 2016; O'Rourke and Reines, 2016).

Lors d'un crible génétique pour identifier des protéines de terminaison de la transcription, plusieurs protéines connues comme étant des « nuclear pre-mRNA downregulation (NRD) proteins » (Steinmetz and Brow, 1998) ont été identifiées. L'une d'elles, Nrd1, est également considérée comme une protéine hnRNP. En effet elle possède un domaine RRM et est capable de se lier à l'ARN. Nrd1 possède également un domaine de liaison au CTD de polII (CID, pour « CTD Interacting Domain), et se fixe sur ce domaine lorsque les Ser5 y sont phosphorylées (Vasiljeva et al., 2008). Cette spécificité de fixation de Nrd1 sur le CTD de polII en début de transcription (marquée par la phospho-Ser5) est une des bases de la sélectivité du complexe NNS pour la terminaison des ARNs plus courts, par exemple les snoRNAs (Gudipati et al., 2008). Nrd1 a la capacité de se fixer sur les séquences terminatrices des snoRNAs via son domaine RRM (Carroll et al., 2004), et comme expliqué précédemment, lorsque Nab3 et Nrd1 interagissent ensemble sur l'ARN, la force d'interaction est amplifiée. La protéine Nrd1 a été identifiée comme étant un terminateur pour plusieurs classe d'ARNs non codants différents, incluant les snoRNAs, les snRNAs et les CUTs (Arigo et al., 2006; Thiebaut et al., 2006). Plusieurs observations de la fixation de Nrd1 et Nab3 sur des ARNs codants ont

été faites chez la levure et ont soulevé l'hypothèse que ces facteurs pourraient participer activement à la biogenèse des pré-mRNAs (Creamer et al., 2011; Jamonnak et al., 2011; Webb et al., 2014a). Une autre hypothèse, pourrait être que ces facteurs de terminaison de la transcription des transcrits non codants participent au contrôle qualité des unités transcriptionnelles codantes en se liant à ses dernières en cas de problème(s) lors de leur biogenèse en mRNPs (Honorine et al., 2011).

La dernière sous-unité du complexe, Sen1, est la moins bien caractérisée des 3 protéines du NNS.

Le gène codant pour l'hélicase Sen1, a été initialement identifié lors d'un crible de perte d'activité endonucléasique d'épissage de l'ARN de transfert (Winey and Culbertson, 1988). Une des particularités de la protéine Sen1 est sa taille de 253 kDa, qui en fait une protéine difficile à étudier. De plus, cette protéine est présente en petite quantité vis à vis de ses partenaires d'interaction, Nab3 et Nrd1 (environ 498 copies/cellule versus 1723 et 1916 molécules par cellule, respectivement (Kulak et al., 2014). Cette dernière observation soulève la question de l'implication de Sen1 dans le complexe NNS. En effet, les proportions des partenaires du complexe n'étant pas stœchiométriques, quel peut être le niveau de saturation du complexe en Sen1 lorsque le NNS entre en action ? Plusieurs études ont montré l'implication de Sen1, en compagnie du reste du NNS, dans la terminaison de la transcription de petits ARNs non codants (Steinmetz et al., 2001). Un obstacle dans la compréhension des mécanismes d'action de Sen1, reste sa taille conséquente. À l'instant où ces lignes sont rédigées, aucune expérience n'a pu confirmer que le domaine identifié comme possédant l'activité hélicase, fonctionne comme telle dans la protéine entière, bien que ce dernier soit fonctionnel lorsqu'il est purifié seul (Porrua and Libri, 2013).

Le complexe NNS fonctionne donc en terminaison de la transcription des petits ARNs non codants, or nous avons vu précédemment que cette étape est, chez la levure, intimement liée à la maturation de l'extrémité 3' de l'ARN. En ce qui concerne les lncRNAs, cette étape de terminaison est liée à la dégradation de ce produit instable (e.g. CUTs décrits au sous-chapitre suivant). Le recrutement du complexe NNS passe donc par un réseau d'interactions multiples autour de la polIII et du transcrit naissant. Sans prendre en compte un ordre d'évènements particuliers, voici un résumé de ce réseau : a/ Nab3 et Nrd1 reconnaissent leurs motifs de recrutement sur l'ARN en cours de synthèse (UCUU pour Nab3 et GUAA pour Nrd1), b/ Nrd1

reconnait l'état Ser5 phosphorylé sur le CTD de polII et c/ Nrd1 et Nab3 peuvent s'associer à d'autres dimer de Nrd1/Nab3 pour former une plateforme ribonucléoprotéique supposée recruter Sen1, qui va alors pouvoir détacher l'ARN de la bulle de transcription (Carroll et al., 2007). Le recrutement de Sen1 pour la terminaison de la transcription est un mécanisme qui reste à éclaircir. En effet, bien que l'on retrouve Sen1 associé avec Nab3 et Nrd1 à l'extrémité 5' des transcrits, et donc également associé au CTD de polII phosphorylé sur ses Ser5, on le trouve également, de manière indépendante du reste du NNS, associé au CTD de polII phosphorylé sur la Ser2 (Chinchilla *et al.*, 2012). Une explication possible, est que Sen1 se charge sur polII via son interaction avec Nab3 et Nrd1 en début de transcription, puis va se lier indépendamment au CTD lorsque ce dernier se fait phosphoryler sur la Ser2 après le départ de Nab3 et Nrd1 (Chinchilla *et al.*, 2012). Cette hypothèse pourrait expliquer l'association de Sen1 à l'extrémité 3' de certains gènes codants sans pour autant déterminer son rôle exact à cette localisation (Arndt et al., 2015).

Le recrutement de plusieurs dimères de Nab3/Nrd1 au site de terminaison de la transcription, où certaines interactions protéine/ARN sont spécifique de la séquence et d'autres non, pourrait expliquer la distribution variable du nombre de séquences spécifique de recrutement ainsi que l'espacement variable entre ces motifs parmi les ARNs cibles (Carroll et al., 2007; Carroll et al., 2004). Ce multi-recrutement de dimères permet également de comprendre la « souplesse » de reconnaissance de sites des RRM de chacune des protéines ainsi que la nécessité de coopération des deux partenaires pour la reconnaissance d'un ARN (Bacikova et al., 2014). Le complexe NNS a été montré capable de se fixer sur les transcrits codants aberrants générés par l'utilisation du facteur bactérien Rho par l'équipe du Dr Rahmouni en 2011 (Honorine et al., 2011). Il y a donc toute une série de combinaisons possibles qui impliquent différentes fonctions de Nab3 et Nrd1 et qui impliquent une certaine plasticité dans la reconnaissance de l'ARN par ces protéines. Il existe possiblement, une compétition pour la terminaison de la transcription entre ce complexe et les complexes CF et CPSF (Porrua and Libri, 2015; Schaughency et al., 2014).

Deux autres complexes protéiques vont prendre en charge et assurer la maturation/dégradation des ARNs « marqués » par Nab3/Nrd1 : Le complexe TRAMP et l'exosome qui seront détaillés plus tard dans ce manuscrit.

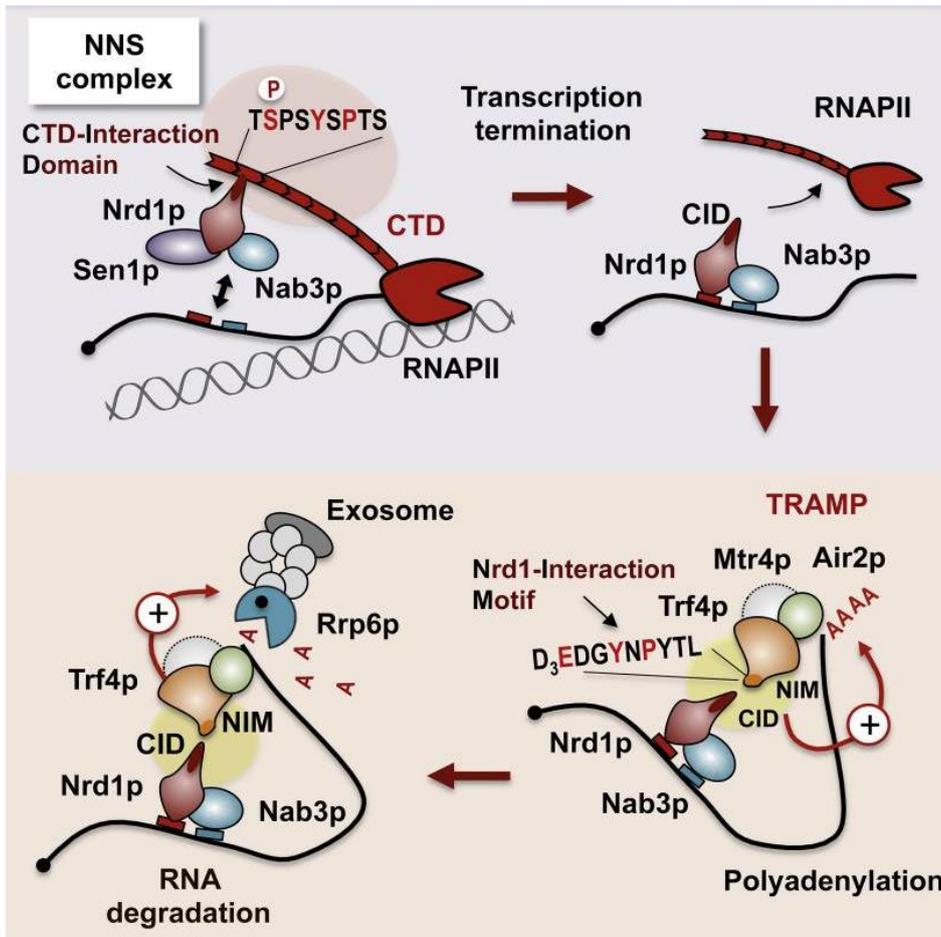


Figure 17 Mode d'action schématisé du complexe NNS. Recrutement en plusieurs étapes clés : reconnaissance de la phospho-ser5 du CTD, fixation sur les motifs spécifiques pour Nab3 et Nrd1, terminaison, liens avec la dégradation dans le cas (ici illustré) d'un lncRNA (Tudek et al., 2014).

1.2.4.2. Les longs ARNs non codants chez la levure

Au cours de cette section, nous allons nous intéresser aux longs ARNs non codants (lncRNAs) qui sont transcrits comme les ARNs messagers et les snoRNAs par polIII. L'exemple d'une classe appelé CUTs (pour Cryptic Unstable Transcripts) sera utilisé comme support à cette présentation. Il existe de nombreuses classes d'ARNs non codants (référéncé dans le tableau Figure 11) ayant des particularités, soit de transcription, soit de sensibilité à la dégradation qui leurs sont propres et qui ont défini leur appellation. Ces dernières ont fait leur apparition dans la littérature entre 2003 et 2005 comme étant des unités transcriptionnelles issues de polIII et dégradés quasiment tout de suite après transcription (Davis and Ares, 2006; Kaplan et al., 2003; Wyers et al., 2005).

Les CUTs ont été mieux caractérisés en 2009 (Neil et al., 2009; Xu et al., 2009) via des techniques de RNA-seq brin-spécifique, donnant une information sur le brin ADN d'origine des ARNs et apportant, de ce fait, une information supplémentaire lorsque deux ARNs sont portés par la même région génomique mais dans deux orientations différentes. De plus, cette nouvelle classe d'ARNs non codants a été révélée dans des conditions de déplétion de l'exonucléase 3'-5' nucléaire Rrp6. Cette dernière information nous donne la raison pour laquelle ces ARNs étaient demeurés invisibles jusqu'ici. En effet il semblerait que ces derniers soient hautement régulés, et dégradés rapidement après leur production.

1.2.4.2.1. Rôles et hypothèses de fonctions des lncRNAs.

Bien qu'intensivement étudiés, les lncRNAs sont assez mal caractérisés, que ce soit du point de vue de leur faible durée de vie ou de leurs fonctions biologiques. Cependant certaines hypothèses existent. Une d'elles serait que la transcription même des lncRNAs pourrait induire ou réprimer l'expression du gène adjacent (Wu et al., 2012) en induisant un remodelage de la chromatine proche (Geisler and Coller, 2013). Une étude récente portant sur l'effet de la transcription de 180 transcrits stables mais non traduits (SUTs), antisens et chevauchant des gènes codant des protéines, n'a pas montré de relation directe entre la transcription des SUTs antisens et le niveau d'expression de gènes codants chevauchants (Huber et al., 2016). Une autre étude visant à identifier des interactions génétiques entre 6 SUTs inter-géniques délétés dans des souches dépourvues de 6 protéines non essentielles a mis en évidence un rôle potentiel de 2 SUTs dans des fonctions cellulaires spécifiques et a également apporté la preuve

que ces derniers pouvaient agir en trans (Kyriakou et al., 2016). De nombreux lncRNAs peuvent se lier à des protéines de liaison à l'ARN différentes de celles se liant aux ARNm, ce qui pourrait leur conférer certaines fonctions spécifiques (Tuck and Tollervey, 2013).

Il est également intéressant de relever un autre phénomène autour de ces lncRNAs, plus lié à l'acte de transcription pervasive qu'à la molécule ribonucléique qui en découle : l'interférence transcriptionnelle. Cette dernière se caractérise par le fait que la transcription d'un lncRNA donné va interférer avec la transcription du mRNA voisin (Nevers *et al.*, 2018). Ce phénomène peut être observé pour deux unités transcriptionnelles issues d'un promoteur bidirectionnel, mais aussi lorsque l'initiation de la transcription du lncRNA s'effectue aux alentours de l'extrémité 3' du gène codant voisin (Muray *et al.*, 2012). Chez la levure, le complexe NNS prévient d'une telle interférence en terminant prématurément la transcription des unités non codantes, empêchant ainsi tout conflit au niveau du promoteur du gène codant voisin (Schulz *et al.*, 2013 ; Castelnuevo et al., 2014).

La plupart de ces études ne se sont intéressées qu'à une fraction de ces nombreux lncRNAs. Des études à l'échelle transcriptomique chez la levure permettraient peut-être de définir la fonction de cette masse fonctionnellement non caractérisée de lncRNAs.

1.2.4.2.2. Régulation particulière de ces lncRNAs

En plus d'avoir une origine génomique originale, ces ARNs non codant pouvant être synthétisés depuis des régions intergéniques, des introns, des activateurs ou encore venir du même promoteur que les gènes codants mais dans une direction opposée, les lncRNAs sont soumis à une régulation stricte. Celle-ci est d'autant plus stricte lorsqu'il s'agit des CUTs dont nous allons parler par la suite. En règle générale, les régions non codantes divergent rapidement au cours de l'évolution. Ce phénomène représente un désavantage certain pour une reconnaissance efficace par le système de QC basée sur des protéines hautement conservées, qui logiquement seraient plus affines à réguler des séquences également conservées (motifs).

Les CUTs ont été identifiés de prime abord dans des souches dépourvues de l'exonucléase 3'-5' Rrp6 de l'exosome ou de la sous-unité Trf4 du complexe TRAMP qui est lui-

même un cofacteur de l'exosome. Ils semblent donc être en permanence sous la surveillance d'un système qui les conduit fatalement à leur dégradation (Wyers et al., 2005; Xu et al., 2009).

Comme indiqué précédemment, les promoteurs eucaryotes sont intrinsèquement bidirectionnels (Andersson et al., 2014). Au niveau de nombre d'entre eux, polII initie la transcription de manière équivalente dans les deux directions, mais les transcrits « en amont » sont rapidement éliminés par la machinerie nucléaire de surveillance. Que ce soit chez la levure, ou chez l'humain, cette surveillance est intimement liée au mécanisme de terminaison précoce de la transcription. Chez *S. cerevisiae*, la transcription des ARNs non codants est terminée par le système NNS (présenté plus haut dans ce manuscrit dans la terminaison de la transcription des snoRNAs), et le transcrit résultant de cette terminaison est pris en charge par le complexe TRAMP et l'exosome pour être dégradé. Les motifs de fixation des protéines Nrd1 et Nab3 ont tendance à être plus représentés dans les transcrits issus d'une transcription en amont des promoteurs bidirectionnels, appelés NUTs, et déplétés dans les séquences des transcrits codants pour des protéines (Schulz et al., 2013). Cette asymétrie semble favoriser la terminaison NNS-dépendante dans les régions en amont des promoteurs, mais ne suffit pas à expliquer la différence de stabilité entre les CUTs/NUTs et les mRNAs. Qui plus est, les motifs de fixation de Nab3 et Nrd1 sont de faible complexité et généralement aussi abondants dans les gènes codants des protéines. Pour parfaire l'ambiguïté, des expériences de pontage ARN-protéine (4tU-seq) montrent que Nrd1 et Nab3 se lient à presque tous les transcrits polII dépendants (Schulz et al., 2013).

Un autre déterminant de cette terminaison/dégradation particulière, est l'état de phosphorylation du CTD de polII. Les stades précoces de la transcription sont marqués par un taux élevé de Ser5 phosphorylées et un faible niveau de phosphorylation des Tyr1, Ser2 et Ser7 du CTD. Cette combinaison, comme expliqué précédemment, est une des conditions favorisant le recrutement du complexe NNS sur les transcrits naissants. Pour la plupart des gènes, polII, qui est en début de phase d'élongation, va marquer une pause aux alentours des 150 premiers nucléotides transcrits, correspondant grossièrement à la position du premier nucléosome. Au fil de la transcription, passé les 150-200 premiers nucléotides transcrits, le code du CTD est graduellement reconfiguré en mode « élongation ». Plus polII avance sur le gène plus la marque de phosphorylation de la Ser5 s'amenuise, rendant de ce fait les transcrits moins sensibles à un recrutement du NNS et à la terminaison qui en découle. En parallèle,

l'augmentation croissante de la phosphorylation des Ser2 facilite le recrutement du complexe canonique de clivage et polyadénylation, CF-CPF. De manière surprenante, il semblerait que ce changement de marque de phosphorylation échoue pendant la transcription des CUTs (Milligan et al., 2016), ce qui laisserait une fenêtre d'action plus grande pour la mise en place du complexe NNS et la terminaison prématurée sous-jacente.

Additionnellement, cet état inchangé de phosphorylation Ser5-P inhibe le recrutement des facteurs qui pourraient faciliter la terminaison au niveau de signaux de polyadénylation cryptique et conduire à la formation d'un ARN stable. Ce faisceau d'informations nous indique que les CUTs sont en quelque sorte destinés à être dégradés par un système passant par une terminaison par le NNS, suivi d'une prise en charge par le TRAMP aboutissant à la dégradation par l'exosome et Rrp6.

L'exosome est un acteur important dans la biogenèse et la surveillance de toutes les classes d'ARNs. Le chapitre suivant s'attelle à présenter ce complexe majeur dans la « vie » des ARNs.

1.3. Contrôle qualité et dégradation

Toutes les étapes de la biogenèse des mRNPs sont soumises à la surveillance stricte de plusieurs acteurs appartenant au système de contrôle qualité (QC) de l'ARN. Tout évènement aberrant, que ce soit d'emballage ou encore de maturation, sera détecté par ce QC qui empêchera d'abord la molécule aberrante d'être exportée vers le cytoplasme et contribuera à sa dégradation. Cette ultime étape de dégradation est assurée par un complexe protéique hautement conservé au cours de l'évolution, appelé exosome. Cette dégradation par l'exosome ne serait pas possible sans le complexe TRAMP, cofacteur qui va préparer les ARNs pour qu'ils soient dégradés par les sous-unités catalytiques de la machinerie de dégradation exosomique. Mais le rôle de l'exosome ne s'arrête pas à la dégradation des produits défectueux de transcription. En effet, cette machinerie est aussi nécessaire à la maturation d'une multitude d'ARNs non codants et est responsable en partie de la courte durée de vie des lncRNAs.

1.3.1. Machinerie canonique de dégradation : l'exosome

La régulation de la maturation des ARNs et la dégradation de ces derniers sont une étape clé de l'expression génique. L'ARN exosome nucléaire est un acteur central dans la surveillance de toutes les classes de transcrits issus de la transcription de toutes les ARN polymérase (Figure 18). L'exosome porte la responsabilité de la maturation de l'extrémité 3' de certaines classes d'ARNs (ARNr, ARNt, sn/snoRNAs) en catalysant non seulement la maturation de l'extrémité 3' de ces classes, mais en dégradant également les transcrits présentant une extrémité 3' incomplète (Kilchert et al., 2016; Mitchell, 2014; Mitchell et al., 1997; Zinder and Lima, 2017) (Figure 19). L'exosome possède également d'autres fonctions dans sa forme cytoplasmique (Sloan et al., 2012). Ces autres fonctions ne seront pas décrites dans ce manuscrit.

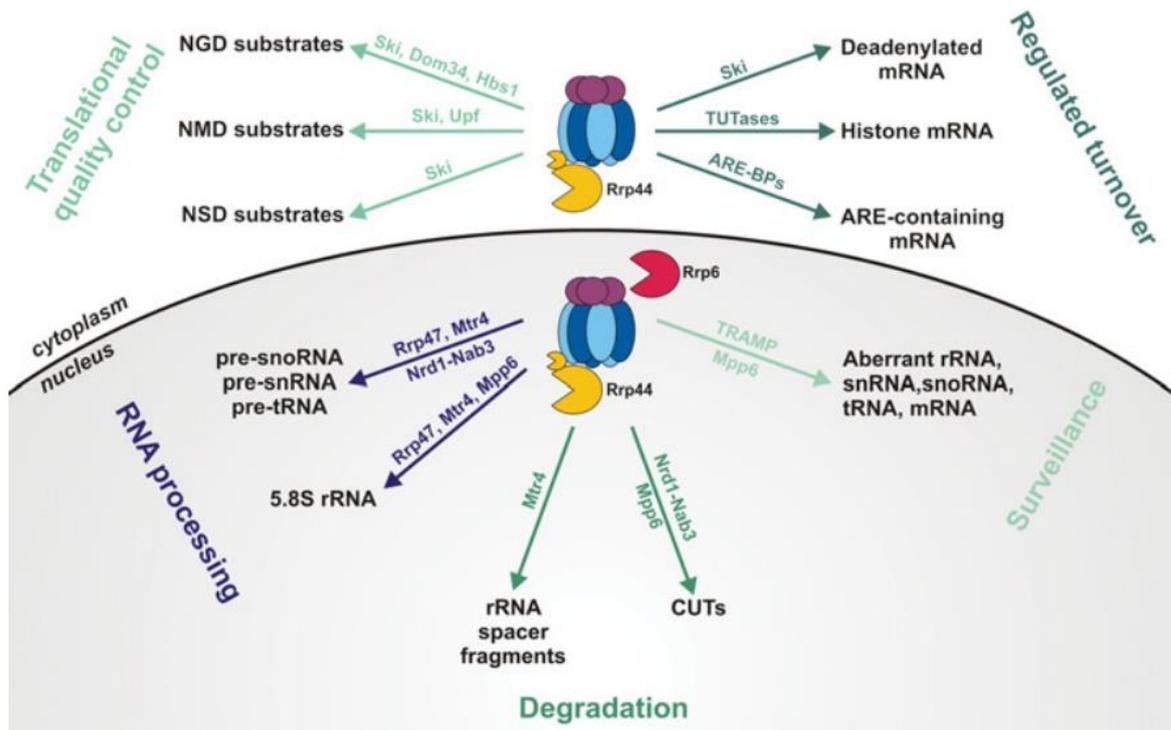


Figure 18 Vue d'ensemble non exhaustive des substrats de l'exosome chez les eucaryotes (Sloan et al., 2012). Schéma présentant les cibles de l'exosome (forme nucléaire en bas de figure et cytoplasmique en partie haute) ainsi que les cofacteurs impliqués (sur les flèches).

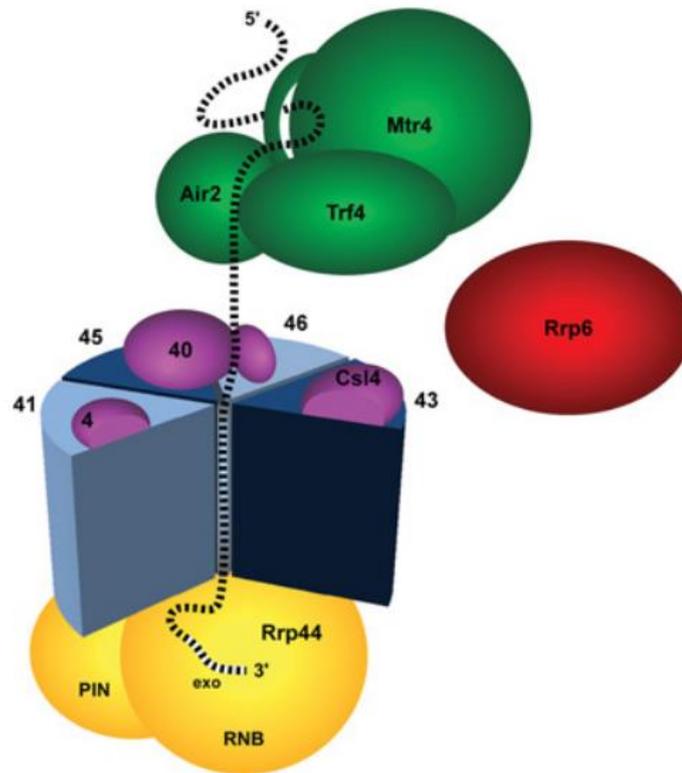


Figure 19 Schéma de la dégradation des ARNs médiés par le complexe TRAMP et vue d'ensemble de la structure de l'exosome (Sloan et al., 2012)..

1.3.1.2. Structure et fonctionnement de l'exosome

L'exosome possède une structure particulière. Cette machinerie est une super-structure comprenant un complexe cœur de 9 protéines, catalytiquement inactif et 2 sous-unités catalytiques Rrp6 et Dis3 (Rrp44) (Mitchell et al., 1997). Le complexe cœur a une forme de baril composé de deux couches différentes. La première couche comprend six protéines à domaine PH (domaine homologue à la pleckstrine) : Rrp41-42-43-45-46 et Mtr3 et va former un anneau (Liu et al., 2006). La deuxième couche comprend trois protéines de « coiffe » qui possèdent des domaines KH/S1 et vont venir stabiliser l'hexamère de protéine à domaines PH. Les deux sous-unités catalytiques vont venir se positionner à chaque extrémité du baril pour former une super-structure à 11 sous-unités, avec Rrp6 au-dessus de la coiffe trimerique de protéines KH/S1, près du pore d'entrée du baril et Dis3 va se placer dessous le baril près du pore de sortie (Figure 19) (Ogami et al., 2018). Rrp6 possède une activité 3'-5' exonucléase alors que Dis3 présente une activité endonucléase en plus de son activité 3'-5' exonucléase. Les deux nucléases possèdent de nombreuses cibles en communs (ARNm, CUTs, SUTs, ARNt, ARNr), indiquant une certaine redondance du système de dégradation. De plus, les activités exo- et endonucléases de Dis3 semblent fonctionner de concert pour la dégradation des ARNs cibles. Enfin l'activité endonucléolytique de Dis3, dans certains cas, fournit un point d'entrée pour Rrp6 en « créant » une extrémité 3' non protégée (Schneider *et al.*, 2012).

Pour que l'activité nucléaire de l'exosome soit maximale, trois cofacteurs supplémentaires sont requis : Mtr4, Rrp47 ainsi que Mpp6. Rrp47 interagit avec Rrp6 et fournit une plateforme pour recruter Mtr4, une hélicase DExH-box essentielle. Mpp6, quant à elle, va se lier à la sous-unité inerte Rrp40 et renforcer l'activité hélicase de Mtr4. Cette dernière activité est responsable de la résolution des structures secondaires de l'ARN à dégrader afin que ce dernier se retrouve dans une forme simple brin pour son entrée dans le canal formé au cœur du baril de l'exosome dans une orientation 3'-5' (Wasmuth *et al.*, 2017 ; Falk *et al.*, 2017).

Dis3 va dégrader les ARNs qui traversent entièrement le canal du baril de l'exosome alors que Rrp6 dégrade, plus ou moins (en fonction de l'action requise : maturation ou élimination du substrat), les ARNs qui entrent dans la coiffe formée par les trois protéines à domaine S1/KH, et ressortent directement pour rejoindre l'exonucléase (Figure 20).

Il semblerait que l'emprunt du canal central du baril de l'exosome par les ARNs se fasse principalement dans le cytoplasme des cellules. Ce phénomène est en partie expliqué par le changement allostérique résultant de la fixation de Rrp6 au complexe lorsque ce dernier est nucléaire (Wasmuth et al., 2014). La conséquence étant l'inhibition de la dégradation des substrats cytoplasmiques uniquement en présence de Rrp6 (Delan-Forino et al., 2017).

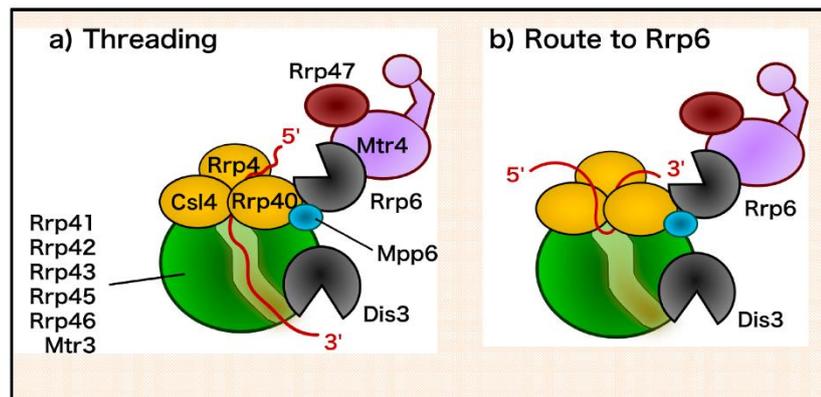


Figure 20 Chemin emprunté par l'ARN substrat à travers l'exosome (Ogami et al., 2018).

1.3.1.2. Cofacteurs de l'exosome

L'exosome a donc la capacité de dégrader une large variété de substrats ARNs. Reste à savoir comment cette machinerie de dégradation fait pour cibler ses substrats. Il existe une multitude de complexes adaptateurs qui vont guider l'exosome pour que ce dernier se charge sur les ARNs sélectionnés.

Les différentes composantes protéiques des adaptateurs de l'exosome sont bien conservées à travers l'évolution de la levure à l'homme. Il est important de noter que la protéine Mtr4 est contenue dans la plupart des complexes adaptateurs, révélant de ce fait le rôle central et primordial de ce facteur dans la formation de ces complexes et leurs fonctions.

1.3.1.2.1. Le complexe TRAMP

Le complexe Mtr4-Trf4/5-Air1/2 de polyadénylation ou TRAMP est, à ce jour, le complexe le mieux caractérisé parmi les différents cofacteurs de l'exosome nucléaire chez *S. cerevisiae*. Un bref historique des découvertes autour du complexe TRAMP nous montre que c'est au travers du contrôle qualité des tRNAs, et plus précisément de l'importance de la polyadénylation du tRNA Met par Trf4 (Kadaba et al., 2004) que cette sous-unité a commencé à être étudiée plus en détail. Peu de temps après, la composition entière du complexe responsable de cette polyadénylation a été élucidée (LaCava et al., 2005; Vanacova et al., 2005) et les protéines Mtr4 et Air1/2 ont été ajoutées à la liste des composantes du complexe.

Air1/2 fournit au complexe la capacité de se lier à l'ARN et est également crucial pour la stabilité du complexe (Fasken et al., 2011) et Mtr4 joue un rôle dans le déroulement de l'ARN lorsque ce dernier présente des structures secondaires et également un rôle dans la modulation de la taille de la queue polyA de ce même ARN substrat (Jia *et al.*, 2011). Le complexe TRAMP est capable de reconnaître un large panel de transcrits, comme les ARNt, les ARNr, les sn/snoRNAs, les ARNs télomériques, les longs ARNs non codants (CUTs) et les pré-ARNm. Ces différentes classes d'ARNs substrats sont tous polyadénylés par Trf4/5.

Bien que le complexe TRAMP présente une capacité intrinsèque de liaison à l'ARN, son efficacité de recrutement aux ARNs substrats est renforcée par le complexe NNS présenté plus tôt dans ce manuscrit. Le recrutement du complexe sur le substrat via le NNS, passe par le domaine CID de Nrd1 qui va être capable d'interagir avec un domaine de Trf4 mimant le CTD de polIII (Tudek et al., 2014). Bien que les interactions Nrd1/Trf4 et Nrd1/CTD de polIII soient mutuellement exclusives, la terminaison de la transcription par le complexe NNS et la dégradation de l'ARN par l'exosome stimulée par le complexe TRAMP sont coordonnées. Il a également été montré que Nab3 pouvait faciliter la fonction du complexe TRAMP indépendamment de Nrd1 (Fasken et al., 2015).

Il est intéressant de noter que la polyadénylation apposée par le complexe TRAMP sert de guide à l'exosome (Jia et al., 2011), et constitue donc une amorce de dégradation. Le rôle de cette « queue » polyA est donc aux antipodes de la queue polyA ajoutée par le complexe de polyadénylation canonique Pap1 en 3' des ARNm qui sert au contraire à la stabilité du mRNA.

1.3.1.2.2. Autres cofacteurs de l'exosome

Nous avons présenté le complexe cofacteur de l'exosome central à de nombreux évènements de dégradation. Comme dit plus tôt, il existe de nombreux cofacteurs de l'exosome, nous allons en citer quelques-uns et brièvement voir leurs fonctionnements.

La protéine Rrp47 est l'un des cofacteurs les plus intimement lié à l'exosome. En effet, Rrp47 semble être indispensable à la maturation par l'exosome des ARNr et des sn/snoRNAs et participerait activement au ciblage de ces substrats (Mitchell et al., 2003). Qui plus est, la stabilité de la protéine Rrp6 semble dépendre de la présence de Rrp47 (Stead et al., 2007; Stuparevic et al., 2013).

Le complexe NNS a déjà été abordé au cours d'un précédent chapitre. Il est l'un des cofacteurs important de l'exosome nucléaire. C'est via la fixation de Nab3 et de Nrd1 à l'ARN que le complexe de dégradation peut être recruté. De même la protéine Mpp6, se fixe sur des séquences d'ARNs riche en pyrimidine et va faciliter la maturation de l'ARN 5.8S en promouvant le recrutement de l'exosome. Une étude a montré l'implication de cette protéine dans la surveillance des pré-ARNr et pré-ARNm ainsi que dans la dégradation des CUTs (Milligan et al., 2008).

En plus de ces cofacteurs séquences dépendants, il existe des partenaires de l'exosome qui vont être structure dépendant. L'endonucléase Rnt1 en est un bon exemple, puisqu'elle va catalyser le clivage de structures ARNs en tige-boucle, et ainsi fournir un point d'entrée à l'exosome sur ces cibles. Ce procédé a été montré au niveau d'ARNr et de sn/snoRNA par plusieurs équipes (Abou Elela and Ares, 1998; Chanfreau et al., 1998; Egecioglu et al., 2006; Vasiljeva and Buratowski, 2006).

Nous avons donc pu voir que le système de contrôle qualité et dégradation est en quelque sorte, en compétition cinétique avec les évènements de biogenèse. Une erreur se produisant dans une étape de maturation de l'ARN naissant en mRNP fonctionnelle sera détectée par le système de contrôle qualité qui pourra alors retenir ce transcrit malformé dans le noyau afin de le dégrader (Schmid and Jensen, 2008).

La suite de ce manuscrit va décrire par quels moyens, nous parvenons dans l'équipe à étendre cette courte fenêtre d'observation d'évènements aberrants. En effet, la compétition cinétique entre la biogénèse de la mRNP et la détection par le système de contrôle qualité d'erreurs potentielles se réalise dans un temps très court en plus d'être des évènements isolés dans le noyau des cellules. La section suivante va donc présenter l'outil qui nous permet d'étudier aisément les mécanismes de QC.

Dans ce manuscrit, seul l'aspect dégradation des mRNP aberrantes est développé, or d'autres mécanismes de contrôle qualité existent dans le noyau des cellules eucaryotes. Ils peuvent consister en la rétention du transcrit aberrant à son site de transcription ou encore sa rétention au niveau du nucléole et aussi au niveau du pore nucléaire. Ces mécanismes de contrôle qualité par rétention sont intimement liés aux protéines d'emballage et d'export qui se fixe à l'ARN en cours de transcription et également à l'exosome. D'après une étude récente de Paul & Montpetit (Paul and Montpetit, 2016), il est possible d'identifier trois procédés de rétention différents basés sur des observations phénotypiques de localisation des ARNs polyA dans différentes souches de levures mutantes. En croisant avec d'autres références bibliographiques, la première classe correspond aux ARNs immatures (pre-mRNAs) qui vont être retenus au niveau du pore nucléaire (Hackmann et al., 2014 ; Palancade et al., 2005) et est défini comme le contrôle qualité dépendant du NPC (« Nuclear Pore Complex »). La seconde classe a été identifiée dans des souches présentant des mutations pour les facteurs de maturation de l'extrémité 3' des ARNm. Lorsque de tels défauts apparaissent, les ARNm sont retenus au niveau de leur site de transcription dans un procédé semblant impliquer la protéine Ldb19, une alpha-arrestine initialement décrite dans d'autres procédés dans la cellule. La troisième classe correspond à un mécanisme de rétention près du nucléole. Ce phénotype de localisation est observé uniquement dans des mutants pour le complexe TRAMP et l'exosome et soulève la possibilité qu'une telle rétention d'ARNs aberrants n'est possible que lors d'une perte d'activité de dégradation de ces ARNs.

1.4. Disséquer les systèmes de contrôle qualité et dégradation nucléaires de l'ARN

Afin d'étudier l'implication des diverses composantes des machineries de QC et de dégradation décrites précédemment, l'équipe du Dr Rahmouni a développé un outil original. En partant de l'observation que les événements de biogenèse d'un ARNm aboutissant à une molécule ribonucléoprotéique aberrante sont rares, l'équipe a développé une approche permettant de créer une grande quantité de mRNPs aberrantes qui serviront de substrats pour les machineries de contrôle qualité et dégradation de l'ARN. De ce fait, en suivant les niveaux relatifs des ARNm perturbés dans différentes souches de levure dépourvues des protéines appartenant à ces machineries, il devient possible de déduire la participation des protéines en question dans les mécanismes de QC et/ou dégradation.

1.4.1. L'hélicase translocase bactérienne Rho

L'étape clé de l'approche employée au sein de l'équipe afin de perturber la biogenèse des mRNPs repose sur l'utilisation du facteur de terminaison de la transcription bactérienne Rho de manière hétérologue dans le noyau des levures. Afin de mieux comprendre ce choix de détourner un des systèmes de terminaison de la transcription chez les bactéries en un outil de perturbation de la biogenèse des mRNPs chez les eucaryotes, nous allons tout d'abord replacer ce facteur bactérien dans son contexte biologique en décrivant sa structure et sa fonction naturelle.

1.4.1.1. Structure et fonction

Le facteur Rho est une hélicase homohexamérique en forme d'anneau. Chacune des sous-unité est faite de 419 acides aminés et aucun homologue fonctionnel dans le règne eucaryote n'a, à ce jour, été découvert (Geiselman et al., 1992; Skordalakes and Berger, 2003, 2006).

La fonction principale de Rho est la terminaison de la transcription. Chez *E. Coli*, environ 50% des événements de transcription sont terminés de façon Rho-dépendante (D'Heygere et al., 2013). Une récente revue rapporte différents rôles, additionnellement à la terminaison de transcription, que pourraient posséder Rho, comme la régulation de l'expression génique, la maintenance de la stabilité du génome ou encore la prévention contre

la formation de R-loops (Boudvillain et al., 2013). La fonction principale de terminaison de la transcription peut être résumé comme suit : tout d'abord, le domaine N-terminal de Rho, appelé PBS pour « *Primary Binding Site* » va pouvoir se fixer aux ARNs au niveau de séquences ribonucléiques riches en cytidine ou sur sa séquence (déterminée in vitro) d'utilisation appelé rut pour « *Rho Utilization* ». Cette fixation est possible lorsque le ribosome se détache de l'ARN en fin de traduction (la transcription et la traduction étant couplés chez les bactéries) ou lorsqu'un codon non-sens est rencontré (Yu et al., 2000), et s'effectue lorsque Rho est dans sa conformation semi-ouverte. La seconde conformation apparaît lorsque l'homohexamère est fermé autour de l'ARN. L'ARN lié à Rho est ensuite positionné dans le canal central de la structure en anneau qui ne peut accueillir qu'une molécule d'ARN à la fois (Skordalakes and Berger, 2003). Cet ARN va alors pouvoir interagir avec le second site de liaison à l'ARN (SBS pour « *Secondary Binding Site* ») situé à l'extrémité C-terminal de chaque monomère, ce qui provoque le changement conformationnel en position fermé (Figure 21). Ce domaine C-terminal porte également l'activité ATPase, nécessaire à la translocation 5'-3' le long de l'ARN naissant (Thomsen and Berger, 2009). Le modèle de translocation supposé est une translocation à pas monotone d'un nucléotide d'avancé par molécule d'ATP hydrolysée (Soares et al., 2014). L'efficacité de terminaison de transcription est dépendante du différentiel de vitesse entre la transcription, assurée par l'ARN polymérase et la translocation sur le brin d'ARN par Rho. C'est le résultat du différentiel entre ces deux cinétiques qui permet de fournir un outil moléculaire à la bactérie pour moduler son taux de terminaison de la transcription, en jouant sur la vitesse de ces deux enzymes (Gocheva et al., 2015). Les mécanismes menant à la terminaison du transcrit restent mal compris, mais le contexte dans lequel il se déroule est lui assez claire : la polymérase marque une pause sur le site de terminaison de la transcription et le facteur Rho va rattraper le complexe de transcription et promouvoir sa dissociation de la matrice ADN (Ciampi, 2006) et la libération du transcrit. Plusieurs modèles ont été proposés pour expliquer ce phénomène. Le premier, soumet l'idée que l'ARN serait tiré de la matrice ADN par l'activité hélicase de Rho (Richardson, 2002) (Figure 22). Le second modèle propose un mécanisme par lequel Rho, via son activité translocase, entrerait en contact avec la polymérase et la « pousserait » de la matrice ADN (Park and Roberts, 2006). Le dernier modèle propose que Rho provoquerait un changement conformationnel allostérique du complexe d'élongation de la transcription qui induirait sa

dissociation et par voie de conséquence marquerait la terminaison de la transcription (Epshtein et al., 2010).

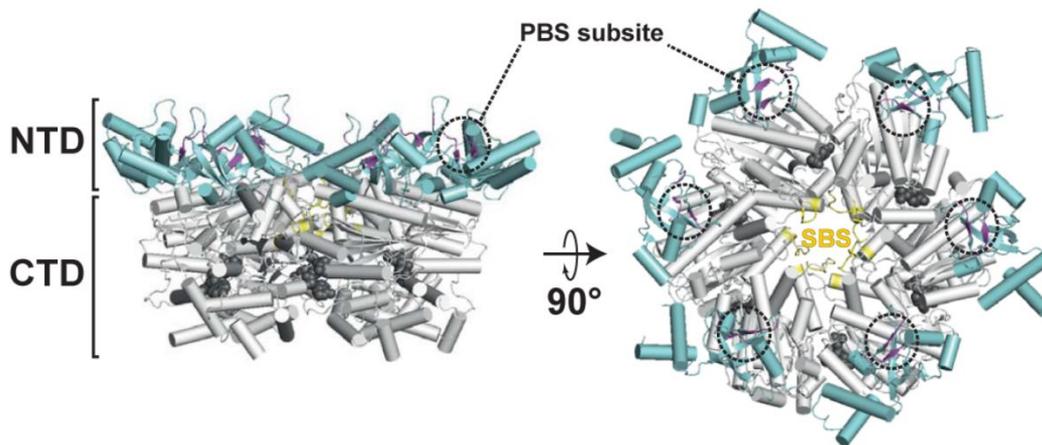


Figure 21 Structure de l'hexamère Rho (Soares et al., 2014). Les domaines N- et C-terminaux (NTD et CTD) sont représentés respectivement en cyan et en blanc. Les résidus du PBS et du SBS sont représentés en violet et jaune respectivement.

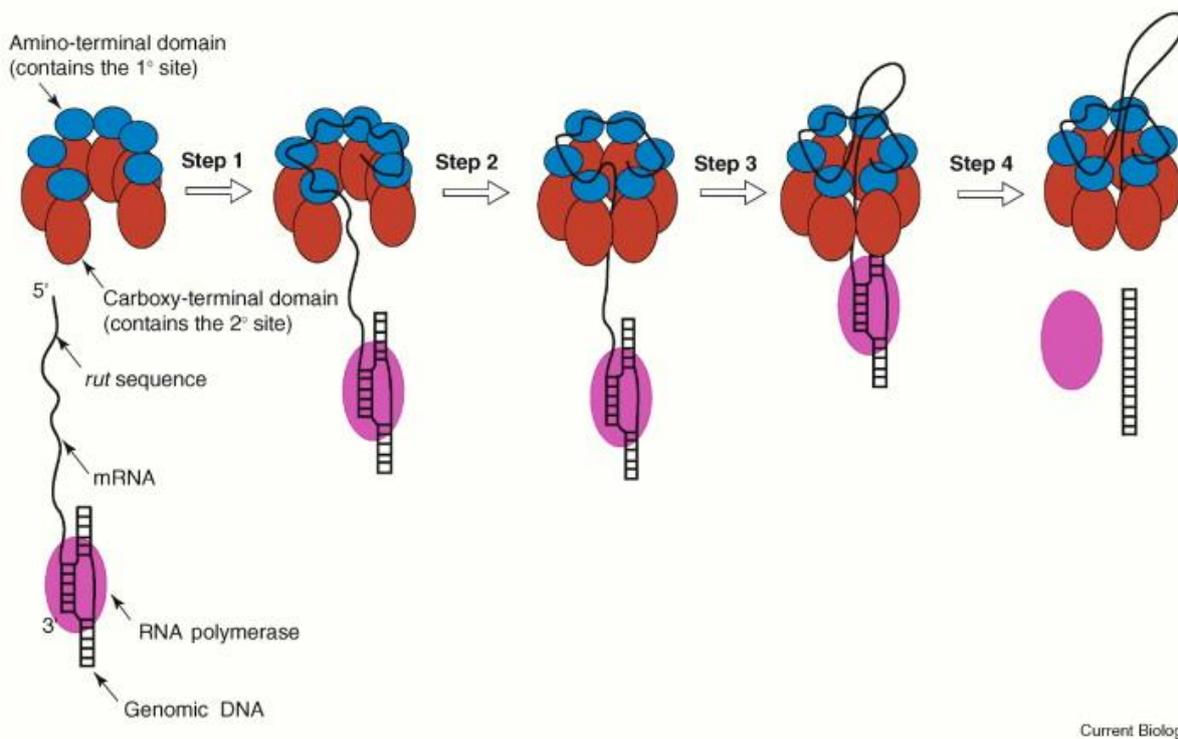


Figure 22 Terminaison de la transcription bactérienne Rho-dépendante (Kaplan and O'Donnell, 2003). 1/ Reconnaissance de la cible. 2/ liaison primaire via le PBS. 3/ liaison au SBS puis changement conformationnel. 4/ Translocation. 5/ Terminaison et dissociation.

1.4.1.2. Détournement du système bactérien pour amplifier la fenêtre d'observation d'évènements rares

Les propriétés intrinsèques du facteur Rho ont été vu par l'équipe du Dr Rahmouni comme étant une opportunité pour étudier le système de contrôle qualité en employant l'hexamère bactérienne comme outil de perturbation de la biogenèse des mRNPs. Fort d'un passif d'étude de la terminaison de la transcription procaryote, l'équipe possédait déjà un robuste fond de connaissance sur Rho et l'a détourné de son rôle canonique chez la bactérie pour l'utiliser dans l'organisme modèle qu'est la levure. Afin de l'utiliser dans la levure de manière contrôlée, la séquence codante de Rho a été placée sous le contrôle d'un promoteur TetO7, réprimé par la doxycycline et le C-terminal de la protéine a été fusionné à un signal de localisation nucléaire (NLS). Cette construction est placée dans un plasmide centromérique exprimant déjà de manière constitutive, le trans-activateur tTA qui contrôle l'expression à partir du promoteur TetO7 (Mosrin-Huaman et al., 2009).

Les différentes études, menées par l'équipe, ont permis de montrer que l'expression de Rho dans la levure *Saccharomyces cerevisiae* est toxique, et se traduit par un retard de croissance des cellules. Ce ralentissement de la croissance semble dépendant du niveau d'expression du facteur bactérien. En effet, lorsque la quantité de doxycycline dans le milieu diminue (donc la trans-activation du gène augmente) progressivement, le retard de croissance se montre de plus en plus important. Une partie de la toxicité observée phénotypiquement peut être expliqué d'un point de vue moléculaire, car certains transcrits de polIII voient leur niveau global réduit lorsque Rho est exprimée. La première hypothèse émise pour expliquer cette diminution du niveau global de certains transcrits fut l'apparition d'évènements de terminaisons précoce de la transcription sous l'effet de Rho, or aucun ARNm terminé prématurément n'a pu être observé. Dans la veine de cette hypothèse, l'impossibilité de détecter des transcrits prématurément terminés, et donc plus courts, a été incriminé à une possible dégradation précoce. L'expérience a donc été répétée dans une souche déplétée de la sous-unité catalytique Rrp6 de l'exosome. De manière surprenante, la souche mutante déplété de Rrp6 et en présence de Rho a vu le phénotype de retard de croissance annulé. Ce dernier résultat s'accompagne d'un retour à la normale des quelques transcrits analysés dont le niveau était diminué dans la souche sauvage en présence de Rho. Ces résultats, qui sont à la base de mes travaux de thèse, posent le constat qu'en présence de Rho, certains ARNm

transcrits dans leurs intégralités sont reconnus comme étant aberrants par le système de contrôle qualité et sont dégradé par l'exosome, ou du moins par une de ses sous-unité catalytique, Rrp6. Afin de faciliter le travail dans une souche dépourvue de Rrp6, l'équipe a montré qu'il était possible d'utiliser une souche dépourvue de Rrp47. En effet, comme présenté plus tôt, la stabilité de la protéine Rrp6 repose sur son interaction avec Rrp47 (Mitchell et al., 2003; Schuch et al., 2014; Stuparevic et al., 2013). La délétion d'une des deux protéines induit la déplétion de l'autre d'environ 90%. Sachant que la souche *rrp6Δ* est connue comme étant sensible à la température et possède également une croissance ralentit ainsi qu'une certaine instabilité génétique, nous avons choisi de nous placer dans une souche *rrp47Δ* afin de dépléter Rrp6.

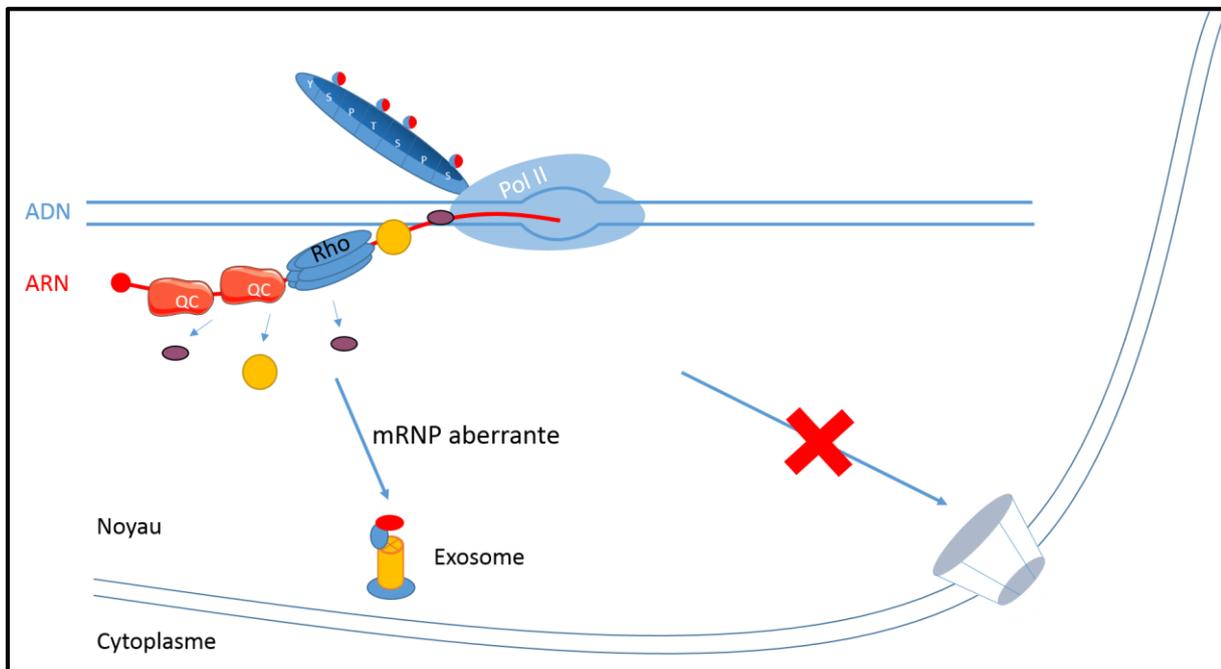


Figure 23 Schéma de la perturbation de la biogenèse des mRNPs par Rho et de la reconnaissance de l'ARN naissant par le système de contrôle qualité. Rho va chasser les protéines qui empaquètent l'ARN naissant, le rendant aberrant et de fait, détectable par les protéines du système de contrôle qualité (simplifiées ici par « QC »).

1.5 Problématique de travail et projets de thèse

Les procédés liés à la transcription, qu'ils soient impliqués dans la maturation de l'ARN naissant, dans sa protection, son empaquetage, son export ou dans le contrôle de la qualité de la particule ribonucléoprotéique, sont des événements compliqués. Ils mettent en scène différents complexes protéiques ayant eux-mêmes, intrinsèquement, le potentiel de se retrouver en contact direct, ou à travers des interactions avec le transcrit, ou encore la polII.

L'étude des mécanismes régissant le contrôle qualité des mRNPs est donc un challenge de part, les différents complexes protéiques impliqués dans ce procédé mais également de part, la faible fenêtre d'observation que le vivant nous accorde. Il faut donc redoubler d'ingéniosité pour proposer des outils capables de nous fournir une vision du comportement des composantes du QC et de leurs cibles. En ce sens, l'outil Rho développé par mon équipe d'accueil est l'une des bases fondamentales de mon travail de thèse.

Mon sujet d'étude se situe dans la continuité des travaux effectués par l'équipe. En partant des observations faites sur le système de contrôle qualité au travers de quelques transcrits codants, mon rôle a été d'étendre ces modèles à l'ensemble du transcriptome via des techniques de RNA-seq. En parallèle de cela, différentes expériences de CHIP-seq ont été menées dans le but de suivre le recrutement des composantes du système de contrôle qualité et dégradation des ARNs sur la chromatine. Différentes questions et hypothèses sont donc posées :

1/ Quels sont les ARNs impactés par notre système de perturbation ? Qu'arrive-t-il aux ARNs non codants (petits et grands) ?

2/ Quel est le devenir des ARNs impactés par Rho ? Peut-on observer un retour à la normale dans une souche dépourvue de Rrp6 ?

3/ Quels sont les co-facteurs de l'exosome et autres protéines impliquées dans la reconnaissance des transcrits aberrants ? Agissent-ils de manière co-transcriptionnelle dans notre modèle de perturbation ?

4/ Comment le système de contrôle qualité agit-il ? Est-il mobile sur le génome ? Peut-on suivre le recrutement de ces composantes au niveau des régions génomiques (ADN) correspondant aux loci d'ARNs perturbés ?

5/ Dans un contexte de stress intense de la biogenèse des mRNPs, les complexes connus comme étant impliqués dans leurs maturation et leurs empaquetages restent-ils en place ? Assurent-ils toujours leurs rôles initiaux ? Leurs compositions changent-elles ?

En parallèle de ce travail de recherche fondamentale, un second projet a été mené et a abouti à une publication en janvier 2018 dans *The Journal Of Antibiotics* (voir Résultat Partie1). Ce projet annexe que j'ai initié dans l'équipe avec l'aide précieuse de ma collègue technicienne Aurélia Le Dantec, a consisté à transposer le modèle de perturbation de la biogenèse des mRNPs dans des lignées cellulaires de mammifères, des ECR293 (lignée dérivée des célèbres HEK293). Nous avons pu observer, qu'à l'instar de notre modèle de levure, les cellules qui expriment Rho, vont commencer à mourir très rapidement après l'induction du facteur bactérien. Nous avons donc saisi l'opportunité d'un tel phénotype dans une lignée humaine, pour essayer de mettre au point un outil de crible pour des molécules visant le facteur bactérien Rho, qui est extrêmement important pour la vie bactérienne et très conservé à travers la plupart des phyla bactériens et spécialement chez les bactéries Gram négatives.

L'absence d'homologues structuraux et fonctionnels de Rho chez les eucaryotes, en font un candidat de premier rang pour des molécules antibactériennes. Il est déjà connu que Rho est la cible de la bicyclomycine (BCM), un produit naturel, isolé en 1972 de deux souches de *Streptomyces* différentes. La BCM est un antibiotique très efficace contre les bactéries Gram négatives. Cet effet antibiotique résulte de l'inhibition de l'activité ATPase de Rho, ce qui rend sa translocation et son activité hélicase impossible (Skordalakes et al., 2005; Zwiefka et al., 1993). En revanche, la BCM n'a pas connu un franc succès d'un point de vue pharmaceutique, de par sa mauvaise absorption par le corps humain lorsque la molécule est administrée par voie gastro-intestinale et est donc un piètre antibiotique contre les infections systémiques. Qui plus est, la BCM est également très faiblement efficace contre les bactéries Gram positive. Il est donc important de pouvoir tester de nouvelles molécules ciblant Rho, et ce de manière rapide et c'est pourquoi nous avons tiré avantage des caractéristiques que nous avons observées en exprimant Rho dans nos cellules.

Notre approche est un outil deux en un pour le crible de nouvelle molécule anti-Rho. En effet, la viabilité cellulaire étant fortement impactés par l'action de Rho dans le noyau, une

inhibition par une molécule testée nous permet de restaurer la viabilité de la cellule. De surcroît, étant donné que le test d'inhibition de l'activité de Rho se réalise directement dans les cellules eucaryotes, nous pouvons dès le premier crible détecter et éliminer des molécules avec un potentiel danger pour les cellules humaines. Les résultats ont d'abord été validés dans la levure (et corrobore avec les observations faites sur le projet principal de la thèse), en utilisant plusieurs Rho de plusieurs souches bactériennes différentes, puis nous avons étendues le test sur les cellules humaines dans les conditions de crible HTS (*High Throughput Screening*).

La suite de ce manuscrit va se subdiviser en 5 points. Tout d'abord seront présentés les résultats inhérents à la publication issue du projet annexe introduit précédemment. Par la suite, les résultats issus de mes travaux principaux de thèses seront décrits, également sous forme d'article (soumis pour publication) et porteront sur le déroutage de la machinerie de contrôle qualité nucléaire des ARNs non codants vers les ARNs codants des protéines, lorsque ces derniers sont aberrants. La suite de mes résultats seront exposés de manière plus classique, avec une partie matériels et méthodes suivie d'une partie résultat, et porteront sur la biogenèse des mRNPs et surtout sur comment les complexes nécessaires à l'emballage et l'export, réagissent face à la perturbation de la biogenèse des molécules qu'ils fixent, ainsi que leurs implications dans les procédés de contrôle qualité. Enfin, une très courte section présentera les travaux d'analyse actuellement en cours. La dernière partie présentée sera une discussion des résultats.

2. Résultat Partie 1 : Article n°1

Au cours de cette section, les résultats obtenus concernent le projet annexe à ma thèse. Cette partie va donc s'intéresser à la manière dont nous avons développé un outil de crible pour de nouveaux agents antibiotique ciblant spécifiquement l'action de l'hélicase translocase bactérienne Rho. Cette partie est présentée sous forme de publication.

Concernant cette publication, j'ai réalisé les manipulations sur les cellules ECR293 (tests de croissance et viabilité cellulaire) exceptée les western blots. J'ai également réalisé l'analyse d'alignement de séquences protéiques des différents facteurs Rho, ainsi que la visualisation de la structure. Pour la partie résultats non-publiés qui suivent l'article, j'ai effectué l'analyse des données de qPCR.

2.1. Article n°1.



Recombinant yeast and human cells as screening tools to search for antibacterial agents targeting the transcription termination factor Rho

Kévin Moreau¹ · Justine Surand¹ · Aurélie Le Dantec¹ · Christine Mosrin-Huaman¹ · Alain Legrand² · A. Rachid Rahmouni¹

Received: 31 August 2017 / Revised: 24 October 2017 / Accepted: 31 October 2017 / Published online: 25 January 2018
© The Author(s) 2018. This article is published with open access

Abstract

The alarming issue of antibiotic resistance expansion requires a continuous search for new and efficient antibacterial agents. Here we describe the design of new tools to screen for target-specific inhibitors of the bacterial Rho factor directly inside eukaryotic cells. Rho factor is a global regulator of gene expression which is essential to most bacteria, especially Gram-negative. Since Rho has no functional or structural homolog in eukaryotes, it constitutes a valuable and well known bacterial target as evidenced by its inhibition by the natural antibiotic, Bicyclomycin. Our screening tools are based on perturbation of mRNA processing and packaging reactions in the nucleus of eukaryotic cells by the RNA-dependent helicase/translocase activity of bacterial Rho factor leading to a growth defect phenotype. In this approach, any compound that impedes Rho activity should restore growth to yeast or human cells expressing Rho protein, providing valuable means to screen for target-specific antibacterial agents within the environment of a eukaryotic cell. The yeast tool expressing *E. coli* Rho factor was validated using Bicyclomycin as the control antibacterial agent. The validation of the screening tool was further extended with a stable human cell line expressing Rho factor conditionally. Finally, we show that Rho factors from different bacterial pathogens can also be designed as yeast-based screening tools which can reveal subtle variations in the functional features of the proteins.

Introduction

The continuous emergence and expansion of antibiotic resistant bacteria constitutes a serious threat for public health. This concern leads to an urgent requirement for the discovery of new and more efficient antibacterial agents, preferentially compounds that are not susceptible to existing mechanisms of resistance [1, 2]. Among the innovative drug discovery approaches, one strategy to discover new antibiotics relies on in vitro biochemical screening of large collections of synthetic or natural compounds searching for binders and/or inhibitors of well-known and relevant

bacterial targets [3]. In contrast to the older whole-cell screening methods that seek inhibition of bacterial cell growth or cell killing, the in vitro biochemical screening strategies avoid the hazardous manipulations of pathogens and hence, they are more amenable to high throughput screening methodologies (HTS). In addition, the interrogation of very large chemical libraries by HTS can be very advantageous by revealing multiple compounds acting on the same bacterial target and which can have additive or synergistic effects when used in combination [4]. However, such target-oriented in vitro screening approach for drug discovery can be very laborious and time-consuming. First, it requires the production and purification of sufficient amounts of the bacterial target to be analyzed, usually a macromolecule or a multicomponent complex. Following biochemical characterization, the approach needs the setup of an in vitro assay to measure the inhibition of the target activity. Within the context of HTS measurements, the assay should be easy-to-use and cost-effective to allow the interrogation of large chemical collections as well as reiterative screening for downstream explorations of

✉ A. Rachid Rahmouni
rahmouni@cnsr-orleans.fr

¹ Centre de Biophysique Moléculaire, UPR 4301 du CNRS, Rue Charles Sadron, 45071 Orléans, France

² University of Orléans, Pole Universitaire Centre Val de Loire, Orléans, France

structure-activity relationship to improve efficacy and optimize pharmacology of potential lead compounds.

Rho factor, which belongs to the family of ring-shaped homo-hexameric helicases, is a global regulator involved in about 50% of the transcription termination events in bacteria such as in the model micro-organism *Escherichia coli*. In addition to its involvement in the regulation of gene expression and the control of pervasive transcription [5–7], Rho plays also a major role in a number of other physiological processes including the maintenance of bacterial genome integrity by silencing foreign DNA expression and preventing the formation of transcription-driven R-loops [8–10]. These important functions explain why Rho is essential for cell growth and is highly conserved across most bacterial phyla, especially in Gram-negative bacteria [11]. Interestingly, the absence of structural and functional homologues in eukaryotes makes Rho a very attractive target for antibacterial drug discovery. Indeed, Rho is the target of the natural product Bicyclomycin (BCM) which was isolated in 1972 from two different *Streptomyces* strains and which is an effective antibiotic against Gram-negative bacteria (Reviewed in [12]). BCM exerts its action by inhibiting the ATPase activity of Rho which is required for its RNA-dependent helicase/translocase functioning [13–15]. However, BCM did not have a large pharmaceutical success because it is poorly absorbed when taken through the gastrointestinal tract and thus, it is a weak antibiotic for systemic infections. Other limitations of BCM as an antibiotic stem from the absence of effective activity towards some bacterial pathogens including Gram-positive strains, as well as the occurrence of mutation-induced resistant cells (reviewed in [12]). Nevertheless, Rho remains a valuable bacterial target for drug discovery as highlighted by a recent study showing that its depletion from *Mycobacterium tuberculosis* leads to rapid cell death both for in vitro cultures and in vivo infections in mice [16].

We took advantage of the functional features of Rho to implement an innovative eukaryotic cell-based approach that can be used as a tool to screen for new antibacterial compounds targeting Rho and to optimize their pharmaceutical efficiency through reiterative screening of lead derivatives. The method is based on perturbation of the co-transcriptional mRNA processing and packaging reactions in the yeast *S. cerevisiae* and human cells by the RNA-dependent helicase/translocase activity of the bacterial Rho factor. The heterologous expression of Rho and its localization in the nucleus of the eukaryotic cells lead to a growth defect phenotype that can be easily monitored. Thus, any compound that impedes Rho activity will restore cell growth, providing a valuable tool to search for efficient target-specific antibacterial agents within the environment of a eukaryotic cell. We report the design and

validation of the approach in a microplate-based system using yeast *S. cerevisiae* strain expressing *E. coli* Rho and with BCM as the control Rho-specific antibacterial agent. The validation of the approach was extended further using a recombinant human cell line expressing the Rho factor conditionally. Yeast-based screening tools were also constructed with Rho factors originating from different bacterial pathogens.

Materials and methods

Yeast cells and plasmids

S. cerevisiae yeast cells (BMA41 *MAT α*) were grown according to standard procedures at 25 °C in synthetic medium lacking some nucleobases and amino acids as necessary for selection and with glucose (2%) as the carbon source. The cell growth was monitored by measuring the optical density (OD) at 595 nm. The construction of the yeast plasmids expressing *E. coli* Rho with or without a Nuclear Localization Signal (NLS), Rho-NLS or Rho-ΔNLS respectively, under the control of the Doxycycline-repressed promoter (TetO₇) was previously described [17]. Briefly, the Rho open reading frame was amplified by PCR from *E. coli* genomic DNA using appropriate primers having or not the SV40 NLS sequence and designed to generate NotI restriction sites at both ends. The PCR fragments were trimmed with Not I and inserted into the pCM185 (TRP1 Marker) or pCM189 (URA3 Marker) vectors which harbor the constitutively expressed transactivator, tTA [18]. The same cloning process was utilized to construct the plasmids expressing Rho factors from different pathogens following PCR amplification of the open reading frames using genomic DNAs obtained from CIRM (Centre International de Ressources Microbiennes, Centre INRA Val de Loire, France). The appropriate plasmid clones were selected after sequencing and verification of protein expression by Western blot. The expression of Rho is repressed by growing the yeast cell in the presence of Doxycycline (Doxy) at the concentration of 5 µg/ml and its omission in the growth medium allowed maximum expression.

For the drop test experiments, ten-fold serial dilutions of yeast cultures were spotted on agar plates with or without Doxy and photographed after 3 days growth at 25 °C. For the Kirby-Bauer test, agar plates with or without 5 µg/ml Doxy were seeded with a lawn of *S. cerevisiae* bearing the Rho expression plasmid. Paper disks were dispensed on the agar surface and spotted with 7 µl of a solution of BCM at the indicated concentration or 7 µl of water for the control. The plates were incubated at 25 °C for 3 days before taking the picture.

Yeast growth in 96-well microplate

The yeast cells transformed with the appropriate Rho expression plasmid were pre-grown overnight under repressing conditions with 5 µg/ml Doxy at 25 °C in a shaker (255 rpm). Three milliliter medium were seeded with the culture and grown for 5 h with 1 µg/ml Doxy at 25 °C in a shaker (255 rpm). At this point, the OD of the culture at 595 nm is usually between 0.7 and 0.8. The cells were then centrifuged for 10 min and washed three to four times with 5 ml of fresh medium lacking Doxy. The cells were resuspended in the medium at a density of 1 OD (equivalent to 2×10^7 cell/ml) and a final dilution at 1/20 was made. Two hundred microliter of the cell suspension were dispensed in each well of a 96-well flat bottom microplate with the addition of Doxy and BCM when needed. Each growth condition sample was performed in triplicate. The microplate was covered with a lid and introduced into a TECAN incubator for growth monitoring with the following parameters: temperature 25 °C, 24 cycles of absorbance monitoring at 595 nm with 1 h interval, shaking for 1 min at 432 rpm each 10 min. The data, time points versus OD, were processed with excel software.

Generation of mammalian cell lines expressing Rho

For the construction of the mammalian plasmids, the open reading frames of Rho-NLS or Rho-ΔNLS were amplified by PCR from the yeast plasmids using appropriate primers designed to generate HindIII and XhoI restriction sites at the upstream and downstream ends of the amplicons, respectively. After restriction cutting, the DNA fragments were inserted between the HindIII and XhoI sites within the poly-linker of the pIND mammalian expression vector (Invitrogen). The integrity of the resulting plasmids (pIND-Rho-NLS and pIND-Rho-ΔNLS) was verified by sequencing.

EcR 293 cells, derived from HEK 293 cells by incorporation of the regulatory ecdysone-inducible mammalian expression system within the vector pVgRXR (Invitrogen), were transfected with 30.4 µl of PEI 10 µM (SIGMA Aldrich) in sterile H₂O mixed with 4.5 µg of pIND-Rho-NLS or the pIND-Rho-ΔNLS plasmid DNA diluted in 1 ml of physiological serum. After 20 min incubation at room temperature, the plasmid/PEI mix was directly added drop by drop into the culture wells containing cells in DMEM Fetal Bovine Serum (FBS) free medium. Cells were incubated for 2 h at 37 °C (5% CO₂), then the medium was replaced by complete DMEM AQmedia™ (SIGMA Aldrich ref: D0819) culture medium containing 1% penicillin/streptomycin (SIGMA Aldrich ref: P0781), 4 µl of Zeocin (Invitrogen ref: ant-zn-1) at 100 mg/ml and 10% FBS (SIGMA Aldrich ref: F7524). Following transfection, stable transfectants were selected by dilution cloning based

on growth in the presence of G418 at 800 µg/ml (resistance gene in pIND) and zeocin at 400 µg/ml (resistance gene in pVgRXR). The growing clones were first verified by PCR-amplicon sequencing then confirmed by Western blot analyses of Rho expression induced after addition of Ponasterone A (Sigma Aldrich ref: P3490) into the medium at 5 µg/ml. Among many positive clones, appropriate cell lines exhibiting the best repression and inducibility of Rho were chosen. Aliquots of the cell lines were frozen for cryopreservation in liquid nitrogen and a new aliquot was used after each 10 passages. For each passage, confluent cells were rinsed with phosphate-buffered saline (PBS) pH 7.4. After cell detachment using trypsin, the cell suspension was used to seed new culture flasks at a 1/10 split ratio.

Mammalian cell viability assay

Rho-expressing EcR 293 cell lines were seeded into a 96-well plate at 7,000 cells per well in 100 µl of DMEM AQmedia™ culture medium. Following cells adhesion, Rho induction was performed by replacing the culture medium by fresh medium containing Ponasterone A at 5 µg/ml. When needed, BCM was directly added to the wells at different concentrations (10, 20 or 50 µg/ml, final concentration) at the 12 h growth time point. Cells viability was assessed at each time point by using PrestoBlue™ Viability Reagent (life technologies ref: A13262) according to the manufacturer's instructions. Briefly, 10 µl of the reagent were added directly to the culture medium in each well. Following 2 h incubation (37 °C, 5% CO₂), the plate was shaken 5 min on a plate shaker then read with a Victor 3 V plate reader (PerkinElmer). Shortly after reading the plate, pictures were taken with a 16 million pixel camera by pipetting the colored media from the wells and pouring them into a new 96-well plate.

Western blotting

Total proteins in whole-cell extracts were obtained as described [19], then resolved on SDS-10% polyacrylamide electrophoresis gels and analyzed by Western blotting according to standard procedures. The Rho protein was detected with rabbit anti-Rho antibodies (a custom preparation from Eurogentec, Belgium). Secondary antibodies (goat anti-rabbit-HRP IgG, Promega) were used and visualization on X-ray films was performed following enhanced chemiluminescence (Clarity Western ECL Substrate from Biorad).

Structure and multi-alignment study

Structure visualization was performed using The PyMOL Molecular Graphics System, Version 1.7.2.1 Schrödinger,

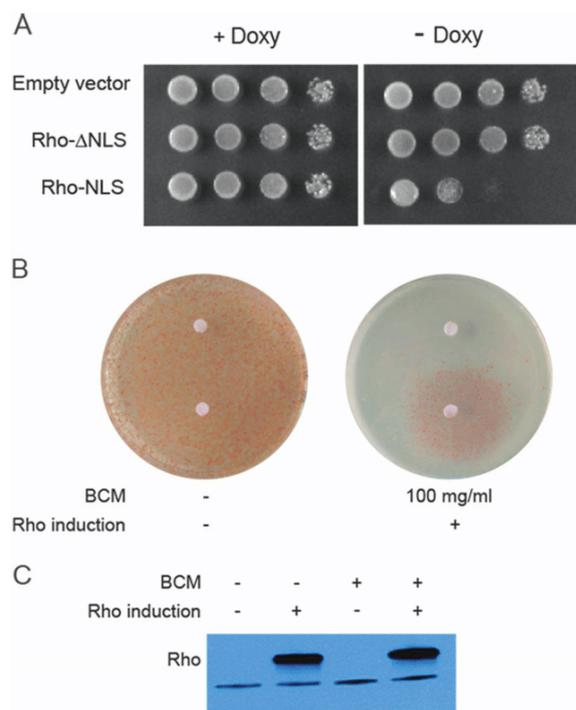


Fig. 1 Rho expression in the nucleus of yeast cells induces a growth defect. **a** Ten-fold serial dilutions of yeast cells transformed with the empty vector or the indicated plasmids expressing Rho-NLS or Rho-ΔNLS were spotted on agar plates in the presence or absence of 5 μg/ml of Doxy and grown at 25 °C for 3 days. **b** Kirby-Bauer disk diffusion test. Agar plates with or without 5 μg/ml Doxy were seeded with a lawn of *S. cerevisiae* bearing the Rho-NLS expression plasmid. Paper disks were dispensed on the agar surface and spotted with 7 μl of a solution of 100 mg/ml of Bicyclomycin (BCM, bottom disk) or 7 μl of water (top disk) for the control. Pictures were taken after 3 days incubation at 25 °C. **c** Western blot using anti-Rho antibodies showing Rho expression upon induction by omission of Doxy in the growth medium in the absence or presence of 100 μg/ml of BCM. The lower band corresponds to a yeast protein cross-reacting non-specifically with the anti-Rho antibodies

LLC. Structural data were downloaded from PDB (accession number 1XPU). Multiple sequence alignments were done using the EBI online tool Clustal Omega with default parameters.

Results

The activity of *E. coli* Rho factor in the nucleus of *S. cerevisiae* impedes cell growth

Bacterial Rho factor functions as a powerful helicase/translocase that uses energy from ATP hydrolysis to track along an RNA chain and can melt nucleic acid base pairing or disrupt RNA-protein complexes present in its path [20, 21]. We have shown previously that the activity of Rho

in the yeast nucleus impairs the co-transcriptional recruitment of mRNA processing and packaging factors yielding mRNAs that are aberrantly processed or depleted of some crucial export proteins. Such defective mRNAs are targeted and eliminated by the nuclear quality control system coupled to the RNA degradation machinery, a process that hinders cell fitness and growth [22–24]. As shown in Fig. 1a with drop tests, ectopic expression in *S. cerevisiae* of *E. coli* Rho factor harboring a Nuclear Localization Signal (Rho-NLS) at its C-terminus leads to a growth defect phenotype. Rho is expressed under the control of a Doxycyclin-repressed TetO₇ promoter within a centromeric plasmid bearing the constitutively expressed trans-activator (tTA). The growth defect is not observed when the ectopically produced Rho is no longer addressed to the cell nucleus by the localization signal (Rho-ΔNLS construct). Interestingly, a similar relief of growth defect is achieved by inhibiting Rho activity with the antibiotic BCM as observed by the Kirby-Bauer disk diffusion test (Fig. 1b). The revival of yeast cells expressing Rho in a zone surrounding the paper disk containing BCM is an indication regarding the target specificity of the antibiotic and the absence of any obvious toxicity towards the eukaryotic host cells. As shown by Western blot, the suppression of the growth defect by BCM is clearly linked to the inhibition of Rho activity and not to a potential shutdown of the protein production (Fig. 1c).

Microplate-based measurements of Rho-induced growth defect in yeast

To adapt our approach to high-throughput screening conditions, we monitored the Rho-induced growth defect in liquid medium using 96-well microplates with 200 μl cultures. The growth measurements were conducted by recording the OD of the cultures at 595 nm every hour over a period of 24 h with vigorous shaking at 25 °C in a TECAN incubator (Fig. 2). The control growth curve of yeast cells containing an empty vector reached the stationary phase plateau at 1.2 OD after 18 to 20 h incubation under both repressing (+Doxy) and inducing (–Doxy) conditions. A similar growth curve profile was observed for the yeast cells harboring the Rho expression plasmid and grown under repressing conditions. In contrast, the growth curve of yeast cells containing the Rho expression plasmid remained flat under inducing conditions, reaching only 0.2 OD after 20 h of growth. In agreement with the disk diffusion test results, the presence of BCM in the medium restores growth to the cells expressing Rho with a curve profile rising up to 0.8 OD when the concentration of the antibiotic was raised to 200 μg/ml. These growth profiles indicate that the rescue is not complete, probably due to a limited intracellular availability of the Rho inhibitor (see discussion). However, the mitigation of Rho-induced

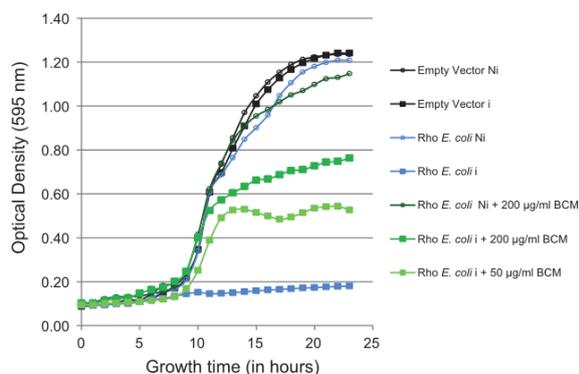


Fig. 2 Growth curves of yeast cells harboring the empty plasmid or the plasmid expressing *E. coli* Rho-NLS. The growth was monitored in a TECAN incubator with 200 µl cultures in 96-well microplates under inducing (i) or non-inducing (Ni) conditions. For the relevant cultures, BCM, at the indicated final concentration, was added in the medium at the start of growth

growth defect by BCM became clearly apparent after only 10 to 12 h growth. Thus, within the context of an HTS screen, it is possible to detect within this time interval the inhibitory effect of a potential compound targeting Rho activity.

Recombinant human cell line as Rho-based screening tool

Conceivably, potential inhibitors of Rho activity may escape detection during a screening process with the yeast tool if they have an antifungal activity that kills the cells. To circumvent this possible limitation, we developed a recombinant human cell line as a complementary Rho-based screening tool. DNA encoding the open reading frame of Rho-NLS or Rho-ΔNLS was shuffled from the yeast plasmids into the pIND mammalian expression vector and the resulting constructs were transfected into EcR 293 cells (HEK 293 cells derivatives which were stably transformed with the regulatory vector pVgRXXR) [25]. Resistance to Zeocin and G418 was used to select for potential positive clones. Stable clonal cell lines were established and the induction of Rho expression by Ponasterone was confirmed by Western blot analyses as described in Materials and Methods and shown in Fig. 3a. The selected cell lines showed a tight regulation of the expression system with the appearance of Rho protein 6 h following induction (Fig. 3b).

To evaluate the effect of Rho expression on cell proliferation, the growth of the selected cell lines was monitored visually on 6-well plates over a period of 7 days. In the absence of Rho expression, the cells grew to at least 90% confluence during the 7 days period, covering nearly the whole surface of the well (Fig. 3c, lane 1). The

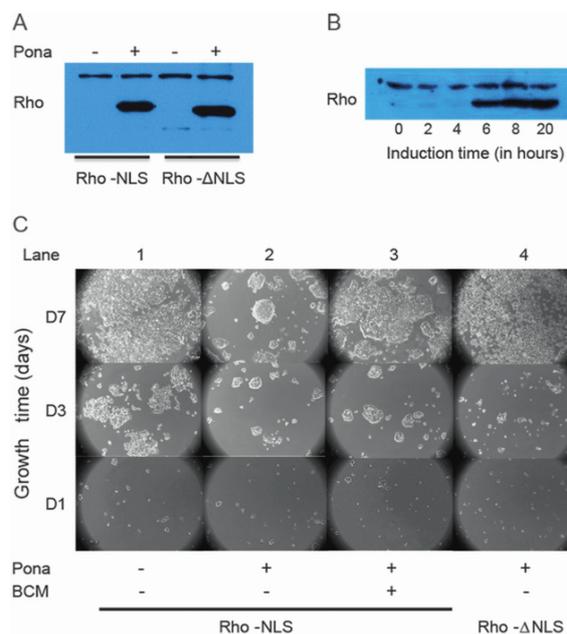


Fig. 3 Rho expression in EcR 293 cell lines induces a cytostatic effect. **a** Western blot showing the conditional expression of Rho-NLS and Rho-ΔNLS by the recombinant EcR 293 cell lines. Pona stands for the inducer Ponasterone which was added to the cell culture medium at 5 µM. The upper band corresponds to a protein cross-reacting non-specifically with the anti-Rho antibodies. **b** Western blot showing the expression of Rho-NLS as a function of time. **c** Photograph showing the growth over 7 days of the recombinant EcR 293 cell lines expressing conditionally Rho-NLS or Rho-ΔNLS. The addition of BCM to the relevant culture at 50 µg/ml final concentration was done 12 h after cells seeding

induction of Rho-NLS expression by Ponasterone had a dramatic effect on growth with the cells reaching only 25–30% confluence after the 7 days incubation (Fig. 3c, lane 2). This cytostatic effect was not observed with the cell line expressing the Rho derivative lacking the NLS (Rho-ΔNLS, lane 4), emphasizing the need for nuclear accumulation of the bacterial factor to impair the growth of the eukaryotic cell. Importantly, the addition of BCM to the culture medium (12 h after cell seeding) alleviates the cytostatic effect of Rho-NLS expression with the cells reaching nearly 80% confluence after the 7 days of growth (Fig. 3c, lane 3). These results show that Rho expression in the nucleus of human cells induces a similar growth defect than in yeast.

Adaptation of the Rho-based human cell line tool to HTS conditions

To adjust the new tool to high-throughput screening conditions, the effect of Rho expression on cell proliferation was assessed quantitatively using the resazurin-based PrestoBlue® Cell Viability Reagent that measures the reducing

environment of viable cells. In the presence of live cells, the color of the reagent turns from blue to red whose intensity can be measured by absorbance-based plate readers. The cell viability measurements were processed at 12, 24, 72, and 96 h after cell seeding. As shown by the plate photograph in Fig. 4a, the detection by the reagent is in complete agreement with the visual observation of the cell growth. In the absence of Rho-NLS expression or with the expression of Rho- Δ NLS, a switch to red color was observed over the growth period. In contrast, the induction of Rho-NLS expression did not show significant color change over time, an indication for low cell density. Again, the effect of BCM in rescuing the cell proliferation is clearly seen by the color indications in the wells. The absorbance values of the samples taken at each time point were plotted to reveal the growth curve profiles (Fig. 4b). The graph reveals that the effect of Rho inhibition can be assessed as early as 36 h after compound addition (48 h after cell seeding) with a larger confidence at the 72 h time point of growth where there is more than twofold difference in the level of cell viability between the samples with and without BCM addition. Together, these results show that the human cell tool is readily appropriate for large scale screening of compound libraries to search for potential inhibitors of bacterial Rho activity.

Yeast growth defect induced by Rho proteins originating from different bacterial pathogens

In general, the amino acids involved in Rho functional features such as RNA binding and ATP hydrolysis are relatively conserved across bacterial phyla. However, some variations in other functionally important motifs that mediate the RNA-dependent helicase/translocase activity have been observed among Rho proteins from different bacterial species ([11] and references cited therein). Having this in mind, we wanted to assess the extent to which the yeast-based screening tool can be generalized to other bacteria and especially some clinically relevant pathogenic strains. To this end, we cloned Rho proteins of *A. baumannii*, *E. coli* ESBL, *E. cloacae*, *E. faecalis*, *P. aeruginosa*, and *S. enterica* in the yeast centromeric expression vector with the NLS at their C-terminus. After transformation into *S. cerevisiae*, the expression of the different Rho proteins was verified by Western blot (Fig. 5a). Although we did not perform quantitative assessment of the expression level for each Rho protein, some Rho species appear to be less expressed compared to the expression level of *E. coli* Rho. Notably, the Western blot reveals lower accumulations for Rho proteins from *A. baumannii*, *E. faecalis*, and *P. aeruginosa*.

The effect of the different Rho proteins on the growth of *S. cerevisiae* was first monitored by serial dilutions on agar

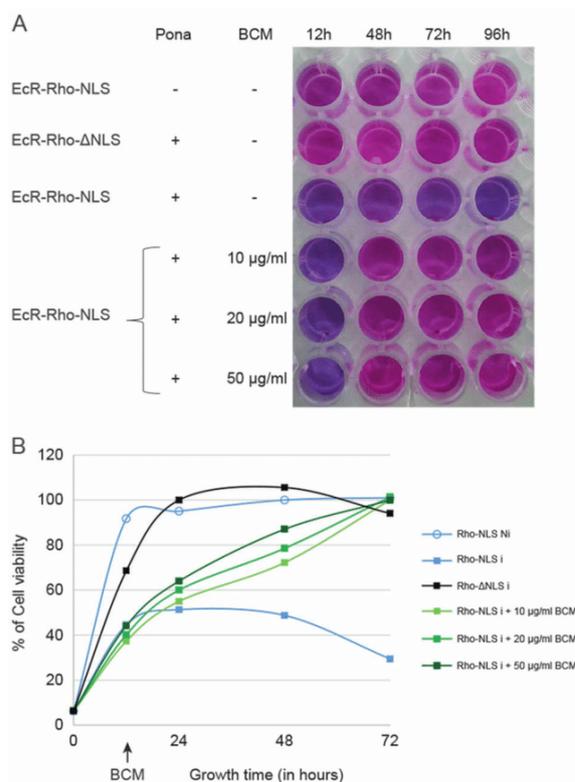


Fig. 4 Quantitative measurements of mammalian cells viability following Rho expression. **a** Picture of a part of a 96-well plate showing the colored media resulting from the reaction of PrestoBlue™ Viability Reagent. BCM was added to the media of the relevant culture wells at the indicated final concentration 12 h after cell seeding. **b** Graphs showing the growth curves obtained by reading the 96-well plate with a Victor 3 V plate reader following revelation by PrestoBlue™. (Ni) stands for non-induced and (i) for induced

plates in the presence or absence of the effector, Doxy (Fig. 5b). Under inducing conditions (–Doxy), the growth of the yeast cells expressing Rho proteins was severely reduced. However, the growth defect seems to be notably less pronounced with the *P. aeruginosa* Rho protein. This trend was also observed under HTS-like growth conditions by monitoring the growth in liquid medium using 96-well microplates with 200 μ l cultures (Fig. 6).

The moderate effect of *P. aeruginosa* Rho on yeast growth may imply a less proficient helicase/translocase activity of the protein. This possibility was explored by comparing the Rho sequence of *P. aeruginosa* to that of *E. coli* (Fig. 7a). Interestingly, the sequence alignment shows a substitution of an important and highly conserved amino acid in the ATP binding pocket for *P. aeruginosa* Rho (Met-186-Ile). Indeed, as illustrated in Fig. 7b, previous structural studies of *E. coli* Rho have shown that the side chains of two conserved amino acids, Met-186 and Phe-355, are responsible for sandwiching the adenine

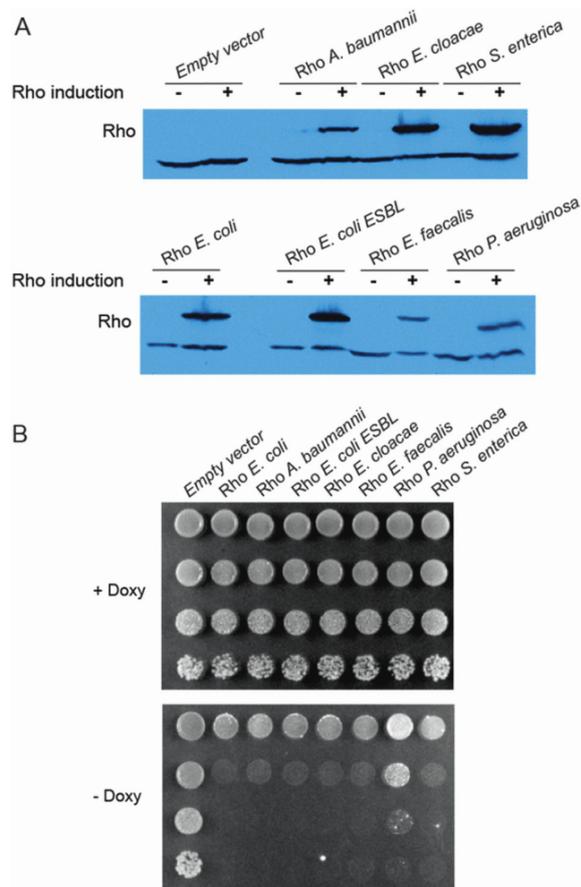


Fig. 5 Yeast growth defect induced by Rho proteins from different bacterial pathogens. **a** Western blot showing the conditional expression in yeast of Rho proteins from the indicated bacterial strains. **b** Ten-fold serial dilutions of cells transformed with the empty vector or the plasmid expressing the indicated Rho variants having an NLS at their C-terminus. The cells were spotted on agar plates with or without 5 µg/ml Doxy and incubated at 25 °C for 3 days

moiety of bound ATP [26]. Thus, one possibility could be that lower binding affinity of ATP may decrease the rate of its hydrolysis which in turn will weaken the strength and/or the pace of Rho translocation along the RNA chain. Such scenario could explain a less efficient interference in the yeast mRNA processing and packaging reactions with ensuing modest impairment to the cell fitness.

Discussion

The development and spread of antibiotic resistance in bacteria is a global threat to public health. One useful way to tackle this concern resides in the search for new drugs that inhibit the activity of well-known and relevant bacterial

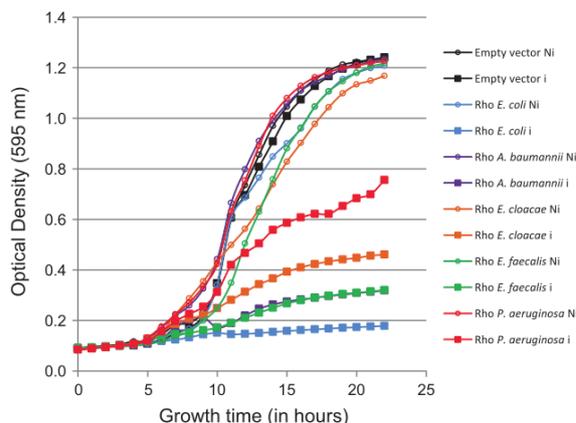
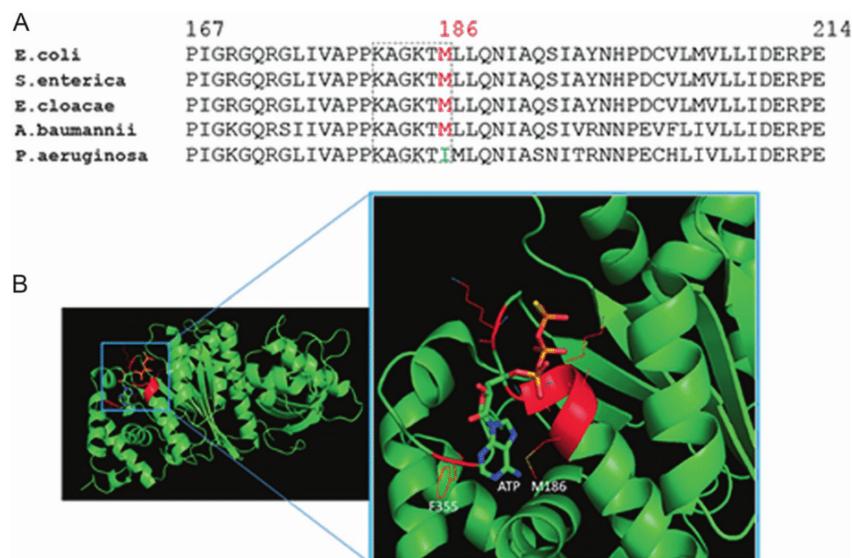


Fig. 6 Growth curves of yeast cells harboring the empty plasmid or the plasmid expressing Rho-NLS proteins originating from the indicated bacterial strains. The growth was monitored in a 96-well plate as indicated in Fig. 2. (Ni) stands for non-induced and (i) for induced

molecular targets. Hence, this requires the design of original methodologies and tools to screen, at high throughput and low cost, large chemical libraries looking for lead compounds [3]. Here, we describe the implementation in yeast and human cells of simple and innovative tools that can be used to screen for Rho-specific antibacterial compounds directly inside a eukaryotic cell. The tools are based on our previous work showing that the ectopic expression of *E. coli* Rho factor in the nucleus of *S. cerevisiae* induces a growth defect phenotype. We sought to extend the approach to screening conditions where any relevant inhibitor of Rho activity should suppress or alleviate the growth defect. The methodology was validated using the Rho-specific natural antibiotic BCM as a positive control.

The yeast tool was adapted to HTS-like conditions by monitoring cell growth and growth recovery in 200 µl liquid cultures within 96-well plates in a semi-automated TECAN incubator. The presence of BCM in the culture medium mitigates the Rho-induced growth defect but did not lead to a complete suppression of the growth deficiency even at the concentration of 200 µg/ml which shows only a small improvement relative to 50 µg/ml. At concentrations above 200 µg/ml, the antibiotic became toxic to the yeast cells (results not shown). Presumably, the uptake of BCM is not optimal in yeast due to the presence of a cell wall. Nevertheless, the BCM-mediated mitigation of Rho-induced growth defect is clearly detectable within the beginning of the exponential phase of growth after 10 to 12 h of incubation. In the context of a fully automated and multiplexed HTS platform, such time frame should be appropriate to interrogate a large chemical library and reveal potential hits in a reasonable time and at low cost. Furthermore, the time necessary to reach the exponential phase can be shortened by changing the growth conditions such as raising the

Fig. 7 Possible implication of a lower affinity of the ATP pocket in the moderate growth defect induced by Rho from *P. aeruginosa*. **a** Multiple sequence alignment showing the mutation Met-186-Ile within the ATP pocket of Rho from *P. aeruginosa* as compared to that of *E. coli* and other bacterial strains. **b** Rho protomer crystal structure PDB 1XPU [26] showing the ATP pocket where the adenine moiety of ATP is stacked between the side chains of the two conserved amino acid residues Met186 and Phe355



temperature of incubation to 30 °C instead of 25 °C used in our experiments which should make the screening process less time-consuming. A subsequent analysis by the Kirby-Bauer diffusion test should be ideal to fine-tune the validation of potential hits regarding their specificity and possible toxicity at high concentration.

The extension of our approach to human cells provides additional strength to the screening tools by analyzing the effect of Rho inhibitors directly in the appropriate cellular environment which should exclude toxic compounds. Thus, despite a longer time frame for growth and detection of an effective inhibitor, the human cell line tool presents many advantages that should complement the yeast based assay. For instance, the cell line tool could be used to reevaluate hits selected by a first round of rapid screens made by the yeast tool. Indeed, with the human cell line tool, Rho inhibition by BCM induces a full relief of the growth defect which should give more confidence in the detection of positive hits. Also, the effect of BCM is observed at a relatively low concentration of the antibiotic, probably due to a higher uptake by the human cell in the absence of a cell wall. These features should increase the likelihood of discovering an efficient antibiotic by detecting numerous positive hits in a screening process. The positive hits will be subsequently submitted to downstream selections regarding major issues linked to bacterial strains such as uptake and avoidance by the efflux system.

Finally, we report that our screening tools can also be designed with Rho proteins originating from different bacterial pathogens. The use of a pathogen-defined tool in a screen may be helpful to discover drug candidates that

inhibit Rho activity by binding to specific regions of the protein that are not shared with other bacterial strains. Interestingly, by testing the growth defect induced by the pathogen-defined tools we show that our approach can reveal subtle features regarding the functioning of a specific Rho factor. Thus, the tools could be utilized in studies connected with the structure/function relationship of Rho as a molecular motor.

Acknowledgements We are grateful to M Si-Tahar and V Hervé for discussions regarding the project and for sharing time slots to use the TECAN incubator. We thank R Washburn and K Schnetz for the gift of BCM. This work was supported by Région Centre Val-de-Loire through a research grant (Anti-Rho-Biotics) to A R Rahmouni and a Ph.D. fellowship to K Moreau.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, and provide a link to the Creative Commons license. You do not have permission under this license to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Bush K, et al. Tackling antibiotic resistance. *Nat Rev Microbiol*. 2011;9:894–6.
2. Lukačičinová M, Bollenbach T. Toward a quantitative understanding of antibiotic resistance evolution. *Curr Opin Biotechnol*. 2017;46:90–7.
3. Brown ED, Wright GD. Antibacterial drug discovery in the resistance era. *Nature*. 2016;529:336–43.
4. Wambaugh MA, Shakya VPS, Lewis AJ, Mulvey MA, Brown JCS. High-throughput identification and rational design of synergistic small-molecule pairs for combating and bypassing antibiotic resistance. *PLoS Biol*. 2017;15:e2001644.
5. Peters JM, et al. Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev*. 2012;26:2621–33.
6. Peters JM, et al. Rho directs widespread termination of intragenic and stable RNA transcription. *Proc Natl Acad Sci USA*. 2009;106:15406–11.
7. Bidnenko V, et al. Termination factor Rho: from the control of pervasive transcription to cell fate determination in *Bacillus subtilis*. *PLoS Genet*. 2017;13:e1006909.
8. Cardinale CJ, et al. E. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science*. 2008;320:935–8.
9. Krishna Leela J, Syeda AH, Anupama K, Gowrishankar J. Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*. *Proc Natl Acad Sci USA*. 2013;110:258–63.
10. Washburn RS, Gottesman ME. Transcription termination maintains chromosome integrity. *Proc Natl Acad Sci USA*. 2011;108:792–7.
11. D'Heygere F, Rabhi M, Boudvillain M. Phyletic distribution and conservation of the bacterial transcription termination factor Rho. *Microbiology*. 2013;159:1423–36.
12. Williams RM, Durham CA. Bicyclomycin: synthetic, mechanistic, and biological studies. *Chem Rev*. 1988;88:511–40.
13. Zwiefka A, Kohn H, Widger WR. Transcription termination factor rho: the site of bicyclomycin inhibition in *Escherichia coli*. *Biochemistry*. 1993;32:3564–70.
14. Lawson MR, Dyer K, Berger JM. Ligand-induced and small-molecule control of substrate loading in a hexameric helicase. *Proc Natl Acad Sci USA*. 2016;113:13714–9.
15. Skordalakes E, Brogan AP, Park BS, Kohn H, Berger JM. Structural mechanism of inhibition of the Rho transcription termination factor by the antibiotic bicyclomycin. *Structure*. 2005;13:99–109.
16. Botella L, Vaubourgeix J, Livny J, Schnappinger D. Depleting *Mycobacterium tuberculosis* of the transcription termination factor Rho causes pervasive transcription and rapid death. *Nat Commun*. 2017;8:14731. <https://doi.org/10.1038/ncomms14731>.
17. Mosrin-Huaman C, Honorine R, Rahmouni AR. Expression of bacterial Rho factor in yeast identifies new factors involved in the functional interplay between transcription and mRNP biogenesis. *Mol Cell Biol*. 2009;29:4033–44.
18. Gari E, Piedrafitia L, Aldea M, Herrero E. A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast*. 1997;13:837–48.
19. Kushnirov VV. Rapid and reliable protein extraction from yeast. *Yeast*. 2000;16:857–60.
20. Schwartz A, Margeat E, Rahmouni AR, Boudvillain M. Transcription termination factor rho can displace streptavidin from biotinylated RNA. *J Biol Chem*. 2007;282:31469–76.
21. Walmacq C, Rahmouni AR, Boudvillain M. Testing the steric exclusion model for hexameric helicases: substrate features that alter RNA-DNA unwinding by the transcription termination factor Rho. *Biochemistry*. 2006;45:5885–95.
22. Honorine R, Mosrin-Huaman C, Hervouet-Coste N, Libri D, Rahmouni AR. Nuclear mRNA quality control in yeast is mediated by Nrd1 co-transcriptional recruitment, as revealed by the targeting of Rho-induced aberrant transcripts. *Nucleic Acids Res*. 2011;39:2809–20.
23. Mosrin-Huaman C, Hervouet-Coste N, Rahmouni AR. Co-transcriptional degradation by the 5'-3' exonuclease Rat1p mediates quality control of HXK1 mRNP biogenesis in *S. cerevisiae*. *RNA Biol*. 2016;13:582–92.
24. Stuparevic I, Mosrin-Huaman C, Hervouet-Coste N, Remenaric M, Rahmouni AR. Cotranscriptional recruitment of RNA exosome cofactors Rrp47p and Mpp6p and two distinct Trf-Air-Mtr4 polyadenylation (TRAMP) complexes assists the exonuclease Rrp6p in the targeting and degradation of an aberrant messenger ribonucleoprotein particle (mRNP) in yeast. *J Biol Chem*. 2013;288:31816–29.
25. No D, Yao TP, Evans RM. Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc Natl Acad Sci USA*. 1996;93:3346–51.
26. Skordalakes E, Berger JM. Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor. *Cell*. 2006;127:553–64.

2.2. Discussion de l'article et perspectives liées

Nous avons donc eu l'occasion de voir de quelle manière le système de crible de molécules au potentiel antibiotique a été développé et les résultats qu'il a pu fournir avec un témoin positif comme la BCM. Nous avons donc un outil de crible rapide et présentant un double niveau de sélection. En effet, le système levure, permet de cribler rapidement des molécules pour leurs effets anti-Rho en suivant la croissance des cellules. Et dans un second temps, les molécules présélectionnées, peuvent être testées sur le modèle de cellules humaines, qui va nous fournir un résultat sur l'efficacité de la molécule dans le modèle cellulaire le plus proche de la réalité biologique d'utilisation des molécules, et en outre révéler en direct la toxicité potentielle de ces produits pour l'homme.

En plus de nous fournir ces résultats intéressants d'un point de vue technico-pharmaceutique, les études menées sur les cellules de mammifères avec le modèle de perturbation Rho nous a donnée de belles perspectives quant à l'étude des systèmes de contrôle qualité de la biogenèse des mRNPs dans des lignées cellulaires humaine.

Effectivement, suite aux expériences pour le système de crible, les niveaux de certains transcrits, ont été testé en présence et en absence de Rho. Nous avons pu observer des résultats similaires à ceux obtenus chez la levure, une diminution de certains ARNm. De plus dans une lignée cellulaire « knock-down » pour la sous-unité catalytique de l'exosome humain (EXOSC10), le niveau des transcrits en question est restauré (Figure 2.2.1-A).

Des expériences de FISH (Fluorescence In Situ Hibribization) visant la queue polyA des ARNm ont également été menées (Figure 2.2.1-B). Les résultats nous indiquent qu'en présence de Rho, les ARNm matures aberrants forment des agrégats à l'intérieur du noyau des cellules. Ces résultats ont été confirmés par des analyses en qPCR réalisées sur des fractions nucléaires et cytoplasmiques (Figure 2.2.1-C), qui nous informent que lorsque les cellules sont soumises à l'effet de Rho, la quantité d'ARN nucléaire augmente comparée à la quantité cytoplasmique, signe d'une rétention nucléaire.

Notre modèle d'étude a donc pu être étendue aux cellules humaines à travers deux lignées distinctes (ECR293 et HeLa). Ces résultats ne sont que préliminaires et demandent à être confirmés et approfondies, mais nous permettent d'espérer pouvoir dans le futur étudier,

dans un modèle humain, les mécanismes de contrôle qualité nucléaire des ARNs avec notre approche de perturbation de leurs biogenèses par Rho.

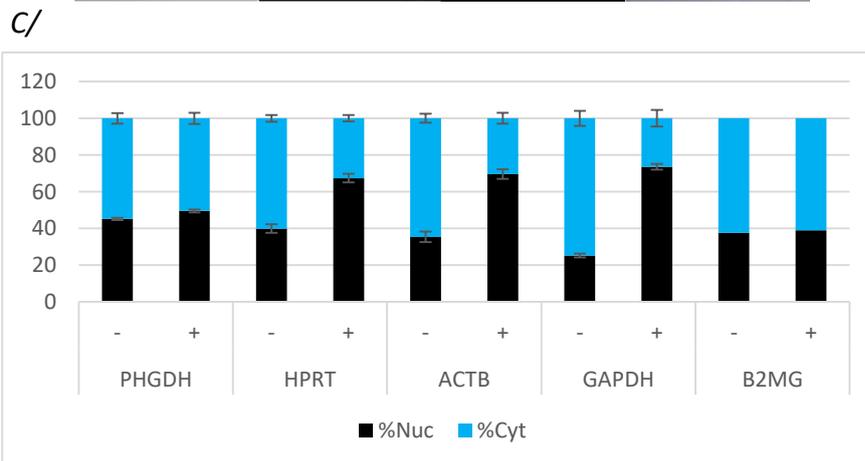
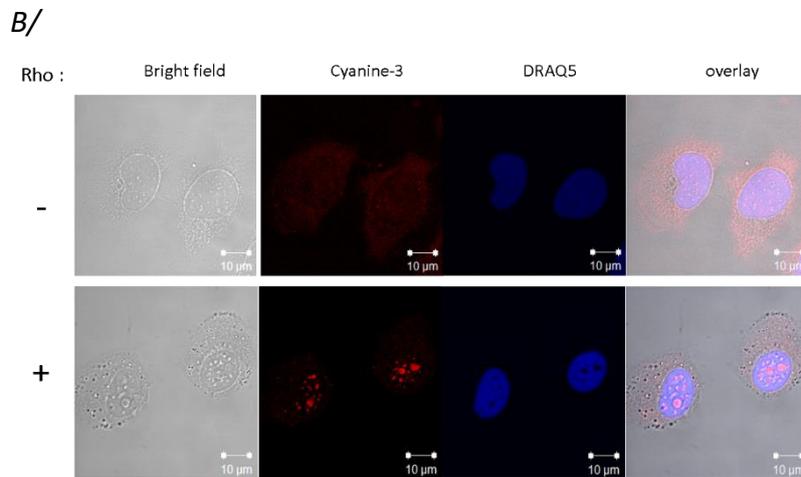
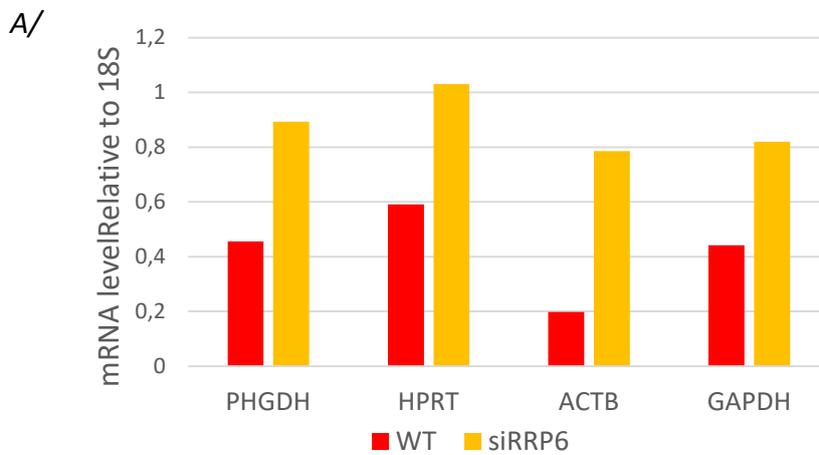


Figure 2.2.1. Effet de Rho sur les ARNs polyadénylés d'une lignée cellulaire humaine. A/RT-qPCR sur 4 ARNs modèles montrant l'impact de Rho (12h d'induction) dans un contexte WT et lors d'une déplétion de Rrp6 par siRNA. Le niveau basal de chaque ARN est fixé à 1. B/ Images de microscopie à fluorescence illustrant la répartition des ARNs polyA en absence et présence de Rho (24h d'induction) via une sonde de FISH poly-dT. Les noyaux sont marqués au DRAQ5. C/ RT-qPCR sur fractions nucléaire et cytoplasmique d'ARN. Les diagrammes empilés indiquent les proportions d'ARNs nucléaires et cytoplasmique en absence ou présence de Rho. Ces résultats sont normalisés sur la quantité totale d'ARN et sur un ARN spécifique de chaque compartiment (snU6 pour le noyau et l'ARNr 18S pour le cytoplasme).

3. Résultat Partie 2 : Article n°2

Cette section concerne la première partie des résultats obtenus au cours de mes principaux travaux de thèse. Les résultats sont présentés sous forme de manuscrit d'article et ce dernier est sous presse dans le journal RNA Biology.

Afin d'introduire cette section et de permettre une bonne compréhension de l'article qui va suivre, nous allons reprendre les principaux points qui vont être abordés.

Concernant cette partie résultat, j'ai réalisé les manipulations sur les levures, de la culture aux extractions d'ARNs. J'ai également réalisé les CHIP de Nab3 et Nrd1. Les librairies d'ADN pour le séquençage des données de CHIP ont également été réalisés par mes soins pour Nab3, Nrd1 et Rrp6. J'ai traité l'intégralité des données issues de séquençage haut débit, supervisé par Mr Benoît Piégu. Enfin, j'ai traité et analysé les données de qPCR.

3.1. Introduction à l'article

Les cellules eucaryotes ont développé, au cours de l'évolution, un système nucléaire de contrôle qualité (QC) des mécanismes qui conduisent à l'empaquetage et la maturation des ARNm en particules ribonucléoprotéiques compétentes pour l'export vers le cytoplasme. Les mRNPs aberrantes sont reconnues par le système de contrôle qualité entraînant leur rétention dans le noyau et les conduisant fatalement à leur dégradation par l'activité exonucléase de la protéine Rrp6. Ces étapes de surveillances de l'intégrité de la mRNP, sont concomitantes à toutes les étapes de la transcription et de la formation de la mRNP et vont donc pouvoir détecter n'importe quel évènement conduisant à une molécule défailante. Afin d'étudier les mécanismes de contrôle qualité, nous utilisons l'outil présenté dans la dernière partie de l'introduction, l'hélicase translocase bactérienne Rho. Ici, les précédentes observations faites par l'équipe via Rho sur quelques gènes modèles ont été étendues à l'ensemble du génome de la levure. Des expériences de séquençage à haut débit de l'ARN (RNA-seq) nous ont permis de déterminer un grand nombre de transcrits sensible à l'action de Rho, ainsi que la portion de ces derniers sauvés lors de la déplétion de Rrp6. Ces résultats sont étendus par des expériences d'immunoprécipitation de la chromatine suivie de séquençage haut débit (CHIP-seq) qui nous ont permis de suivre les mouvements de différents acteurs du système de contrôle qualité. Nos résultats montrent que la perturbation de la

biogénèse des mRNPs induit un basculement de ces acteurs du QC des loci produisant des ARNs non codants vers les loci correspondants aux ARNm dont la biogénèse est perturbée.

3.2. Manuscrit de l'article n°2

Perturbation of mRNP biogenesis reveals a dynamic landscape of the Rrp6-dependent surveillance machinery trafficking along the yeast genome

Kévin Moreau¹, Aurélia Le Dantec¹, Christine Mosrin-Huaman¹, Yves Bigot²,

Benoit Piégu² and A. Rachid Rahmouni^{1, 3}

¹Centre de Biophysique Moléculaire, UPR 4301 du CNRS, Rue Charles Sadron, 45071 Orléans, France. ²Physiologie de la Reproduction et des Comportements UMR7247 INRA-CNRS, 37380 Nouzilly, France.

³Corresponding author: rahmouni@cns-orleans.fr

Running title: Genome-wide mapping of QC components

Keywords: mRNP biogenesis, mRNP Quality Control, Yeast, RNA-seq, CHIP-seq

ABSTRACT

Eukaryotic cells have evolved a nuclear quality control (QC) system to monitor the co-transcriptional mRNA processing and packaging reactions that lead to the formation of export-competent ribonucleoprotein particles (mRNPs). Aberrant mRNPs that fail to pass the QC steps are retained in the nucleus and eliminated by the exonuclease activity of Rrp6. It is still unclear how the surveillance system is precisely coordinated both physically and functionally with the transcription machinery to detect the faulty events that may arise at each step of transcript elongation and mRNP formation. To dissect the QC mechanism, we previously implemented a powerful assay based on global perturbation of mRNP biogenesis in yeast by the bacterial Rho helicase. By monitoring model genes, we have shown that the QC process is coordinated by Nrd1, a component of the NNS complex (Nrd1-Nab3-Sen1) involved in termination, processing and decay of ncRNAs which is recruited by the CTD of RNAP II. Here, we have extended our investigations by analyzing the QC behavior over the whole yeast genome. We performed high-throughput RNA sequencing (RNA-seq) to survey a large collection of mRNPs whose biogenesis is affected by Rho action and which can be rescued upon Rrp6 depletion. This genome-wide perspective was extended by generating high resolution binding landscapes (ChIP-seq) of QC components along the yeast chromosomes before and after perturbation of mRNP biogenesis. Our results show that perturbation of mRNP biogenesis redistributes the QC components over the genome with a significant hijacking of Nrd1 and Nab3 from genomic loci producing ncRNAs to Rho-affected protein-coding genes, triggering termination and processing defects of ncRNAs.

INTRODUCTION

Expression of protein-coding genes in eukaryotic cells is a multistep process in which the genetic message is transcribed from DNA into a pre-mRNA molecule that undergoes several modification reactions such as 5'-end capping, splicing, 3'-end cleavage and polyadenylation while being assembled into a ribonucleoprotein particle (mRNP) that will be exported into the cytoplasm for translation. The current view is that the nascent transcript emerging from the transcription complex is sequentially loaded with a variety of protein factors that ensure its integrity and mediate its maturation and transport from the site of transcription to the nuclear pore for export [1,2]. Like most complex biological activities, the process of co-transcriptional mRNP biogenesis is error prone and thus, the cells have evolved a surveillance system that detects malformed mRNPs resulting from inappropriate or inefficient processing and packaging reactions [3,4]. Aberrant mRNPs that fail to pass the quality control (QC) steps are targeted for destruction by the RNA degradation machinery associated with the nuclear exosome complex [5,6]. The interdependence and coordination between transcript synthesis and the other activities of mRNP biogenesis, QC and nuclear export are facilitated by the C-terminal domain (CTD) of the largest subunit of RNA polymerase II (RNAP II) that serves as a platform for sequential recruitment of the various acting factors.

Although the real frequency of mRNP biogenesis defects occurring under physiological context is actually unknown, its putative prevalence does not allow a direct study of the QC mechanism within the cellular framework. Thus, most insights into the QC process have come from studies focusing on yeast *S. cerevisiae* mutants with defects in the THO-Sub2 complex that mediates mRNP assembly and export. It was shown that the mechanism of nuclear retention and degradation of defective mRNPs depends on the presence of the nucleus-specific exosome-associated exonuclease Rrp6 as well as components of the exosome activating complex TRAMP (Trf4-Air2-Mtr4). However,

these analyses which relied on depletion or mutation of specific components of the THO-Sub2 complex were unable to provide an integrated and functional view of the QC system, especially regarding the co-transcriptional recognition of aberrancies and its role in the subsequent triggering of mRNA decay. Moreover, the observed effects were limited mostly to a subset of transcripts produced under heat shock conditions such as HSP104 [7-9]. Another unanswered question is whether the nuclear-restricted exonuclease Rrp6 operates independently to degrade aberrant transcripts or whether it requires a physical association with the core exosome for this specific activity.

In recent years, we implemented an experimental approach in which the general perturbation of mRNP biogenesis in *S. cerevisiae* by the RNA-dependent helicase/translocase activity of the bacterial Rho factor generates sufficient amounts of defective mRNPs as substrates to investigate the QC process (reviewed in [10]). Our analyses of a set of model mRNAs such as PMA1 substantiated the key role of the 3'-5' exonuclease Rrp6 in the elimination of defective mRNPs. Furthermore, the results revealed that the QC process is coordinated by Nrd1, a component of the early termination complex (Nrd1-Nab3-Sen1) which is recruited by the CTD of RNAP II through its CTD-interacting domain (CID). The targeting and degradation of Rho-induced aberrant PMA1 mRNP is associated with a large increase of Nrd1 recruitment to the transcription complex with a concomitant enrichment of Rrp6 together with its cofactor Rrp47 as well as components of the TRAMP complex [11,12].

In the present work, we sought to investigate the general nature of the previous observations by analyzing the mRNP QC activities over the whole yeast genome. We performed high-throughput RNA sequencing (RNA-seq) to survey a broad collection of mRNPs whose biogenesis is affected by Rho action. Among the 1,015 transcripts that showed decreased levels in the presence of Rho, almost a half (491 transcripts) were rescued to nearly their original levels by the

depletion of Rrp6. This genome-wide perspective was extended by generating high resolution maps (ChIP-seq) of the distribution of most QC components along the yeast chromosomes before and after perturbation of mRNP biogenesis by Rho. Our results show that the perturbation of mRNP biogenesis redistributes the binding of QC components over the genome with a large shift of Nrd1 and Nab3 from gene loci producing noncoding RNAs (ncRNAs) prone to processing and decay (CUTs, SUTs and snoRNAs) to protein-coding genes. Nrd1 and Nab3 are apparently titrated out from ncRNAs genomic features by a large recruitment to Rho-affected mRNA gene loci, leading to transcriptional termination defects of ncRNAs. The ChIP profiles display statistically significant overlapping of occupancy peaks suggesting a co-localization of the QC components over discrete portions of the Rho-affected mRNA genes.

RESULTS

Rho-based perturbation of mRNP biogenesis in yeast

In our experimental approach, large amounts of defective mRNPs can be generated by the activity of bacterial Rho factor in yeast cells to serve as substrates to study the QC mechanism (reviewed in [10]). Rho expressed from a centromeric plasmid under the control of the Doxycycline (Doxy) repressible TetO7 promoter is directed to the nucleus by a nuclear localization signal fused to the C-terminus of the protein. Rho is a hexameric helicase/translocase that binds to naked, C-rich and preferentially unstructured portions of RNA. This binding activates the ATPase activity of the protein leading to its translocation along the RNA chain in a 5' to 3' direction. Our previous investigations with a model gene (*PMA1*) showed that Rho action along the nascent transcript interferes with the normal recruitment of mRNA processing and packaging factors yielding mRNPs that are recognized as defective and eliminated by the QC apparatus in a process involving the 3'-5' exonuclease activity of Rrp6. By monitoring the

rescue of the *PMA1* mRNA level in different deletion strains, we showed that the targeting and degradation of Rho-induced aberrant mRNPs by Rrp6 is stimulated by the cofactors Rrp47 and Mpp6 as well as the TRAMP complex [12]. An interesting finding made in those studies, which was also reported by others [13-15], is that Rrp47 plays an important role in a mutual protein stabilization with Rrp6, highlighting a close physical association between the two protein partners [15]. In effect, removal of one partner by gene deletion induces a depletion of the other partner by more than 90% [12]. Given that the *rrp6Δ* strain is known to be temperature sensitive and to have a slow growth as well as a slight genetic instability, we used in the present work the *rrp47Δ* strain to deplete Rrp6 when needed for the rescue of the Rho-affected transcripts.

High throughput determination of Rho-induced aberrant mRNPs targeted by the Rrp6-dependent QC

To survey the effect of Rho on a large collection of mRNAs, we performed stranded RNA-seq using rRNA-depleted RNAs isolated from wild-type and *rrp47Δ* yeast cells grown in the presence (Rho-repressing conditions) or the absence (Rho-inducing conditions) of the effector, Doxy. For each condition, two biologically independent samples were used for libraries preparation and sequencing. In general, the results of the duplicates were highly correlated with Spearman correlation coefficients above 0.98 (Figure S1). The RNA-seq single reads were aligned on the *S. cerevisiae* genome and the mapped reads matching annotated transcripts were distributed into different RNA biotypes. About 60% of the mapped transcripts were protein-coding mRNAs and the remaining 40% represented ncRNAs including CUTs, SUTs, XUTs, snRNAs and snoRNAs (Figure S2A). Upon Rho expression in the wild-type cells, a substantial fraction of the transcripts within all RNA biotypes showed variations, either up or down, in their steady-state levels as illustrated by the scatter plots (Figure S2B).

We first focused our analyses on the Rho-induced variations in the abundance of mRNAs to have insights into the fate of defective mRNPs. We calculated the statistical significance of the variations (mRNA ratios between +Rho and -Rho conditions) and plotted the log₂ fold change of the level of each mRNA as a function of its log₂ mean expression (Figure 1A). Among the 5,132 mRNAs mapped with the RNA-seq results, 1,015 mRNAs showed decreased levels upon Rho expression in wild-type cells with a log₂ fold change ≤ -0.5 (blue dots in Figure 1A). A nearly symmetrical fraction of mRNAs had increased steady-state levels in the presence of Rho with a log₂ fold change above 0.5 (red dots in Figure 1A). The remaining part of mRNAs did not show statistically significant variations in the presence of Rho (grey dots). As expected, we

interpret these results as an indication that Rho activity in yeast does not affect the biogenesis of all mRNPs. Indeed, all nascent transcripts may not proffer RNA segments suitable for Rho loading and activation. Alternatively, the Rho-induced higher levels observed for a fraction of mRNAs could be explained by a trans-regulation effect where the Rho-mediated reduction in the level of one mRNA or ncRNA up-regulates the expression of one or several transcripts. When the statistical significance calculations (ratios between +Rho and -Rho mRNA levels) and plotting were applied to the RNA-seq datasets of the mRNAs produced in the *rrp47* Δ strain, about half (491 transcripts) of the mRNAs that were diminished by Rho action in wild-type cells were partially reverted close to their normal levels by the deletion of *RRP47* gene, i.e. depletion of Rrp6 (grey dots in

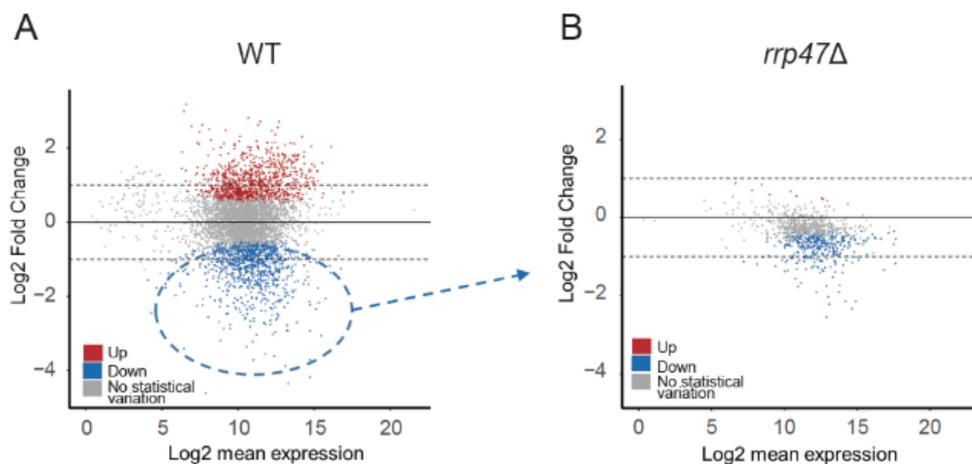


Figure 1: Fate of a subpopulation of mRNAs down-regulated by Rho action in Wt cells and their rescue in a strain with a deletion of *RRP47*, ie depleted of Rrp6. **A)** MAplot of the log₂FoldChange ratios of the mRNAs between -Rho and + Rho conditions (log₂(+/-Rho)) from WT cells distinguishes three subpopulations: statistically upregulated in red, statistically downregulated in blue and no statistical variation in grey. The dashed blue circle surrounds the population of interest which is down-regulated. **B)** MAplot as in A but for the mRNAs obtained from *rrp47* Δ cells and focuses only on the subset of mRNAs that were down-regulated in WT cells as determined in A.

Figure 1B). Four of these mRNAs were randomly picked for individual validation by RT-qPCR and showed relative recoveries of their levels in the *rrp47* Δ strain (Figure S3). The remaining half of the Rho-mediated down-regulated mRNAs (524 transcripts) did not show any significant recovery in the *rrp47* Δ strain, according to our cutoff value (blue dots in Figure 1B).

To have a global view on these sorting results, we used the RNA-seq datasets obtained under the different conditions to generate averaged meta-transcript profiles (Figure 2). The profiles for the whole mapped mRNAs did not show any significant variations among the different conditions, reflecting the averaging of the RNA-seq raw

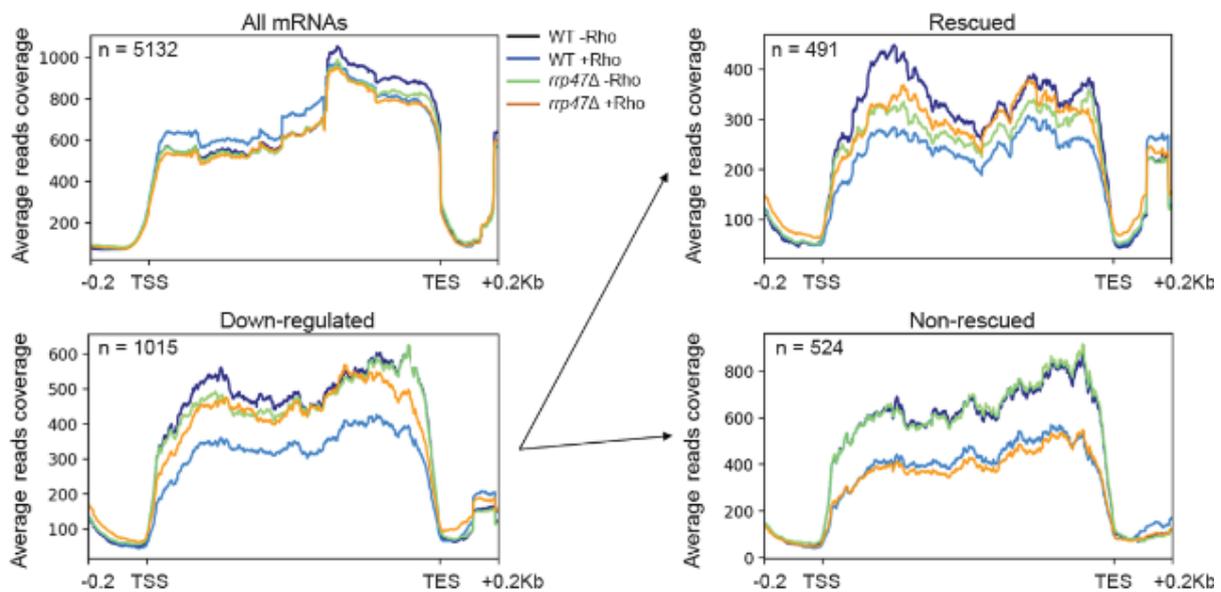


Figure 2: Meta-transcript profiles variations of mRNAs obtained from WT and *rrp47Δ* cells in the absence or presence of Rho. The top left graph shows the average reads coverage for all analyzed mRNAs. The bottom left graph represents the down-regulated mRNAs subpopulation as determined in Figure 1A. The two right graphs display the two groups of the down-regulated mRNAs sorted according to their rescue in the *rrp47Δ* genetic background (Rescued) or not (Non-rescued) as determined in Figure 1B. For all the graphs, the dark blue line represents mRNAs obtained from WT cells in the absence of Rho whereas the light blue gives the meta-profiles of the mRNAs obtained from the same cells but in the presence of Rho. Light green and orange display the meta-profiles of the mRNAs obtained from the *rrp47Δ* strain in the absence and presence of Rho, respectively.

data. For the whole group of down-regulated mRNAs (1,015 transcripts), the action of Rho in the wild-type cells leads to a general decrease in the intensity of the averaged meta-transcript, indicating that, on average, a fraction of corresponding mRNPs has been recognized as aberrant and eliminated. This conclusion is supported by the absence of any marked effect of Rho on the level of the averaged profile for mRNAs produced in the *rrp47Δ* strain. Sorting the group of down-regulated mRNAs into two categories (rescued and non-rescued by Rrp6 depletion) showed two different trends regarding the Rho-induced variations as well as the shape of the averaged meta-transcript profiles. The pool of rescued mRNAs appears to contain relatively less expressed transcripts with a higher average reads coverage at the 5' end of the metagene which could be ascribed to the accumulation of short transcripts resulting from RNAP II pausing at the early

part of the genes. Such paused transcription complexes are likely terminated by the action of Rho as reported previously [16] which could explain the absence of their full recovery in the *rrp47Δ* strain. The meta-transcript profiles obtained with the pool of non-rescued mRNAs reveal an averaged higher expression of the transcripts with a clear absence of rescue in the *rrp47Δ* strain. We infer that the absence of recovery in the *rrp47Δ* strain for this mRNA group is likely an indication that the corresponding Rho-induced defective mRNPs are targeted by the alternative Rat1-dependent QC pathway, as we reported previously for HXK1 mRNP [17]. An in-depth comparative analysis between the two mRNA populations and their specific processing by the two QC pathways (Rrp6-dependent and Rat1-dependent) will be published elsewhere (Moreau et al. manuscript in preparation). Altogether, these results

show that a set of at least 491 yeast mRNPs whose biogenesis was affected by Rho activity are recognized and targeted for destruction by the Rrp6-dependent QC machinery.

Perturbation of mRNP biogenesis by Rho affects the Nrd1-dependent processing and termination of ncRNAs

To investigate further the response of Rrp6-dependent QC to Rho-mediated perturbation

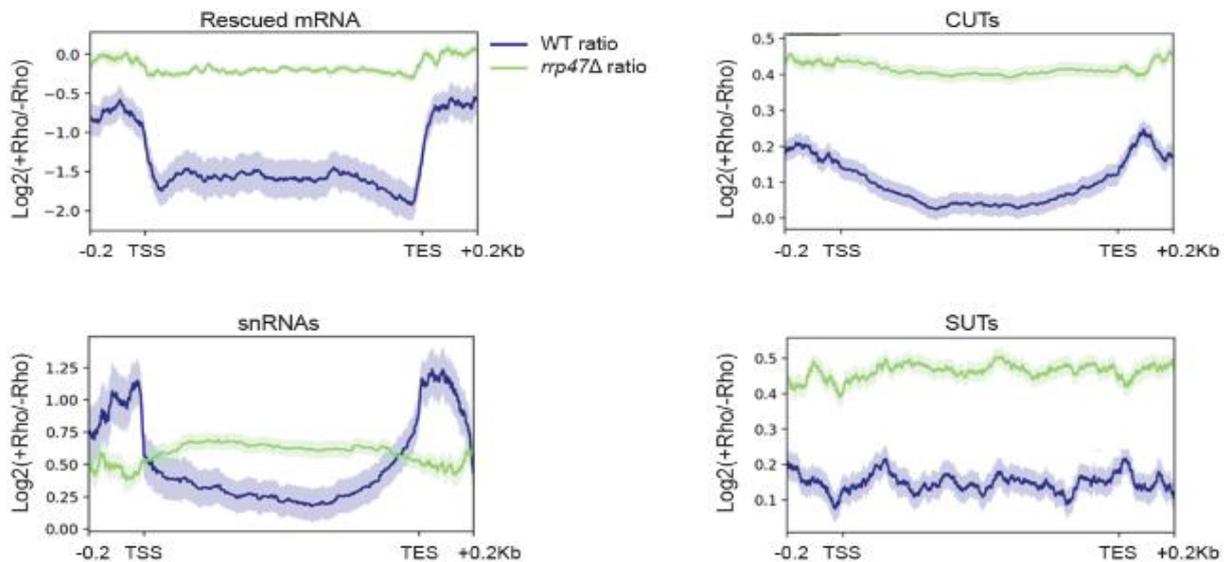


Figure 3: Differential effect of Rho on mRNAs and ncRNAs. The Log₂ ratios of RNA levels between +Rho and -Rho conditions are plotted on the metagenes for each RNA biotype as indicated on the top of each graph. Blue lines are for RNAs extracted from Wt cells and green lines are for RNAs obtained from the *rrp47Δ* strain. Standard error is represented by the lightened area around each line.

of the yeast transcriptome, we extended the comparative analyses of the RNA-seq datasets to ncRNAs including CUTs, SUTs and snoRNAs which are known to be degraded and/or processed by Rrp6 [18-20]. We generated averaged meta-transcript profiles for each ncRNA category produced in the absence or presence of Rho and in wild-type or *rrp47Δ* cells. In contrast to the Rho-induced decreasing trend observed for the affected mRNAs, the metaprofiles of the ncRNAs showed rather a slight but significant increase of the averaged steady-state levels in the presence of Rho. The Rho-induced stabilization of the ncRNAs is more pronounced in the *rrp47Δ* strain (Figure S4). This differential effect of Rho between mRNAs and ncRNAs is clearly illustrated when the ratios of RNA levels between +Rho and -Rho conditions are plotted as log₂ fold change values on the metagenes (Figure 3). Whereas the fold change values for the rescued mRNAs are contained between 0 for the *rrp47Δ* strain

and around -2 for the wild-type, the log₂ fold changes for the ncRNAs are all positive spanning between 0.1 for the wild-type and 0.5 for the deletion strain. These results suggest that the perturbation of the transcriptome by Rho activity leads to a stabilization of ncRNAs either directly or indirectly. As expected, this RNA stabilization is further increased upon depletion of the exonuclease Rrp6 in the *rrp47Δ* strain.

A visual inspection of the reads coverage in a 100 bp window flanking the annotated termination sites indicated that, in addition to the general increase of the reads signals, Rho induces also termination defects of the ncRNAs. The presence of extended transcripts arising from termination defects is more obvious when analyzing independently transcribed snoRNA genomic features, although the presence of such transcript extensions could also be seen for some isolated

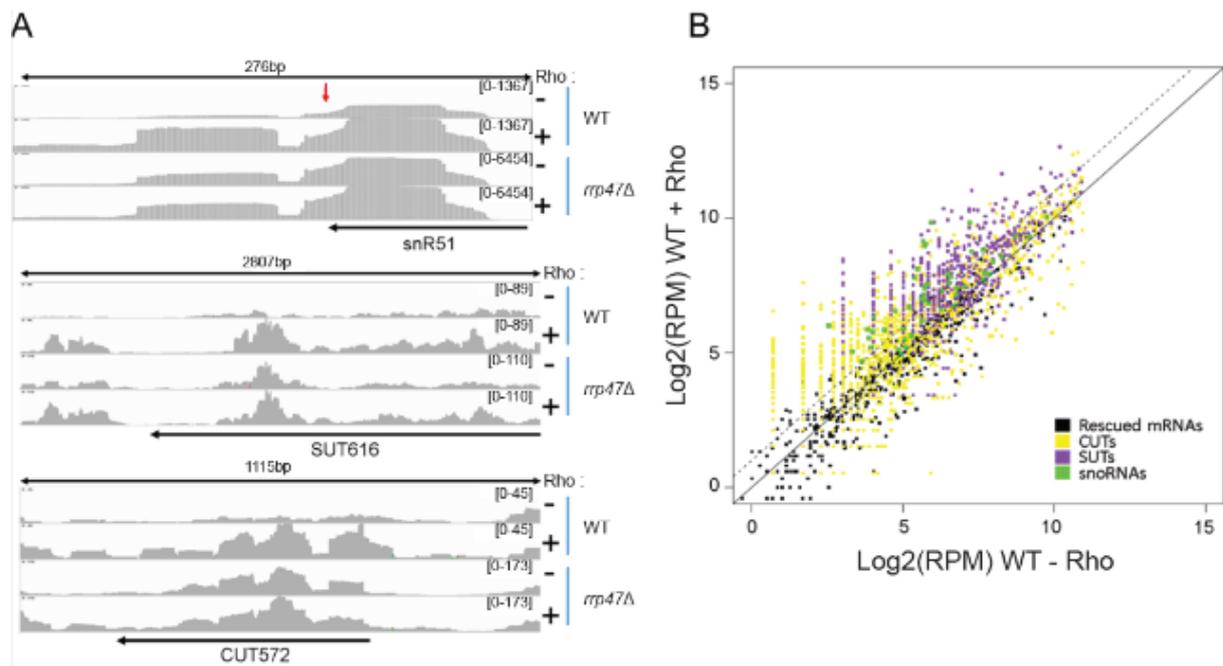


Figure 4: Rho activity in yeast affects the termination and processing of ncRNAs. **A)** IGV snapshots of raw RNA-seq reads focusing on a snoRNA, a SUT and a CUT. Each snapshot shows the reads coverage of RNAs extracted from WT or *rrp47Δ* strains and with or without Rho induction. The scale for the number of reads obtained from the raw RNA-seq data is shown in brackets on the right of each snapshot. Red arrow for snR51 indicates the annotated termination site for the snoRNA. **B)** Scatter plot of Log₂ normalized reads counts in a 100 bp region downstream of annotated genomic features for RNAs extracted from WT cells under Rho perturbation versus the same measure for RNAs obtained from unperturbed cells. Dashed line shows the cutoff of L2FC = 1. The color code for the analyzed RNA biotypes is inside the graph.

CUTs and SUTs (Figure 4A and Figure S5). Note, however, that the scales of the reads counts in Figure 4A and Figure S5 differ between the wild-type and the *rrp47Δ* strain for each ncRNA category, reflecting the substantial additional stabilization of these RNAs upon depletion of Rrp6. To have a global view of the Rho-induced termination defects on all the studied RNAs, we plotted the normalized reads counts in the 100 bp region downstream from each annotated transcription unit obtained for the wild-type strain in the absence of Rho versus the same measure obtained in the presence of Rho (Figure 4B). The threshold for a significance of termination read-through was set as log₂ fold change ≥ 1 (dotted line) and with p-values below 0.05. The plot revealed that the termination of mRNAs was barely affected by Rho (only 5% of the transcripts above the cutoff), whereas the termination defects were observed for 56% of snoRNAs, 35% of CUTs and 50% of SUTs.

High resolution mapping of Rrp6-dependent QC components association with chromatin

Our previous work on a model mRNA transcript (PMA1) revealed that the co-transcriptional targeting of Rho-induced aberrant mRNPs is mediated by a stimulation of Nrd1 recruitment to the gene locus with a concomitant enrichment by components of the Rrp6-dependent QC. To explore the general nature of the process, we analyzed the genome-wide association with chromatin in - Rho and + Rho conditions of Nrd1, Nab3, Rrp6 and Trf4. Yeast strains harboring C-terminally tagged proteins expressed from their chromosomal locations under their own promoters were used for chromatin immunoprecipitation followed by single-end deep sequencing (ChIP-seq). The raw data obtained from two independent biological replicates were first verified for their correlation ($R \geq 0.95$) and alignment on the

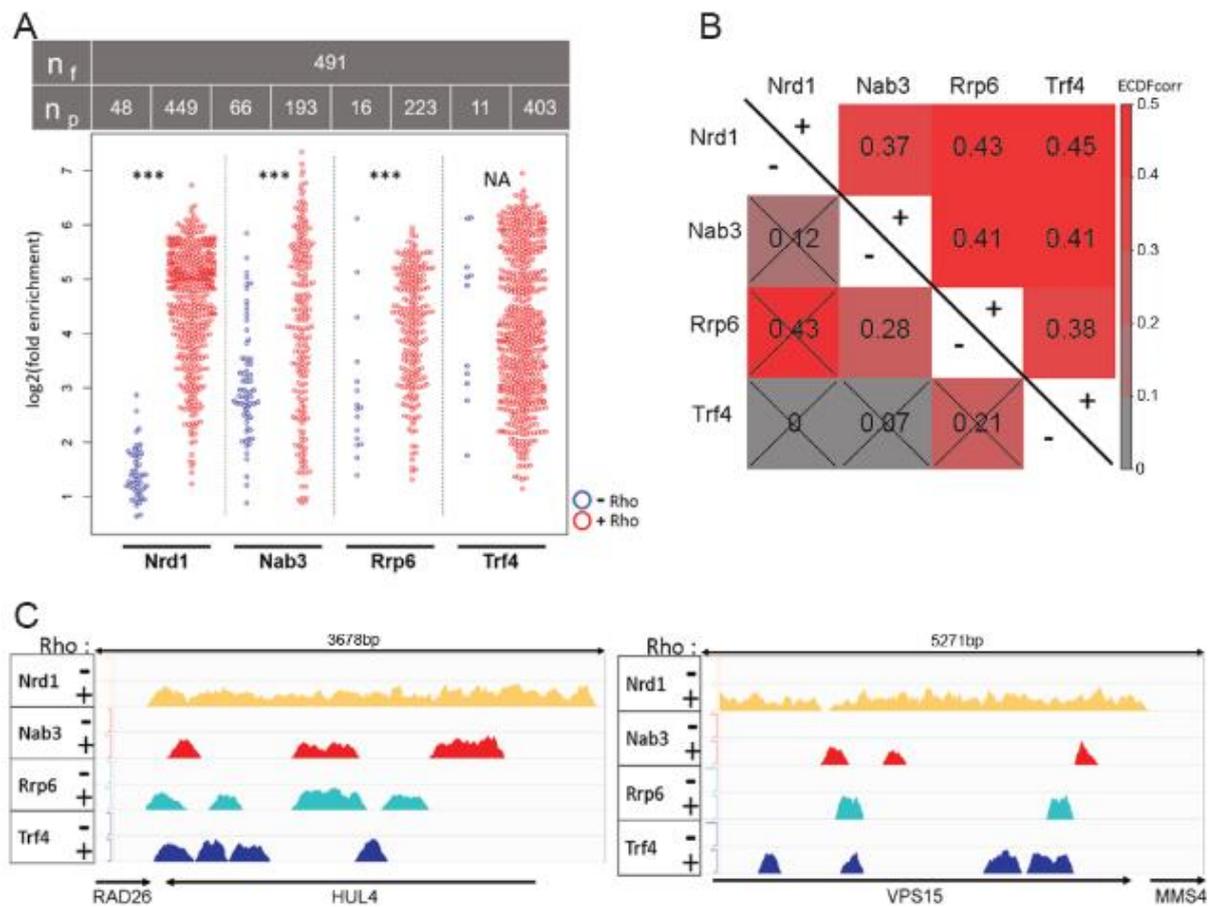


Figure 5: Rho activity mediates the stimulation of massive recruitment of QC components to Rrp6-sensitive mRNAs genomic loci. **A)** Beeswarm plot of \log_2 fold enrichment detected on the rescued mRNAs genomic loci obtained by ChIP-seq experiments for each indicated QC protein. The ChIP signals are shown as open dots in blue for -Rho and in red for +Rho conditions. Student t test was used to assess the significance of enrichment differences between the -Rho and +Rho conditions with the following code $0.05 > * > 0.01 > ** > 0.001 > ***$. (NA) stands for not applicable. The table above the graphs gives the number of analysed features (n_f) and the number of loci possessing at least one peak for each proteins (n_p) in -Rho or +Rho conditions. **B)** Correlogram of Empirical Distribution Cumulative Function correlation area (calculated with genomericorr package) for each protein pairs over the analysed genomic loci under Rho perturbation (“+” symbols) or not (“-” symbols). The crossed boxes indicate absence of correlation significance. **C)** IGV snapshots of ChIP-seq peaks mapped over two rescued mRNAs genomic loci. The ChIP signals for each protein are shown both for -Rho and +Rho conditions.

yeast genome. Following this quality examination, the data were used for peak calling by calculating reads ratios between IP and Input in a comparative mode where only peak signals detected simultaneously in the two replicates were considered. The piled-up peak signals obtained for - Rho and + Rho conditions were subsequently subjected to a step of differential analysis where only peaks appearing for one condition but not for the other were scored [21]. This differential ChIP-seq analysis is more appropriate to determine differences between conditions when

analyzing large populations of peaks. The processed data for the genomic features of the 491 mRNAs that were affected by Rho and rescued upon Rrp6 depletion are presented as a Beeswarm plot in Figure 5A. The results reveal a strong Rho-mediated stimulation of Nrd1 and Nab3 recruitment to the genes, with 91% of the genes showing at least one peak for Nrd1 and 39% for Nab3. Moreover, the scoring results show also a higher level of occupancy by the two proteins in the presence of Rho. The Rho-induced large enrichment to chromatin within the Rho-affected genes is also clearly

seen for Rrp6 and Trf4 with 45% and 82% of the genes showing at least one peak, respectively.

These ChIP-seq results indicate that the Rrp6-dependent QC components are involved physically in the recognition and targeting of aberrant mRNPs at the co-transcriptional level. Indeed, close visual examinations of individual genomic loci, as shown by representative examples in Figure 5C and Figure S6, highlight the Rho-mediated stimulation of QC recruitment to chromatin. The landscape of Nrd1 association with chromatin reveals broad occupancy profiles that cover the whole body of the affected genes. This underlines the known coupling of Nrd1 with transcribing RNAP II following its recruitment by the CTD (see discussion). Alternatively, the occupancy profiles for Nab3, Rrp6 and Trf4 display rather sharp peaks with significant cross-overlapping, suggesting a possible co-localization of the QC components over portions of the Rho-affected genes. To elucidate a potential spatial correlation between the occupancy sites for the four proteins, we conducted a global analysis of the peaks distributions over the 491 Rho-affected mRNA genes. We used the *genometricorr* package to assess Empirical Distribution Cumulative Function (ECDF) correlations and evaluated the uniformity of the distributions with the Kolmogorov-Smirnov test. The pair-wise comparisons of the peak signals for the four ChIPed proteins revealed that their relative distributions deviate from independence (Figure S7). Moreover, the averaged relative distances between the peaks of protein pairs over the analyzed genomic loci showed significant positive correlation under Rho expressing conditions compared to repressing conditions (Figure 5B). This suggests a close spatial localization of the QC proteins within chromatin upon co-transcriptional recognition and targeting of the aberrant mRNPs.

Rho-mediated ncRNAs processing and termination defects are linked to loss of Nrd1 and Nab3 recruitment to the genomic features

Previous studies have reported that nuclear depletion of Nrd1 by the “anchor-away”

method gives rise to extended ncRNAs as a result of termination and processing defects [22,23]. In particular, defects in the Nrd1-dependent termination pathway have been shown to induce the extension of snoRNA transcripts into downstream sites similar to our observations described above [22,24-26]. These results led us to consider the possibility that the effect of Rho on processing and termination of ncRNAs may stem somehow from a specific depletion of Nrd1 within these genomic loci. To explore this hypothesis, we extended our ChIP-seq data processing to ncRNAs genomic features. The distribution of the ChIP-seq signals detected over the analyzed ncRNAs genomic loci are presented as Beeswarm plots with the CUTs and SUTs pooled together (Figure 6A) and snoRNAs depicted alone (Figure 6B). In the absence of Rho, Nrd1 enrichment signals were detected over 67% of the CUTs + SUTs and 87% of the snoRNA genomic loci. Some of the enrichment signals, in particular those over the snoRNAs, are relatively strong. A similar trend was observed for the ChIP-seq analysis obtained with Nab3, supporting the known close association of the two proteins at the ncRNAs termination sites. Remarkably, the presence of Rho leads to a drastic decrease of both the number of Nrd1 and Nab3 peaks detected as well as their intensity over the CUTs and SUTs. For the snoRNAs loci, however, the action of Rho seems to differ somewhat leading to a significant decrease of the enrichment intensity for Nrd1 but only a slight reduction in the number of peaks detected (Figure 6B). Additional analyses showed that the QC components (Rrp6 and Trf4) are also recruited to ncRNAs loci (Figure 6 A and B). For Rrp6, however, the intensity of the peak signals is relatively low and does not change significantly between -Rho and +Rho conditions. In contrast, Rho activity appears to stimulate the recruitment of Trf4 to ncRNAs loci in a similar way than for the Rho-affected mRNA genes.

These results together with those of the RNA-seq analysis described above substantiate the known key role of Nrd1 and its partner Nab3 in the termination and processing of ncRNAs.

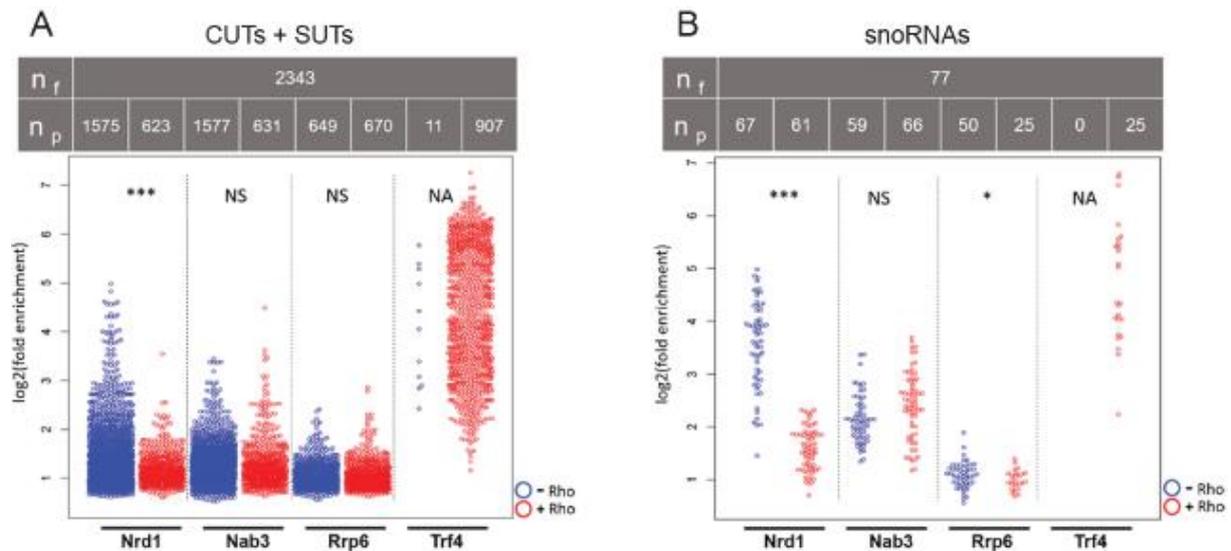


Figure 6: Eviction of QC components from ncRNAs genomic features under conditions of perturbation of mRNPs biogenesis by Rho. Beeswarm plots of log₂ fold enrichment obtained by ChIP-seq experiments for each indicated QC protein detected on CUTs and SUTs in (A) and on snoRNAs genomic loci in (B). The symbols and labelling are the same as in Figure 5A. (NS) stands for non-significant and (NA) for not applicable.

Moreover, the data highlight a link between Rho activity in yeast and a remodeling of the Nrd1-dependent termination machinery within the ncRNAs genomic features leading to transcriptional read-through. To monitor the interplay between the loss of QC proteins from ncRNAs loci and their enrichment within mRNA genes as a result of Rho action, we sought to visualize the protein distributions over the whole genome. This global view was obtained using the Omic Circos software package that generates a circular layout of the 16 yeast chromosomes wreathed with the ChIP-seq

peaks detected for the four proteins (Figure 7). Comparisons between the plots obtained for the ncRNAs and Rho-affected mRNAs under -Rho and +Rho conditions show clearly a wide shift reminiscent of a migration of Nrd1 and Nab3 from ncRNAs genomic positions to mRNA genes in the presence of Rho. This suggests that Nrd1 and Nab3 are probably titrated-out from ncRNAs genomic features by their attachment to Rho-affected mRNA genes, generating extended ncRNAs. In contrast, it appears that the Rho-induced stimulation of Rrp6 recruitment to mRNA genes does not

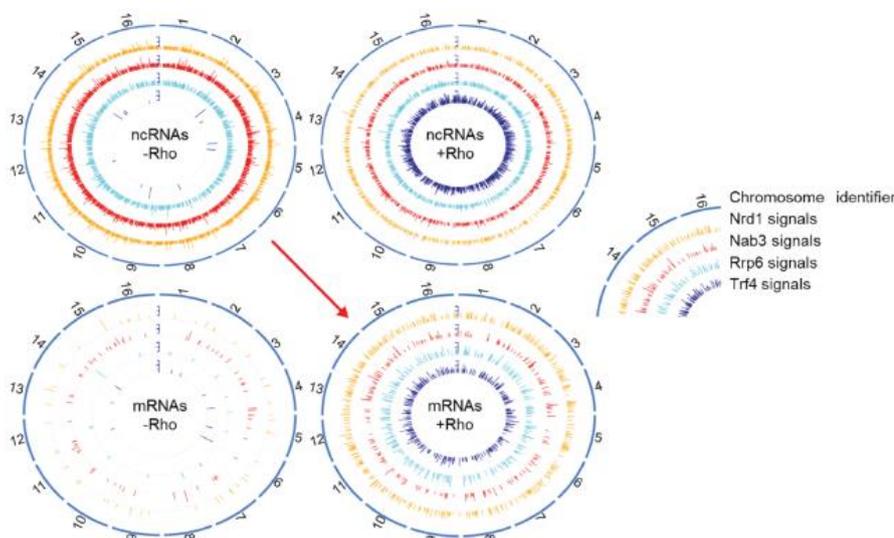


Figure 7: Summary of the observed dynamic landscape of QC components over the yeast genome. Circos plots summarising the ChIP-seq results for the four QC proteins over the 16 yeast chromosomes analysed under -Rho (left plots) and +Rho (right plots) conditions. The two top plots show the ChIP signals detected for the ncRNAs genomic loci (CUTs, SUTs and snoRNAs). The two plots at the bottom show the ChIP signals detected for the genomic features of the Rho-affected mRNAs that were rescued by Rrp6 depletion. The red arrow symbolizes the hijacking of Nrd1 and Nab3 from ncRNAs genomic loci to the Rho-affected mRNA genes upon perturbation of mRNP biogenesis by Rho.

depopulate significantly its presence within ncRNAs loci. The plots show also the large recruitment of Trf4 to both ncRNAs and mRNA genomic features in the presence of Rho, substantiating the role of the TRAMP complex in the targeting of aberrant mRNPs as well as extended ncRNAs [27].

DISCUSSION

Current views of nuclear mRNP QC mechanism suggest that aberrant mRNPs with processing and packaging defects are targeted co-transcriptionally by the surveillance system triggering eventually their degradation by the catalytic activity of the 3'-5' exonuclease Rrp6. However, this assumption is based mostly on the observation that defective mRNPs made in yeast mutants with packaging and export flaws are retained within their site of transcription prior to their elimination [7,28,29]. Thus, it is still unclear how the surveillance system is precisely coordinated both physically and functionally with the transcription machinery to detect the faulty events that may arise at each step of mRNP formation. By implementing an experimental approach in which bacterial Rho factor disturbs mRNP biogenesis in yeast (reviewed in [10]), we conducted a genome-wide examination of mRNPs that are affected by Rho during their formation leading to their targeting by the QC system. This global analysis of the fate of defective mRNPs was accompanied with high resolution mapping of Rrp6-dependent QC components association with chromatin. The study provides an integrated view of how the activation of the surveillance system displaces the QC components from their regular functional locations within termination sites of ncRNAs genomic loci to populate Rho-affected mRNA genes and target co-transcriptionally the defective mRNPs.

Our RNA-seq analysis identified 1,015 mRNAs that were affected by Rho activity leading to a QC-mediated reduction in their steady-state levels. However, we focused our present study on the 491 transcripts whose targeting was found to be mediated by the Rrp6-dependent QC (Figure 2). This group of

transcripts showed a relatively lower level of overall expression in the absence of Rho perturbation as revealed by the comparison between their averaged meta-transcript profile and the averaged profile obtained with the remaining pool of 524 transcripts that we expect to be targeted by the alternative Rat1-dependent QC pathway ([17] and manuscript in preparation). Thus, there is clearly a difference at the transcriptional level that distinguishes the two pools of transcripts.

An interesting feature for the pool of 491 transcripts targeted by the Rrp6-dependent QC is the presence of a comparatively higher reads density within the 5' end region of the averaged meta-transcript profile. This enrichment of transcripts at the 5' end of the genes may indicate the presence of a significant population of paused transcription complexes and/or premature termination as previously revealed by the NET-seq analysis [30] and supported by a more recent study using the RNA crosslinking (CRAC) method [31]. We note, however, that the peak of RNA reads at the 5' end of the averaged profile is probably accentuated in our case by the filtering of the 491 transcripts based on their susceptibility to the Rrp6-dependent QC following Rho action. Obviously, this is an indication for a common feature shared by the selected subset of genes such as a higher level of Ser5 phosphorylation of the CTD of RNAP II within the promoter proximal pausing region that was also seen in the CRAC study [31]. Interestingly, the Nrd1 component of the Nrd1-Nab3-Sen1 (NNS) complex, involved in termination, processing and decay of ncRNAs, binds preferentially to Ser5 phosphorylated CTD of RNAP II via its CTD-interacting domain (CID) [32,33]. Nrd1 interacts also with RNA by recognizing sequence elements in the nascent transcript in association with Nab3 [34,35]. The recent genome-wide studies of RNA-protein crosslinking in yeast have shown that, in addition to their implication in the metabolism of ncRNAs, Nrd1 and Nab3 play also an important role in the regulation of many mRNA coding genes by remodeling gene expression during nutritional shift or during a stress-response process [27,36-38]. For instance, the study by Bresson et al. [27] revealed that upon

glucose withdrawal, growth-related genes showed reduced expression with a concomitant increase of Nab3 binding to the transcripts, suggesting that this subset of mRNAs represent specific targets for the NNS.

Our previous studies on two model genes (*PMA1* and *PGK1*) have shown that the targeting and elimination of Rho-induced aberrant mRNPs is mediated by a stimulation of Nrd1 co-transcriptional recruitment to the genes. The results led us to suggest that this recruitment may promote the attachment of Rrp6-dependent QC components to the nascent transcript affected by Rho action [11,12]. Our present ChIP-seq data provide a strong support to this suggestion by revealing a combined association of Nrd1, Nab3, Rrp6 and Trf4 to chromatin within a large collection of Rho-affected genes (Figure 5 and Figure S6). The ChIP profiles reveal a broad occupancy of Nrd1 that covers the whole gene locus reflecting its association with RNAP II while transcribing along the gene. Alternatively, the ChIP profiles of the Nrd1 partner, Nab3, show rather relatively sharp and discrete peaks within the body of the genes. These profiles agree with previous RNA-protein crosslinking results indicating a clustering of the peaks for Nrd1 but not for Nab3 [36]. The ChIP-seq profiles of Rrp6 and Trf4 show also relatively sharp and discrete peaks within the body of the genes with statistically significant co-localization of the occupancy regions that also overlap with Nab3 peaks (Figure 5 and Figure S6). Together, these chromatin association profiles suggest that the targeting of Rho-induced aberrant mRNPs by the Rrp6-dependent QC relies on the co-transcriptional recruitment and stabilization of a complex made by Nrd1 and Nab3. In our view, Nrd1 might be weakly anchored to the transcription complex through the interaction of its CID with the CTD of RNAP II within the promoter proximal part the gene, a process that could be more common for a subset of genes such as the pool of 491 genes studied here. This primary interaction of Nrd1 is stabilized further by binding in association with Nab3 to specific RNA segments uncovered by Rho action, a fact that mimics a faulty event in mRNP packaging. The importance of Nrd1-Nab3 association in

this early step of the targeting process is also inferred from our previous studies with model genes. In effect, we showed that the targeting of aberrant mRNPs as well as the growth defect phenotype induced by Rho were suppressed when the Nrd1 and Nab3 coupling was flawed by a specific mutation in the Nab3 RNA-binding domain (*nab3-11* allele in [11]) or a deletion of the Nab3-association domain within Nrd1 (*nrd1 Δ NID* in [17] and in Figure S8). Thus, the formation of the Nrd1-Nab3 dimer and its stabilization by RNA binding play a key role in the recognition and targeting of aberrant mRNPs.

The statistically relevant co-localization of the four crosslinked proteins within discrete regions of the chromatin indicates that they are part of the same complex as suggested by others [39,40]. Indeed, previous immunoprecipitation studies have detected physical interaction between Nrd1 and Rrp6, although the precise interacting domains within the proteins are still unknown [19]. More recently, Trf4 was shown to interact physically and functionally with the CID of Nrd1 in a mutually exclusive manner regarding the CTD of transcribing RNAP II [40,41]. Thus, the stabilization of the Nrd1-Nab3 dimer by RNA binding at the first step of the targeting of aberrant mRNPs should liberate the CID from the CTD allowing the tethering of Rrp6 as well as the TRAMP complex via Trf4 interaction. The formation of such appended complex may constitute a second step of the QC process at which the mRNP is labelled as aberrant. Subsequently, upon transcription completion and the freeing of the transcript 3' end, the labeled mRNP would be prevented from export and directed for destruction either by the direct action of Rrp6 alone as was suggested previously [42,43] or following a coupling of Rrp6 with the exosome.

According to the recent and probably more accurate protein copy-number estimations [44], the number of Nrd1 and Nab3 molecules per yeast cell are of similar magnitude (1,916 and 1,723, respectively). Remarkably, our ChIP-seq results reveal the binding landscape of the two NNS components over the yeast genome with nearly all the

molecules localized within ncRNAs genomic loci in the absence of Rho perturbation. In effect, as revealed in Figure 6 A and B, the number of features having at least one peak for Nrd1 (1575 for CUTs +SUTs and 67 for snoRNAs) and the number of features having at least one peak for Nab3 (1577 for CUTs +SUTs and 59 for snoRNAs) are close to the number of molecules per cell. The perturbation of mRNP biogenesis by Rho action appears to disturb this distribution by titrating the two components of the termination and processing complex from their genomic locations. This hijacking of the two NNS components by Rho-affected mRNA genes generates read-through transcripts of ncRNAs as revealed by our RNA-seq results. Also, the Rho-induced large recruitment of Trf4 to the ncRNAs genomic features is a real signature that extended read-through transcripts generated by the local lack of Nrd1 and Nab3 are being discarded following polyadenylation by the TRAMP complex as previously observed following total nuclear depletion of the NNS components [22,23,45,46]. Together, these results provide a global view of the subtle balance of the NNS components within the yeast cell that can be readily perturbed upon a nutritional shift or by stress-induced factors such as bacterial Rho in our experimental approach.

MATERIALS AND METHODS

Yeast strains and plasmids

All the *S. cerevisiae* yeast stains used in this study are derived from the wild-type strain BMA41 (*MATa* or *MATα ade2-1 ura3-1 leu2-3,112 his3-11,15 trp1Δ can1-100*) [47]. They have been previously described [11,12,17] except *NRD1-MYC* and *NAB3-MYC* which were newly created. To construct these new strains expressing C-terminally Myc-tagged Nrd1 or Nab3 proteins, PCR products were amplified with the forward and reverse primers (listed in Table S1) on plasmid template pYM19 [48] and integrated into the *NRD1* locus or the *NAB3* locus by homologous recombination. Selected transformants were analyzed by PCR amplification and sequencing of the PCR

products then immunoblotting was used to verify the presence of the tag.

S. cerevisiae cells were grown according to standard procedures at 25 °C in synthetic complete medium with bases and amino acids omitted as necessary for selection and with glucose (2%) as carbon source. The cell growth was monitored by measuring the optical density at 600 nm. The pCM185 vector (CEN, TRP1) from Euroscarf was used as backbone to create the Doxycycline-regulated (Tet off) Rho expression construct (pCM185-Rho-NLS) which was described previously [49]. Strains transformed by Rho expression vector were grown under repressing conditions (5 µg/mL of Doxycycline) at 25°C in YNB medium supplemented with SC-TRP (Synthetic Complete Supplement Mixture from MP Biomedicals) and glucose as carbon source (final concentration 2%). To allow the expression of Rho, the yeast cells were grown in the same medium but without Doxycycline. Growth was performed during 16 hours incubation at 25°C.

RNA extraction and High-throughput sequencing

Total RNA extraction was performed using the acidic hot phenol/chloroform method as previously described [49]. Quantification of 18S transcript or the randomly picked transcripts by reverse transcription coupled to semi quantitative or real-time quantitative PCR was done using Maxima First strand cDNA synthesis Kit from Thermo Scientific and specific oligonucleotides (described in supplementary Table 2). Libraries preparations were made by I2BC high-throughput sequencing platform (Gif sur Yvette) using Illumina ScriptSeq protocol after RiboZero treatment. All samples were processed in duplicates.

Chromatin Immunoprecipitation (ChIP) and NGS libraries preparation

ChIP was performed as described previously [11,12] with some modifications. Following crosslinking and sonication of cell lysates in 1.2 mL of FA140 buffer (50 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1%

Sodium deoxycholate and 1:100 Protease Inhibitor Cocktail from Promega), the samples were centrifuged at 2500g and the supernatant recovered. 20 μ L of supernatant genomic DNA was used as Input and the remaining was mixed with antibody solution (anti c-MYC sc-40X from Santa Cruz Biotechnology). Tubes were incubated overnight at 4°C, then protein G PLUS-agarose beads (Sigma) were added and rotated for 2 hours at 4°C followed by successive washing steps: 2 times with FA140 lysis buffer, 2 times with FA360 lysis buffer (similar to FA140 but with 360 mM NaCl instead of 140 mM), 1 time with washing buffer and 1 time with TE. DNA was eluted twice at 65°C in elution buffer during 10 minutes. Samples (regardless of Input or IP) were reverse crosslinked at 65°C overnight. Proteinase K digestion of the eluates was performed prior to purification of the genomic DNA samples using the ThermoFisher GeneJET PCR Purification Kit.

ChIP libraries preparation was performed following the NEBNext Ultra II DNA Library Prep instructions from New England Biolabs. All ChIP libraries samples were made in duplicates. The sequencing was performed by I2BC high-throughput sequencing platform (Gif sur Yvette) using Illumina ScriptSeq protocol.

Data processing and Computational Analysis

RNA-seq reads were quality controlled (using FastQC v 0.11.5), the adapters were trimmed using Cutadapt 1.15, and then reads were uniquely aligned to *Saccharomyces cerevisiae* genome V64.1.1 (downloaded from SGD) with bwa 0.7.12 (bwa mem -T 30 -t 6) with default settings except for the minimal quality. Samtools 0.1.19 was used to assess the alignment quality. Reads per genes were counted with featureCounts V 1.5.0-p1 (from the subreads suite) with the strand-specific option (-s 1). The genome annotations were downloaded from SGD and ncRNAs annotations were taken from Xu et al. [20] for SUTs and from Neil et al. [18] for CUTs.

Differential expression analyses were performed under R environment (V 3.4.4) using DESeq2 package. For each genetic background

(WT or *rrp47 Δ*) duplicates from Rho-induced conditions (+Rho) were compared to duplicates from Rho repressed conditions (-Rho). The statistically down-regulated genes (p-value < 0.05 and L2FC < -0.5) from WT background in +Rho conditions (1015 mRNAs) were isolated and their compartments were analyzed in the mutant conditions (491 were rescued). BigWig files were generated as following using bedtools V2.25.0 (genomecov and bigwigCompare), and UCSC tool bedGraphToBigWig. After computing genome coverage (with genomecov), bedgraph outputs were converted to bigwig files using bedGraphToBigWig. Meta-profiles plots were made using bedtools V2.25.0 with computeMatrix (scale-regions mode, with a bin size of 1 base) on BigWig files and plotProfile for graphical output.

ChIP-seq reads were submitted to the same processing steps than RNA-seq reads except for the strand specificity. Peak calling was performed using PePr V1.1.21 [50] which was chosen because it allows performing analysis of duplicates directly and in a comparative mode when needed. Differential peak calling was used for focusing on mRNAs loci enriched in +Rho condition compare to -Rho condition. Classical peak calling was used to analyse the other RNAs biotypes loci. Manipulations on peak caller outputs were done mostly with R. Figures were built with ggplot2 package and Circos plots were done using OmicCircos package. Plots of the distributions of peaks signals by RNA biotypes loci and by proteins were made using beeswarm package.

Statistical Analyses

Spearman correlation test was used to control the correlation between biological replicates. Student t test (bidirectional) was used to compare levels of L2FC between RNA-seq conditions and between -Rho and +Rho ChIP-seq Log2(average signals). NS > 0.05 > * > 0.01 > ** > 0.001 > ***. Jaccard and permutations tests were done using the genomericorr package in order to determine if the distribution of peaks of two “ChIPed” proteins

on the yeast genome overlap significantly (not due to hazard).

Visualization

All snapshots used for direct visualization of reads coverage or ChIP peaks distributions were taken with IGV browser.

ACKNOWLEDGEMENTS

This work was supported by the CNRS recurrent funding and has been sponsored in part by support from Région Centre-Val de Loire through research grants (Anti-Rho-Biotics, and ARD2020 Biomédicaments) to A.R. Rahmouni. K. Moreau was a recipient of a PhD fellowship from the Région Centre-Val de Loire. The genome wide analyses have benefited from the processing workflow and expertise of the High-throughput Sequencing Platform of I2BC, CNRS Gif sur Yvette.

REFERENCES

1. Bentley DL (2014) Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* 15: 163-175.
2. Luna R, Gaillard H, Gonzalez-Aguilera C, Aguilera A (2008) Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. *Chromosoma* 117: 319-331.
3. Doma MK, Parker R (2007) RNA quality control in eukaryotes. *Cell* 131: 660-668.
4. Fasken MB, Corbett AH (2005) Process or perish: quality control in mRNA biogenesis. *Nat Struct Mol Biol* 12: 482-488.
5. Schmid M, Jensen TH (2013) Transcription-associated quality control of mRNP. *Biochim Biophys Acta* 1829: 158-168.
6. Singh P, Saha U, Paira S, Das B (2018) Nuclear mRNA Surveillance Mechanisms: Function and Links to Human Disease. *Journal of Molecular Biology* 430: 1993-2013.
7. Rougemaille M, Gudipati RK, Olesen JR, Thomsen R, Seraphin B, et al. (2007) Dissecting

mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. *EMBO J* 26: 2317-2326.

8. Schmid M, Jensen TH (2008) Quality control of mRNP in the nucleus. *Chromosoma* 117: 419-429.
9. Villa T, Rougemaille M, Libri D (2008) Nuclear quality control of RNA polymerase II ribonucleoproteins in yeast: Tilting the balance to shape the transcriptome. *Biochim Biophys Acta* 1779: 524-531.
10. Mosrin-Huaman C, Hervouet-Coste N, Le Dantec, A., Stuparevic, I., Rahmouni, A.R. (2014) Bacterial Rho helicase: a new tool to dissect mRNP biogenesis and quality control in yeast. *Trends in Cell & Molecular Biology* 9: 79-93.
11. Honorine R, Mosrin-Huaman C, Hervouet-Coste N, Libri D, Rahmouni AR (2011) Nuclear mRNA quality control in yeast is mediated by Nrd1 co-transcriptional recruitment, as revealed by the targeting of Rho-induced aberrant transcripts. *Nucleic Acids Res* 39: 2809-2820.
12. Stuparevic I, Mosrin-Huaman C, Hervouet-Coste N, Remenaric M, Rahmouni AR (2013) Cotranscriptional recruitment of RNA exosome cofactors Rrp47p and Mpp6p and two distinct Trf-Air-Mtr4 polyadenylation (TRAMP) complexes assists the exonuclease Rrp6p in the targeting and degradation of an aberrant messenger ribonucleoprotein particle (mRNP) in yeast. *J Biol Chem* 288: 31816-31829.
13. Feigenbutz M, Garland W, Turner M, Mitchell P (2013) The Exosome Cofactor Rrp47 Is Critical for the Stability and Normal Expression of Its Associated Exoribonuclease Rrp6 in *Saccharomyces cerevisiae*. *PLOS ONE* 8: e80752.
14. Feigenbutz M, Jones R, Besong TM, Harding SE, Mitchell P (2013) Assembly of the yeast exoribonuclease Rrp6 with its associated cofactor Rrp47 occurs in the nucleus and is critical for the controlled expression of Rrp47. *J Biol Chem* 288: 15959-15970.
15. Schuch B, Feigenbutz M, Makino DL, Falk S, Basquin C, et al. (2014) The exosome-binding

factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. The EMBO Journal 33: 2829-2846.

16. Lang WH, Platt T, Reeder RH (1998) Escherichia coli rho factor induces release of yeast RNA polymerase II but not polymerase I or III. Proc Natl Acad Sci U S A 95: 4900-4905.

17. Mosrin-Huaman C, Hervouet-Coste N, Rahmouni AR (2016) Co-transcriptional degradation by the 5'-3' exonuclease Rat1p mediates quality control of HXK1 mRNP biogenesis in *S. cerevisiae*. RNA Biology 13: 582-592.

18. Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, et al. (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457: 1038.

19. Vasiljeva L, Buratowski S (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. Mol Cell 21: 239-248.

20. Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, et al. (2009) Bidirectional promoters generate pervasive transcription in yeast. Nature 457: 1033-1037.

21. Steinhauser S, Kurzawa N, Eils R, Herrmann C (2016) A comprehensive comparison of tools for differential ChIP-seq analysis. Brief Bioinform 17: 953-966.

22. Schaugency P, Merran J, Corden JL (2014) Genome-Wide Mapping of Yeast RNA Polymerase II Termination. PLOS Genetics 10: e1004632.

23. Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, et al. (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell 155: 1075-1087.

24. Steinmetz EJ, Conrad NK, Brow DA, Corden JL (2001) RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. Nature 413: 327-331.

25. Steinmetz EJ, Ng SB, Cloute JP, Brow DA (2006) cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. Mol Cell Biol 26: 2688-2696.

26. Chen X, Poorey K, Carver MN, Müller U, Bekiranov S, et al. (2017) Transcriptomes of six mutants in the Sen1 pathway reveal combinatorial control of transcription termination across the *Saccharomyces cerevisiae* genome. PLOS Genetics 13: e1006863.

27. Bresson S, Tuck A, Staneva D, Tollervey D (2017) Nuclear RNA Decay Pathways Aid Rapid Remodeling of Gene Expression in Yeast. Molecular Cell 65: 787-800.e785.

28. Babour A, Shen Q, Dos-Santos J, Murray S, Gay A, et al. (2016) The Chromatin Remodeler ISW1 Is a Quality Control Factor that Surveys Nuclear mRNP Biogenesis. Cell 167: 1201-1214.e1215.

29. Rougemaille M, Villa T, Gudipati RK, Libri D (2008) mRNA journey to the cytoplasm: attire required. Biol Cell 100: 327-342.

30. Churchman LS, Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469: 368-373.

31. Milligan L, Huynh-Thu VA, Delan-Forino C, Tuck A, Petfalski E, et al. (2016) Strand-specific, high-resolution mapping of modified RNA polymerase II. Molecular Systems Biology 12.

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

33. Vasiljeva L, Kim M, Mutschler H, Buratowski S, 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.

34. Conrad NK, Wilson SM, Steinmetz EJ, Patturajan M, Brow DA, et al. (2000) A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. Genetics 154: 557-571.

35. Steinmetz EJ, Brow DA (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.
36. Creamer TJ, Darby MM, Jamonnak N, Schaughency P, Hao H, et al. (2011) Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet* 7: e1002329.
37. van Nues R, Schweikert G, de Leau E, Selega A, Langford A, et al. (2017) Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress. *Nature Communications* 8: 12.
38. Webb S, Hector RD, Kudla G, Granneman S (2014) PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biology* 15: R8.
39. Fox MJ, Gao H, Smith-Kinnaman WR, Liu Y, Mosley AL (2015) The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway. *PLoS Genet* 11: e1004999.
40. Kim K, Heo D-h, Kim I, Suh J-Y, Kim M (2016) Exosome Cofactors Connect Transcription Termination to RNA Processing by Guiding Terminated Transcripts to the Appropriate Exonuclease within the Nuclear Exosome. *The Journal of Biological Chemistry* 291: 13229-13242.
41. Tudek A, Porrua O, Kabzinski T, Lidschreiber M, Kubicek K, et al. (2014) Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol Cell* 55: 467-481.
42. Callahan KP, Butler JS (2008) Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p. *Nucleic Acids Res* 36: 6645-6655.
43. Callahan KP, Butler JS (2010) TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6. *J Biol Chem* 285: 3540-3547.
44. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Methods* 11: 319.
45. Grzechnik P, Kufel J (2008) Polyadenylation Linked to Transcription Termination Directs the Processing of snoRNA Precursors in Yeast. *Molecular Cell* 32: 247-258.
46. Jamonnak N, Creamer TJ, Darby MM, Schaughency P, Wheelan SJ, et al. (2011) Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA* 17: 2011-2025.
47. Baudin-Baillieu A, Guillemet E, Cullin C, Lacroute F (1997) Construction of a yeast strain deleted for the TRP1 promoter and coding region that enhances the efficiency of the polymerase chain reaction-disruption method. *Yeast* 13: 353-356.
48. Janke C, Magiera MM, Rathfelder N, Taxis C, Reber S, et al. (2004) A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* 21: 947-962.
49. Mosrin-Huaman C, Honorine R, Rahmouni AR (2009) Expression of bacterial Rho factor in yeast identifies new factors involved in the functional interplay between transcription and mRNP biogenesis. *Mol Cell Biol* 29: 4033-4044.
50. Zhang Y, Lin YH, Johnson TD, Rozek LS, Sartor MA (2014) PePr: a peak-calling prioritization pipeline to identify consistent or differential peaks from replicated ChIP-Seq data. *Bioinformatics* 30: 2568-2575.

SUPPLEMENTARY FIGURES

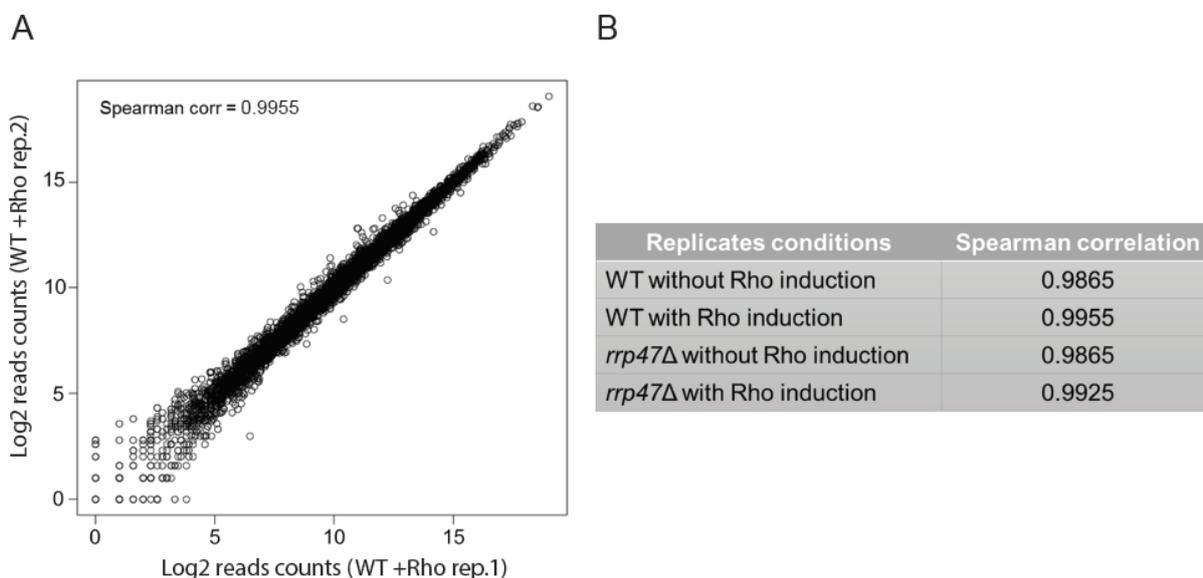


Figure S1: Spearman correlation between biological replicates of RNA-seq datasets. **A)** Scatterplot illustrating the high correlation between two biological replicates regarding the log2 reads count in both replicates for each analysed transcript. **B)** Table indicating the spearman correlation coefficients obtained for each couple of replicates.

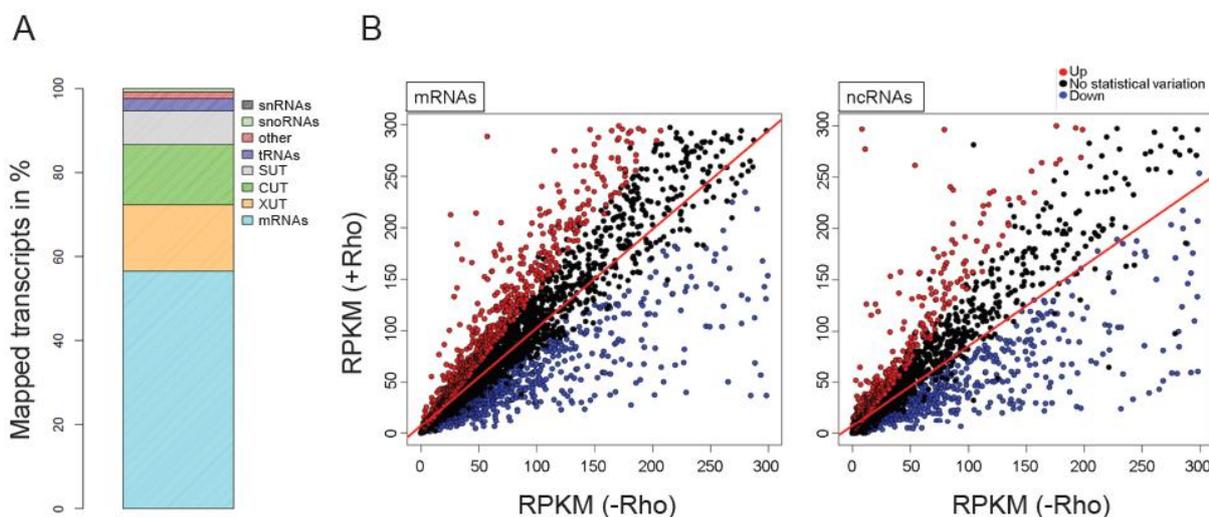


Figure S2: Working parameters assessment. **A)** Stacked barchart giving the percentage of aligned transcripts for each RNA biotype relative to total transcripts mapped by the RNA-seq results. **B)** Scatter plot of reads per kilobase per million comparing the steady-state levels of the transcripts from WT cells under the -Rho conditions versus the same measure under +Rho conditions. The two plots show separately mRNAs and ncRNAs (CUTs, SUTs and snoRNAs). The up-regulated transcripts are presented in red, the down-regulated in blue and the transcripts that show no statistically significant variation are shown in black.

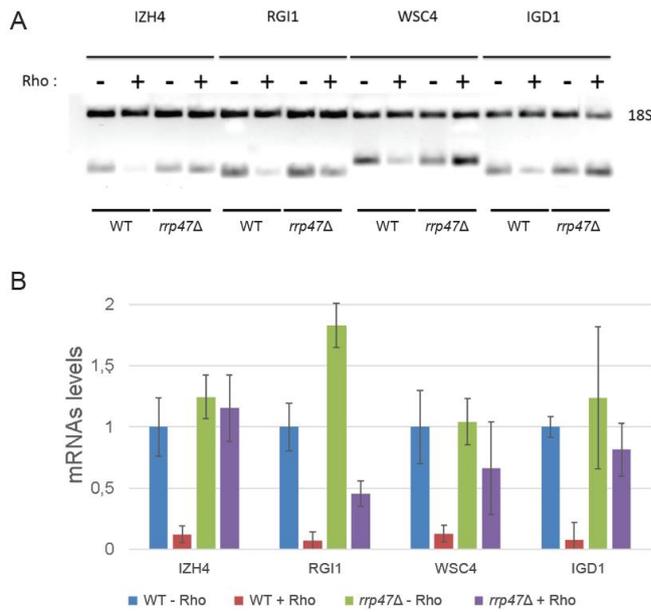


Figure S3: Validation of RNA-seq results by RT-PCR and RT-qPCR analyses of individual transcripts. **A)** Agarose gel showing semi-quantitative RT-PCR analysis performed on total RNA using specific primers to IZH4, RGI1, WSC4, IGD1 mRNAs and 18S rRNA (primers are designed to give PCR products with different sizes). The total RNA analysed was obtained from the two indicated strains and in minus or plus Rho conditions as depicted. **B)** Histogram showing the levels of analysed mRNAs relative to 18S rRNA determined by quantitative RT-PCR and by setting the values to 1 for the WT BMA41 strain in the absence of Rho expression (WT-Rho). Analysed transcripts were picked randomly among the down-regulated subpopulation determined with DESeq2 on RNAseq data.

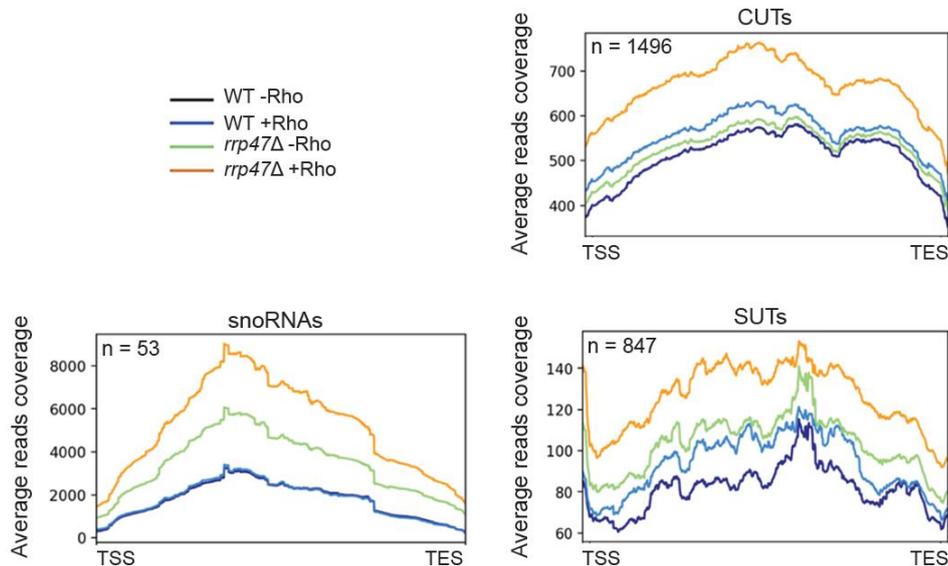


Figure S4: Validation of RNA-seq results by RT-PCR and RT-qPCR analyses of individual transcripts. **A)** Agarose gel showing semi-quantitative RT-PCR analysis performed on total RNA using specific primers to IZH4, RGI1, WSC4, IGD1 mRNAs and 18S rRNA (primers are designed to give PCR products with different sizes). The total RNA analysed was obtained from the two indicated strains and in minus or plus Rho conditions as depicted. **B)** Histogram showing the levels of analysed mRNAs relative to 18S rRNA determined by quantitative RT-PCR and by setting the values to 1 for the WT BMA41 strain in the absence of Rho expression (WT-Rho). Analysed transcripts were picked randomly among the down-regulated subpopulation determined with DESeq2 on RNAseq data.

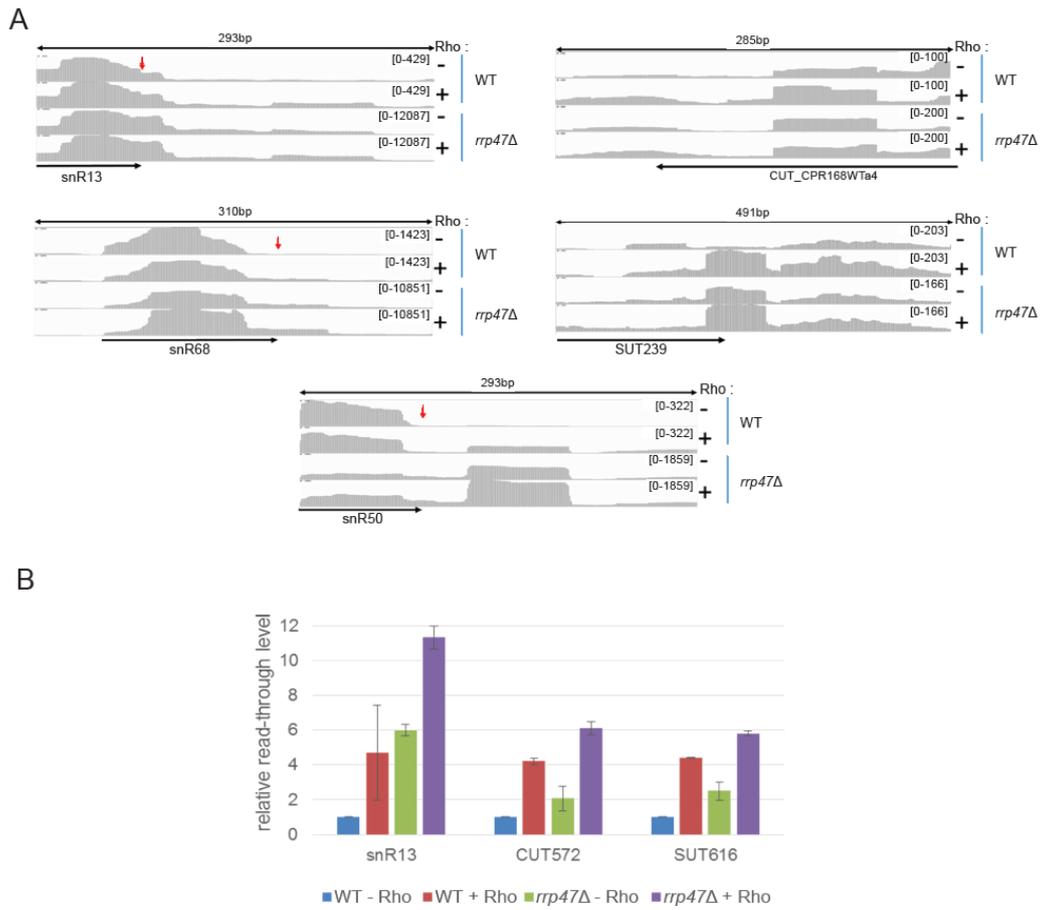


Figure S5: Rho activity in yeast affects the termination and processing of ncRNAs. **A)** Supplemental examples for transcriptional read-through given as in Figure 4 for three other snoRNAs, one CUT and one SUT. **B)** Rho-induced transcriptional read-through for individual ncRNAs as measured by RT-qPCR analyses using primers locations downstream from the annotated termination sites. The data was processed and presented as in Figure S3.

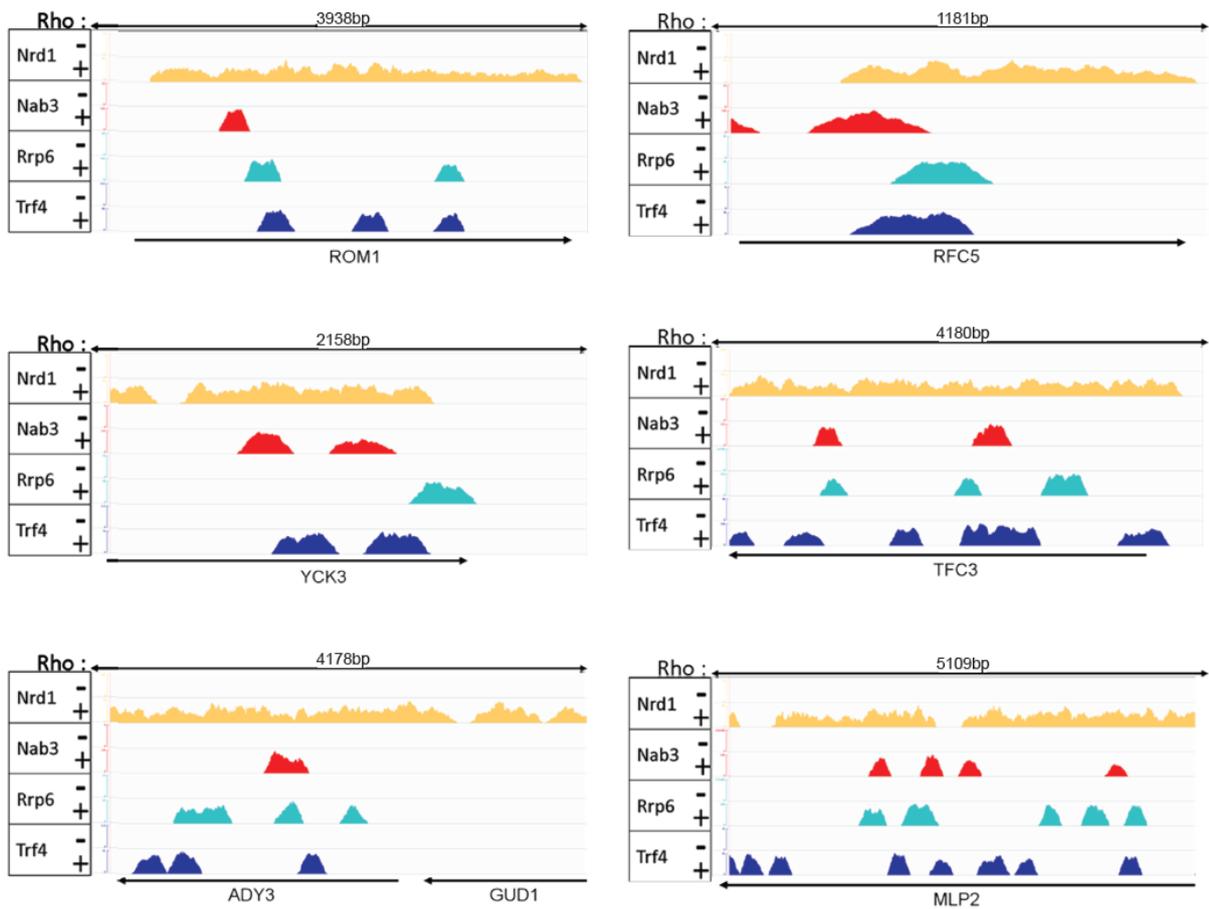


Figure S6: ChIP-seq signals for the four QC proteins over rescued mRNAs loci given as supplemental examples for Figure 5C.

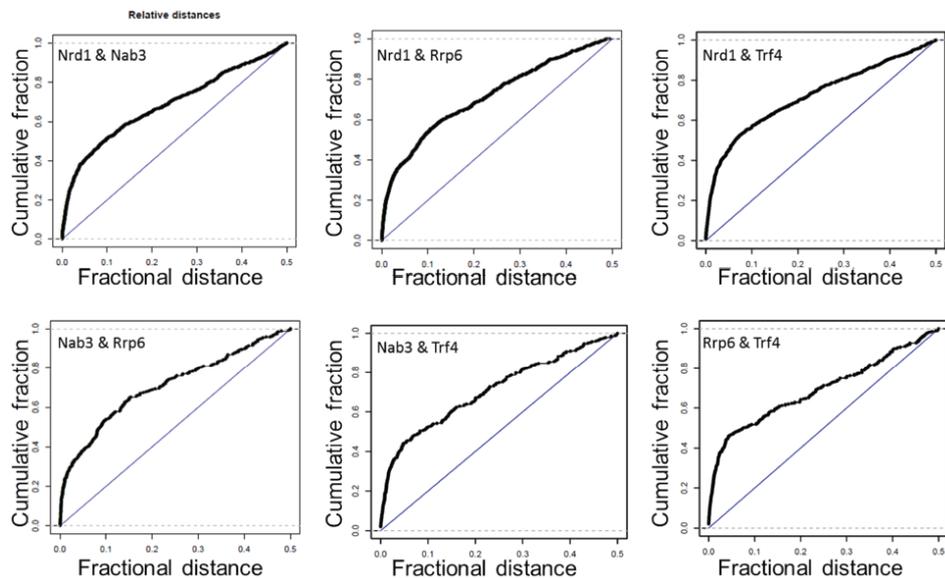


Figure S7: Relative distance test plot used to build figure5-B. This test measures whether two sets of genomic positions are closer together or further apart than expected. The correlation measures used in Figure 5 were extracted from the area between the blue line (uniform, null distribution [0,0.5]) and the black curve (observed distribution of relative distances between two sets of genomic positions).

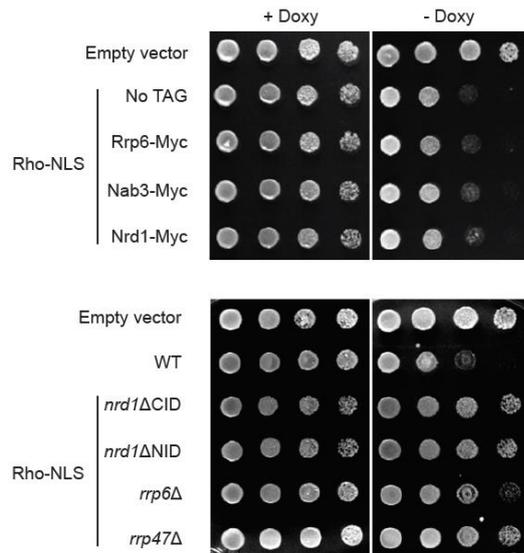


Figure S8: A) Ten-fold serial dilutions of representative yeast cells showing that the strains harboring the tagged proteins grow normally as the wild-type in the absence of Rho and maintain their growth defect sensitivity in the presence of Rho. **B)** Ten-fold serial dilutions of representative yeast cells showing that deletion of or alteration in the QC components (Nrd1, Rrp6 or Rrp47) alleviate the Rho-induced growth defect

Targets	Purpose	Oligonucleotides	
		Forward	Reverse
NAB3	C-terminal tagging	5'-GGCAATAATGTTCAAAGTCTATTAGATAGTTTAGCAAACCTA CAAAAACGTACGCTGCAGGTCGAC	5'-GAATTC AAGTATAATGTACAAGAAATGGAAAAGATTGAAAAAGGG AGTCTAATCGATGAATTCGAGCTCG
NRD1	C-terminal tagging	5'-CAATTGAATCTTTGATGAATATGCTTAACCAACAGCAGCA GCAACAACAACAAAGCCGTACGCTGCAGGTCGAC	5'-AGTTTATGTACTATGAGCAAATAAAGGGTGGAGTAAAGATCTTAAT CGATGAATTCGAGCTCG

Table S1: Oligonucleotides used for endogenous tagging of NAB3 and NRD1

Targets	Purpose	Amplicons	Oligonucleotides	
			Forward	Reverse
18S	Semi-qPCR	707	5'-GAACTTTGGGCCCGGTTG	5'-AGGTTAAGGTCTCGTTCG
	qPCR	197	5'-TTGACGGAAGGGCACCACCA	<i>ibid</i>
IZH4	Semi-qPCR	271	5'-CAAAGCCCAGTCAAATGCGA	5'-TTGTAGTGGTGGTGGGGAAC
	qPCR		<i>ibid</i>	<i>ibid</i>
RGI1	Semi-qPCR	250	5'-AAGAAGTTCGTGCGTCACCT	5'-TTGCGCCTACAGGATGTGTT
	qPCR		<i>ibid</i>	<i>ibid</i>
WSC4	Semi-qPCR	315	5'-AGTTGCCATCGTACAGGGG	5'-CGGTTGTCAATGGGGTCGTA
	qPCR		<i>ibid</i>	<i>ibid</i>
IGD1	Semi-qPCR	255	5'-AGCACGTCACCCACATTTGA	5'-GCTACTCCTACGTTCCGGCTC
	qPCR		<i>ibid</i>	<i>ibid</i>
snR13	qPCR	146	5'-CCTGTCTTTCTGTCGCTTCC	5'-AGTACGTTGGGTTTGCTTG
CUT572	qPCR	103	5'-CCACATTCAACAATAACTCG	5'-TTAGCTTCCATAAAACCAAT
SUT616	qPCR	145	5'-TGTGCAAACGGCAGCTAT	5'-GCCTGACACAATCGTCCCA

Table S2: Oligonucleotides used for qPCR and semi-qPCR

4. Résultats Partie 3 : Dissection des mécanismes menant à la détection des mRNPs aberrantes par le QC.

Concernant ma contribution à l'obtention des résultats de cette section, j'ai réalisé l'ensemble des analyses de données issues de séquençage haut débit. Les CHIP ont été réalisés par le Dr Christine Mosrin-Huaman. Le test en goutte ainsi que les qPCRs ont été réalisés par Mme Aurélia Le Dantec.

4.1. Introduction et mise en contexte

La biogenèse des mRNPs est un procédé se produisant en cours de transcription et regroupant de nombreuses fonctions qui vont de la pose de la coiffe en 5' du transcrit naissant au clivage et à la polyadénylation en 3', en passant par l'épissage et l'empaquetage conduisant ensuite à l'export de la mRNP (Aguilera, 2005; Luna et al., 2008; Mitchell and Parker, 2014). Tous ces mécanismes de maturation et d'empaquetage sont étroitement liés entre eux, ainsi que soumis à une surveillance stricte d'un système qui assure que des ARNs non conformes seront dégradés. Le complexe THO, conservé à travers l'évolution, va se fixer à l'ARNm naissant lors de la phase d'élongation de la transcription pour l'empaqueter et permettre le recrutement d'autres facteurs de maturation et d'export via des interactions protéine-protéine.

Le complexe THO, est composé de 4 protéines interagissant fortement entre elles : Tho2, Hpr1, Mft1 et Thp2 (Chavez et al., 2000). La stabilité du complexe repose sur l'assemblage des 4 sous-unités. La déplétion d'un des composants provoque le désassemblage du complexe et entraîne une dégradation des autres sous-unités. Ainsi, certains mutants sur une sous-unité présentent un phénotype de déficience de croissance à 37°C (Huertas et al., 2006; Libri et al., 2002).

Au cours de la biogenèse de la mRNP, THO interagit avec une multitude de protéines qui peuvent contribuer à la fixation du complexe au pré-mRNA, ou à l'inverse se servent de THO pour se lier au transcrit. Parmi ces interactants se trouvent Sub2, Yra1 et Tex1, qui une fois liées à THO vont former le complexe d'export des ARNm (TREX). Le modèle actuel désigne le complexe THO/TREX comme connectant l'élongation du transcrit à la maturation de

l'extrémité 3' de l'ARN et l'export de ce dernier vers le cytoplasme (Aguilera, 2005). La sous-unité Yra1 du complexe TREX, va pouvoir recruter les protéines Mex67 et Mtr2. Le couple Mex67-Mtr2 formant un récepteur interagissant avec les protéines du pore nucléaire. Le recrutement de Mex67 au site de transcription se fait via son domaine C-terminal d'interaction avec l'ubiquitine (« UBA domain ») qui va se fixer à la sous-unité Hpr1 du complexe THO lorsque cette dernière est ubiquitinylé (Gwizdek et al., 2006). Cette étape est suivie par un transfert de Mex67 sur le transcrite de manière concomitante avec ses partenaires Yra1, Npl3 et Nab2. Ces derniers servant alors d'adaptateurs de liaison à l'ARN pour Mex67. Au moment du passage par le pore nucléaire, Yra1 est ubiquitinylé, provoquant sa dissociation de la mRNP, et permettant le passage à travers le pore. Deux protéines paralogues riches en serine-arginine (« SR rich »), Gbp2 et Hrb1 sont également connues pour se lier au complexe THO et interviennent dans la surveillance de l'épissage de l'ARNm. Lorsque l'ARNm est correctement épissé, Gbp2 et Hrb1 contribuent au recrutement de Mex67 sur le transcrite pour permettre son export. A l'inverse, si le transcrite est mal ou non épissé, les deux protéines participent au recrutement du complexe TRAMP pour préparer ce dernier à la dégradation par l'exosome (Hackmann et al., 2014). Un dernier complexe d'importance doit être mentionné, bien qu'il n'interagisse pas physiquement avec le complexe THO, il possède un rôle primordial dans la même voie d'export des mRNP dont la transcription est dépendante du complexe SAGA : le complexe TREX-2. Ce complexe est composé de cinq sous-unités, les protéines Thp1, Sac3, Sus1, Cdc31 et Sem1. Le complexe TREX-2 interagit physiquement avec les protéines du pore nucléaire ainsi qu'avec Mex67 (Köhler et Hurt, 2007 ; Gonzales-Aguilera et al., 2008 ; Faza et al., 2009). Le complexe TREX-2 est également capable d'influencer l'expression génique via les médiateurs de transcription.

Comme dit précédemment il existe également des protéines et complexes protéiques facilitant le recrutement du complexe THO/TREX sur l'ARN en cours de transcription. Le complexe Prp19, initialement décrit pour son rôle dans l'épissage est également un complexe permettant le recrutement de THO sur l'ARN. Le complexe Prp19 n'est pas présent qu'au niveau des transcrits épissés mais bien sur tous les sites de transcriptions actifs, et semble être une condition nécessaire au recrutement du complexe THO/TREX (Chanarat *et al.*, 2012). Le facteur de transcription Mud2, également décrit comme étant un facteur d'épissage, s'est récemment vu attribué un rôle important pour que le complexe Prp19 se retrouve au niveau

des sites de transcriptions actifs (Minocha *et al.*, 2018). La protéine Mud2 et le complexe Prp19 sont donc des prérequis à la mise en place du complexe THO/TREX sur l'ARN en cours de transcription.

Au-delà de son rôle dans l'empaquetage de l'ARNm de manière co-transcriptionnelle, le complexe THO va également intervenir sur la préservation de l'intégrité de la molécule ribonucléique en prévenant la formation d'hybride ADN:ARN. Ces hybrides sont en partie responsables des troubles de la transcription et de l'instabilité génomique observés dans des souches de levures mutantes pour des composants du complexe THO (Aguilera and Garcia-Muse, (2012); Huertas and Aguilera, 2003).

Certaines études ont amélioré notre connaissance de la connexion entre le complexe THO et la bonne formation de la mRNP. Un mutant ponctuel, *hpr1-101*, qui ne compromet pas la stabilité du complexe THO ou son recrutement à la chromatine mais qui empêche le recrutement de la sous-unité Sub2 du complexe TREX, montre des défauts de transcription comparables à une déplétion entière de la protéine Hpr1. Cependant, il ne déclenche pas de phénomène d'hyper-recombinaison (Huertas et al., 2006) et ces défauts de transcription sont indépendants de la formation de R-loops. Ceci ajoute une importance supplémentaire au complexe THO dans les phénomènes de transcription et de biogenèse des mRNPs (Gomez-Gonzalez and Aguilera, 2009).

Le recrutement du complexe THO sur les ARNs de manière co-transcriptionnelle est possible via l'interaction du complexe avec le CTD phosphorylé de polIII. *In vivo* le recrutement du complexe s'effectue lorsque le CTD de polIII est phosphorylé sur la Ser2 et/ou la Tyr1. Cependant les phosphorylations de la Ser2 et de la Tyr1 semblent interdépendantes *in vivo*, ce qui rend difficile de déterminer par mutagenèse lequel des deux résidus est important dans le recrutement de THO. Cependant, le complexe THO est faiblement recruté sur les gènes *sn/snoRNAs in vivo* où le taux de phospho-Ser2 est faible alors que le taux de phospho-Ser7 du CTD est élevé. Ceci suggère fortement que le recrutement de THO nécessite la phosphorylation du CTD en Ser2 (Meinel et al., 2013). De plus, la présence du complexe THO augmente de manière croissante de l'extrémité 5' vers l'extrémité 3' des gènes, à la manière de la marque phospho-Ser2 du CTD (Katahira, 2015; Meinel et al., 2013). Cette caractéristique d'augmentation de la présence du complexe THO de l'extrémité 5' vers l'extrémité 3' des gènes transcrits le distingue des autres facteurs de transcription recrutés par le CTD de polIII

lorsque ce dernier est phosphorylé sur la Ser2 (Abruzzi et al., 2004; Gomez-Gonzalez et al.). Un autre trait caractéristique du recrutement du complexe THO repose sur le domaine C-terminale désordonné de la sous-unité Tho2. En effet, ce domaine de Tho2 a été montré comme capable de se lier aux acides nucléiques (ADN et/ou ARN), et il va faciliter le recrutement du complexe au niveau de la chromatine des gènes transcrits (Pena et al., 2012). Il existe donc deux voies de recrutement du complexe THO qui demandent de plus amples études pour pouvoir former un modèle unifié de recrutement.

Le procédé de surveillance de la biogenèse des mRNPs a également été caractérisé dans des mutants du complexe THO/TREX (Rougemaille et al., 2007; Saguez et al., 2008). Il a été montré que des mutants des composants du complexe THO conduisaient à la dégradation des transcrits mal empaquetés par le système de surveillance et de dégradation (Saguez et al., 2008). Il faut également garder à l'esprit que le complexe THO doit se fixer à l'ARN naissant pendant l'élongation tout en augmentant sa présence de manière croissante de l'extrémité 5' vers l'extrémité 3' du gène (Meinel et al., 2013) pour que l'export puisse se faire. Le complexe THO est donc indispensable à la bonne formation de la mRNP, ainsi qu'à sa protection contre la dégradation. Il est également un prérequis pour son export correct vers le cytoplasme.

Nous avons pu voir dans les résultats de la Partie 2 comment le système de contrôle qualité et dégradation passant par Rrp6, est guidé vers les ARN messagers rendus aberrants par l'action de Rho. Il semble que l'action de Rrp6 soit facilitée par le basculement massif de Nrd1 des loci d'ARNs non codants vers les loci d'ARNs messagers aberrants. Ce déplacement de Nrd1 s'accompagne d'un recrutement également accentué de Nab3 et Trf4 sur ces mêmes loci. Bien que cet événement semble être une des clés de voute du système de détection des malformations de la mRNP, d'autres protéines ou complexes protéiques présents dès le début de l'élongation de la transcription et intervenant dans la mise en place de la particule ribonucléoprotéique, comme le complexe THO, pourraient avoir le rôle de déclencheur/détecteur de l'aberrance.

Afin d'investiguer leur implication, nous avons suivis par CHIP-seq certains composants du complexe THO dans notre modèle de perturbation de la biogenèse des mRNPs afin de voir si l'action de Rho interférait avec la bonne mise en place du complexe sur ses cibles. En effet, si le complexe THO est déplacé des ARNs en cours de transcription, cela pourrait expliquer en partie, les résultats que nous obtenons avec l'utilisation du perturbateur Rho.

Mes résultats ont montré en effet, qu'une partie du complexe THO n'est plus présente sur les 1015 loci transcrits en ARNm (voir Résultats partie 2 pour plus de détails sur ces ARNs) qui sont rendus aberrants par Rho. De manière surprenante, la sous-unité Tho2 présente un profil de recrutement démultiplié au niveau de la chromatine des ARNs analysés. Ces résultats indiqueraient que Tho2, indépendamment du complexe THO, va se lier sur ces transcrits aberrants au cours de leur transcription et potentiellement participer à leur surveillance. Afin de renforcer cette hypothèse, une analyse en parallèle des résultats obtenus en CHIP-seq de Rrp6 a été effectuée. L'hypothèse forte qui en découle est qu'il semble exister un lien plus ou moins direct entre Tho2 et Rrp6 au niveau des loci des transcrits aberrants. Pour approfondir l'étude de ce point, des expériences de CHIP-seq de Rrp6 ont été effectuées à partir de souches dépourvues de Mft1 ou Tho2 ou exprimant une protéine Tho2 plus ou moins tronquée sur son domaine C-terminal (proposé comme étant responsable de l'interaction de Tho2 avec les acides nucléiques). Mes résultats montrent que le guidage de Rrp6 sur ces cibles à dégrader passe par la capacité de Tho2 à se lier aux acides nucléiques.

4.2. Matériel et méthodes

Les expériences de CHIP et le traitement des données CHIP-seq obtenus ont été réalisées en suivant les mêmes protocoles que ceux explicités dans la publication présentée dans la partie 2 des Résultats. Les souches mutantes utilisées ont également été obtenues comme décrit dans cette section. Ci-dessous, un tableau des souches utilisées pour les différentes expériences de CHIP-seq et RT-qPCR est présenté. Dans la suite du texte, les protéines immuno-précipitées ont été fusionnées avec un tag « MYC » en C-terminal. Les résultats issus du traitement des données de CHIP-seq sont obtenus en utilisant le programme de « peak calling » PePr dans son mode comparatif. Les pics qui ressortent dans ce type d'analyse sont ceux qui sont enrichies dans une des 2 conditions comparées.

Souches	Fond génétique	Source
BMA41	<i>MATa or MATa ade2-1 ura3-1 leu2-3,112 his3-11,15 trp1D can1-100</i>	Baudin-Baillieu et al., 1997
THO2-MYC	Comme BMA41 avec <i>Tho2-MYC::HIS3</i>	Cette étude
HPR1-MYC	Comme BMA41 avec <i>Hpr1-MYC::HIS3</i>	Cette étude
MFT1-MYC	Comme BMA41 avec <i>Mft1-MYC::HIS3</i>	Cette étude
<i>mft1Δ</i>	Comme BMA41 avec <i>mft1Δ::KAN</i>	Cette étude
<i>tho2Δ</i>	Comme BMA41 avec <i>tho2Δ::KAN</i>	Cette étude
<i>rrp6Δ</i>	Comme BMA41 avec <i>rrp6Δ::KAN</i>	Mosrin-Huaman et al., 2009
<i>mft1Δ</i> THO2-MYC	Comme <i>mft1Δ</i> avec <i>Tho2-MYC::HIS3</i>	Cette étude
<i>mft1Δ</i> HPR1-MYC	Comme <i>mft1Δ</i> avec <i>Hpr1-MYC::HIS3</i>	Cette étude
<i>mft1Δ</i> THP2-MYC	Comme <i>mft1Δ</i> avec <i>Thp2-MYC::HIS3</i>	Cette étude
<i>mft1Δ</i> RRP6-MYC	Comme <i>mft1Δ</i> avec <i>Rrp6-MYC::HIS3</i>	Cette étude
<i>tho2Δ</i> MFT1-MYC	Comme <i>tho2Δ</i> avec <i>Mft1-MYC::HIS3</i>	Cette étude
<i>tho2Δ</i> HPR1-MYC	Comme <i>tho2Δ</i> avec <i>Hpr1-MYC::NAT</i>	Cette étude
<i>tho2Δ</i> THP2-MYC	Comme <i>tho2Δ</i> avec <i>Thp2-MYC::NAT</i>	Cette étude
<i>tho2Δ</i> RRP6-MYC	Comme <i>tho2Δ</i> avec <i>Rrp6-MYC::KAN</i>	Cette étude
<i>tho2ΔCter14</i>	Comme BMA41 avec <i>tho2ΔCter (1-1408)-MYC::KAN</i>	Cette étude
<i>tho2ΔCter12</i>	Comme BMA41 avec <i>tho2ΔCter (1-1271)-MYC::KAN</i>	Cette étude
<i>tho2ΔCter14</i> RRP6-MYC	Comme BMA41 avec <i>tho2ΔCter (1-1408)-HA3x::hph Rrp6-MYC::KAN</i>	Cette étude
<i>tho2ΔCter12</i> RRP6-MYC	Comme BMA41 avec <i>tho2ΔCter (1-1271)-HA3x::hph Rrp6-MYC::KAN</i>	Cette étude

Tableau 4.2.1. Présentations des différentes souches utilisées dans cette section de résultats.

4.3. Résultats

Fort des résultats précédemment obtenus par RNA-seq, la première étape de l'analyse a consisté à définir comment se comporte le complexe vis-à-vis de notre système de perturbation au niveau des ARNm dont la biogenèse est compromise. La figure 4.3.1-A présente l'analyse de corrélation de Pearson entre les pics de liaison à la chromatine de chaque protéine du complexe THO étudiée avec ou sans perturbation par Rho. Les résultats indiquent qu'au niveau de la chromatine, les 3 protéines du complexe THO sont fortement corrélées du point de vue de leur présence sur les mêmes loci en absence de Rho. En revanche, lorsque la biogenèse des mRNPs est perturbée par Rho, la co-localisation de ses 3 protéines disparaît (Figure 4.3.1-A). Cette observation est confirmée par une inspection visuelle du génome qui montre que les pics de CHIP qui se co-localisent en condition normale ne co-localisent plus du tout en condition de biogenèse perturbée (Figure 4.3.1-B).

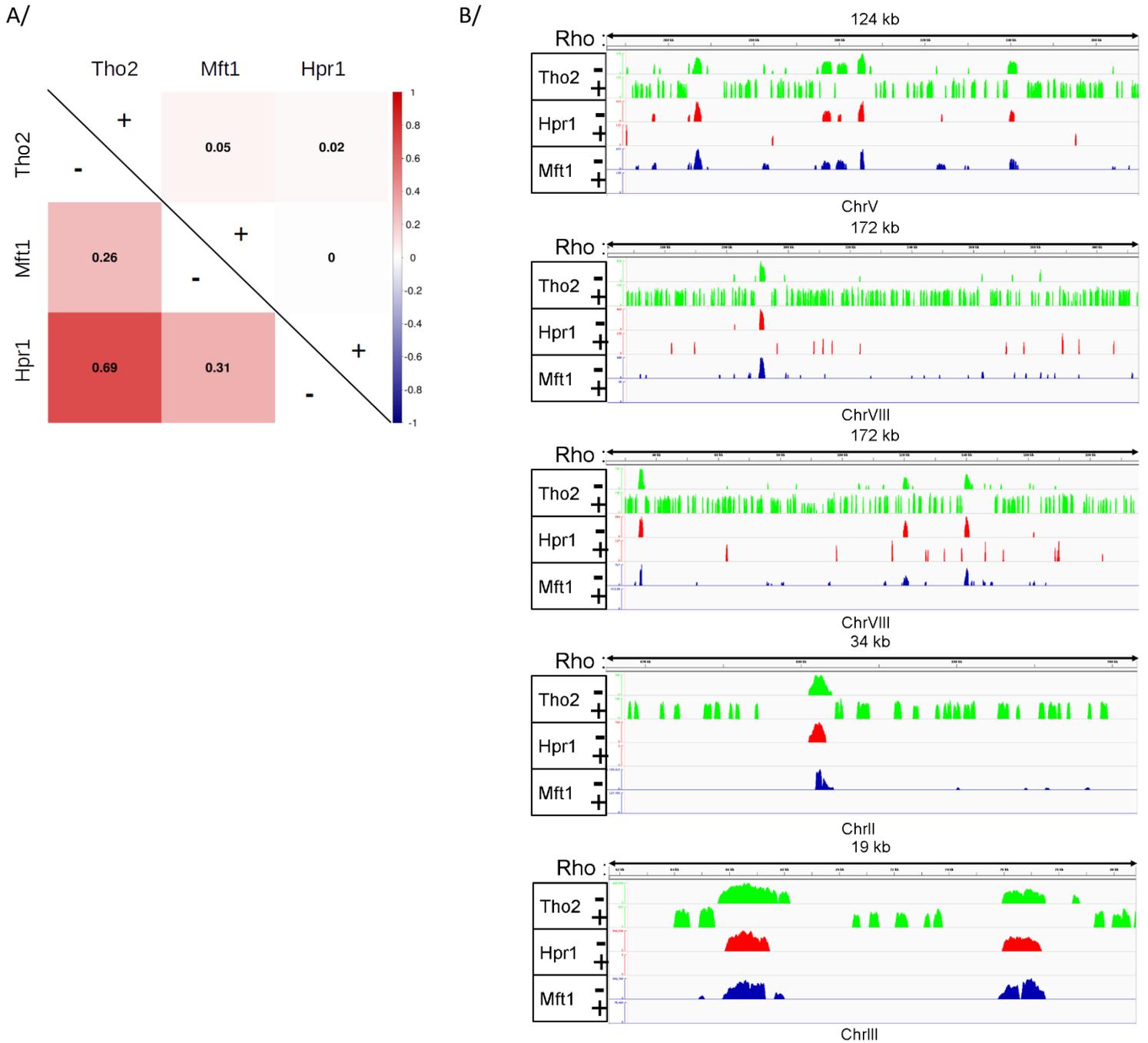


Figure 4.3.1. Corrélation entre les pics de ChIP-seq de Tho2, Mft1 et Hpr1 avant et après induction de Rho. A/ Corrélation de Pearson entre les 3 protéines d'intérêt. La partie inférieure gauche du corrélogramme nous montre les résultats obtenus sans perturbation par Rho et le panneau supérieur nous montre les résultats avec perturbation par Rho. B/ Visualisation des résultats de ChIP-seq en absence ou présence de Rho via IGV à cinq endroits différents du génome de *S. cerevisiae*. Mft1 est en bleue, Hpr1 en rouge et Tho2 en vert.

L'inspection visuelle des profils de recrutement lorsque Rho est induit et perturbe la biogenèse des mRNPs, révèle des résultats surprenants et originaux. En effet, le profil de répartition des pics observés en condition normale est radicalement différent de celui en condition de perturbation. Rho, de par son activité peut déloger le complexe THO, ou du moins une partie, de l'ARN qu'il empaquète. La donnée importante dans ce cas de figure est que la protéine Tho2 est la seule protéine du complexe à être retrouvée sur la chromatine en condition de perturbation. De plus, son profil de recrutement est démultiplié (Figure 4.3.1-B).

Ce résultat est confirmé lorsque l'on s'intéresse au profil de recrutement des 3 protéines d'intérêt au niveau des 491 loci d'ARNm impactés par Rho et sauvés en déplétion de la machinerie de dégradation dépendante de Rrp6 (pour plus d'informations sur ces ARNs spécifiques, se référer à la section Résultat Partie 2). La figure 4.3.2. renseigne sur le nombre ainsi que l'intensité des pics de CHIP de chacune des protéines du complexe THO étudiées en absence et en présence de Rho au niveau de ces 491 loci spécifiques. Ce qui est observable, de manière surprenante, est la tendance inverse observée entre Tho2 et les deux autres protéines, Mft1 et Hpr1. Alors que les deux dernières semblent être évincées de leurs sites de fixation, Tho2 présente un profil de sur-recrutement lorsque les mRNPs correspondant à ces loci voient leurs biogenèses perturbées. Bien que le profil de recrutement de Tho2 sur la chromatine semble être général en présence de Rho, nous nous focaliserons par la suite uniquement sur les loci correspondant aux ARNm dont la biogenèse est perturbée par Rho.

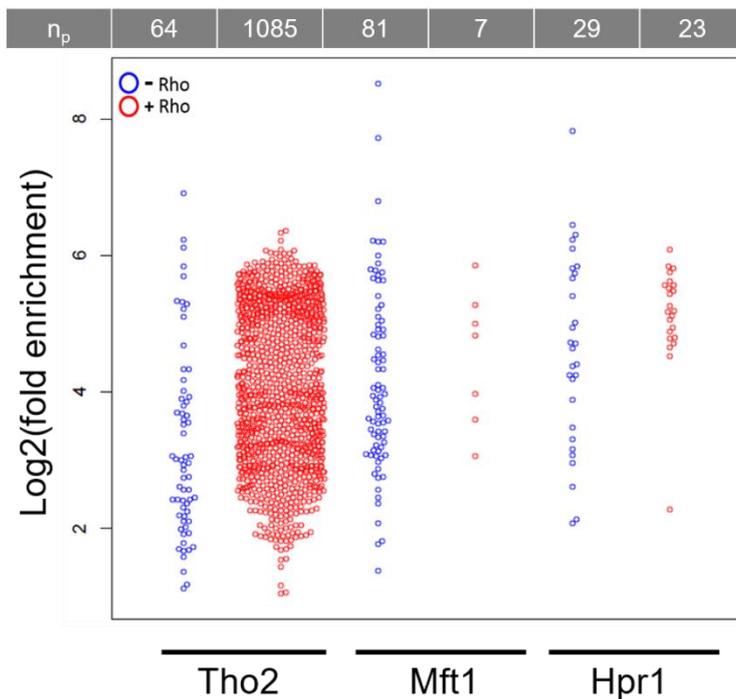


Figure 4.3.2. Graphique de distribution des pics des trois protéines du complexe THO en condition normale (bleu) et en condition de perturbation de la biogenèse des mRNPs (rouge) au niveau des 491 loci dont les ARNm sont sauvés par la délétion de RRP47 (i.e. déplétion de Rrp6). Le tableau au-dessus du graphique indique le nombre de pic pour chaque protéine, dans chaque condition détectés sur l'ensemble des 491 loci.

Ce sur-recrutement intense de la protéine Tho2, ainsi que le comportement inverse des autres composantes du complexe, permet d'envisager deux hypothèses possibles : 1/ Rho va désassembler le complexe THO de l'ARN, et la sous-unité Tho2 libre va se retrouver recrutée en grande quantité au niveau des loci présentant des ARNs non conformes. 2/ L'apparition d'ARNs aberrants est un signal d'alarme qui va déclencher un procédé de marquage des ARNs par Tho2, en dehors de son complexe canonique, et contribuer au contrôle de la qualité de ces derniers. Ces deux hypothèses ne sont pas mutuellement exclusives. De surcroît, la protéine Tho2 est la seule du complexe THO possédant un domaine de liaison aux acides nucléiques. Cette dernière information renforce l'hypothèse selon laquelle Tho2 pourrait jouer un rôle direct sur les transcrits, ou sur la chromatine, en dehors de son complexe canonique.

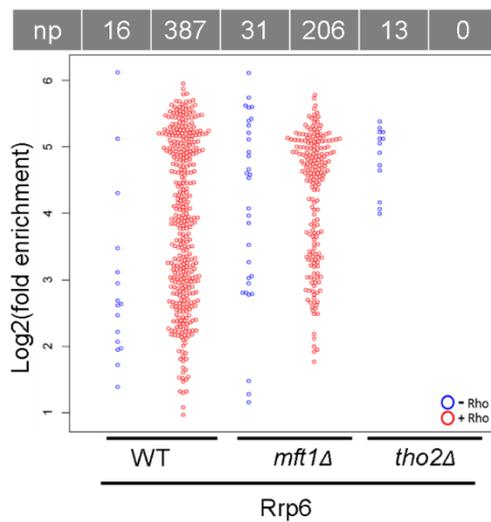
Le modèle qui gravite autour de ces propositions est que la protéine Tho2, de par son sur-recrutement, peut contribuer au marquage des transcrits aberrants et ainsi participer aux mécanismes de contrôle de la qualité des ARNs.

Afin de tester ce modèle, un comparatif des signaux de CHIP de Rrp6 dans un contexte WT et dans différents mutants du complexe THO a été réalisé. Si Tho2 intervient dans des

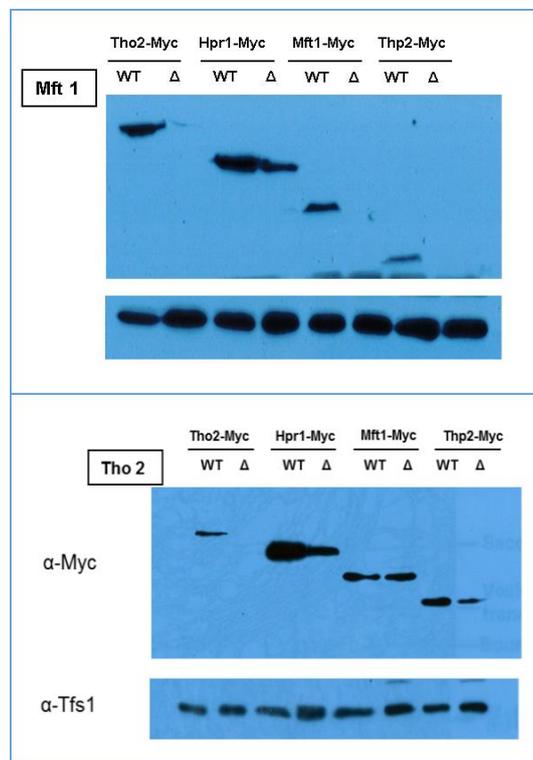
procédés liés au QC des mRNPs aberrantes, alors nous pouvons parier que la déplétion de cette protéine aura un impact sur le profil de recrutement de Rrp6, dont nous avons montré un sur-recrutement au niveau des même loci dans la publication précédente. La figure 4.3.3. illustre ces résultats.

De prime abord, le profil de recrutement de Rrp6 présente des différences notables en condition de perturbation de la biogenèse des mRNPs comparé à la condition normale. En effet, le profil de recrutement de Rrp6 sur les loci d'ARNm aberrants est démultiplié en condition de perturbation par Rho. En comparant ce résultat avec celui obtenu dans une souche dépourvue de Mft1, peu de différences sont observables. Il semblerait que le sur-recrutement Rho-dépendant observé en WT soit moins important dans cette souche. En comparant maintenant la condition WT et *tho2Δ* du point de vue du recrutement de Rrp6, les résultats montrent une disparition totale de signal dans la condition de perturbation (Figure 4.3.3-A). Il est donc clair que le ciblage des ARNs aberrants par Rrp6 passe par la présence de Tho2 sur ces derniers. La diminution relative de la présence de Rrp6 sur ses cibles dans la souche *mft1Δ* peut être expliquée comme un effet secondaire résultant de la déstabilisation de Tho2 en absence de Mft1. Les résultats de Western blots (Figure 4.3.3-B) illustrent le phénomène d'interdépendance des protéines du complexe THO. En effet, lorsque la sous-unité Mft1 est déplétée, on observe une déstabilisation des autres protéines du complexe. De même lorsque c'est la protéine Tho2 qui est déplétée, le reste des protéines du complexe sont retrouvées en moins grande quantité. Par voie de conséquence, si le recrutement de Rrp6 sur ces cibles passe par Tho2, la déstabilisation partielle de cette dernière dans une souche *mft1Δ* explique les observations faites pour Rrp6. La diminution du nombre et de l'intensité de signal des pics Rrp6 est bien liée à une dépendance envers Tho2 et non pas à une diminution de la quantité de la protéine dans les souches mutantes, comme attesté par le western blot figure 4.3.3-C. De surcroit, cette dépendance semble être marquée par le fait qu'une grande proportion de loci d'ARNm étudiés voit la présence de pics Rrp6 (85%) et Tho2 (53%) simultanément (figure 4.3.3-D). Et plus de 50% des pics Rrp6 chevauchent les coordonnées de pics Tho2 au niveau de ces loci (figure 4.3.3-E).

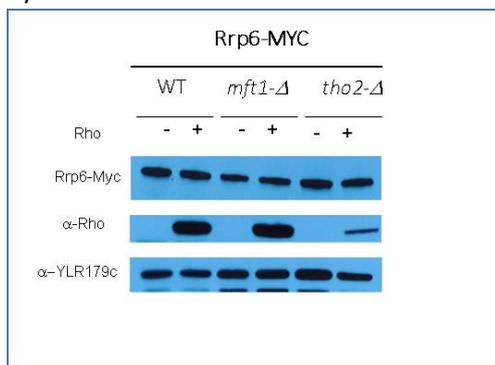
A/



B/



C/



D/



E/

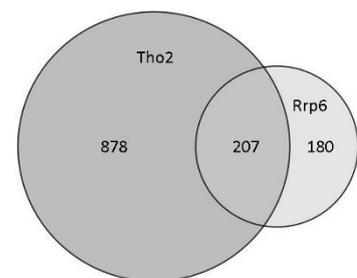


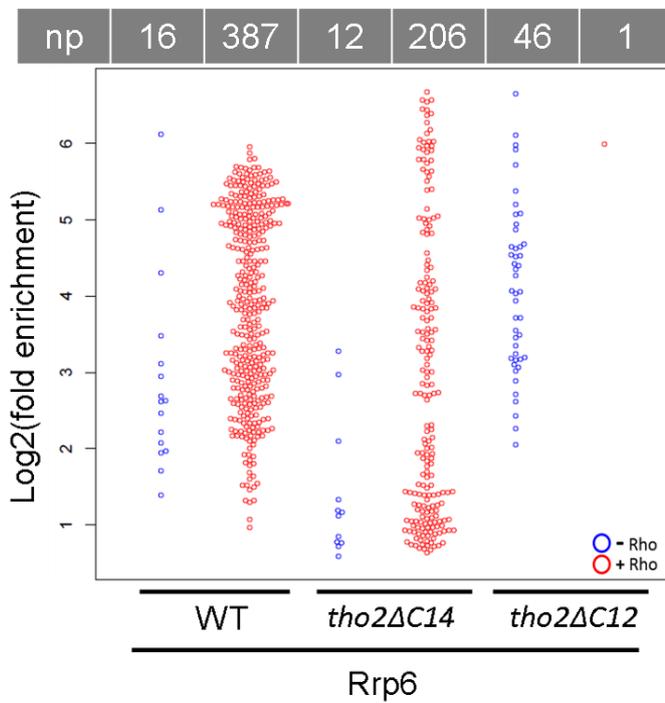
Figure 4.3.3 Implication des protéines du complexe THO dans le recrutement de Rrp6 sur les loci des ARNs aberrants. A/ Graphique de distribution des signaux de ChIP-seq observés pour Rrp6 dans différents mutants du complexe THO. Le code couleurs est le même que précédemment, ainsi que la signification du tableau au-dessus du graphique. B/ Western blot des composantes du complexe THO dans des conditions WT ou mutante pour certaines sous-unités. C/ Western blot de Rrp6 dans différents mutants du complexe THO. D/ Digramme de Venn représentant le nombre de loci possédant à la fois au moins un signal Tho2 et un signal Rrp6. E/ Diagramme de Venn représentant le nombre de pics Tho2 et Rrp6 dont les coordonnées se chevauchent.

La dépendance vis-à-vis de Tho2 du recrutement de Rrp6 aux loci d'ARNs aberrants a donc été établie. Pour aller plus loin dans la mécanistique de cette dépendance, nous nous sommes demandé si Tho2 devait se lier aux acides nucléiques pour permettre à Rrp6 d'être recruté au niveau des loci d'ARNm aberrants.

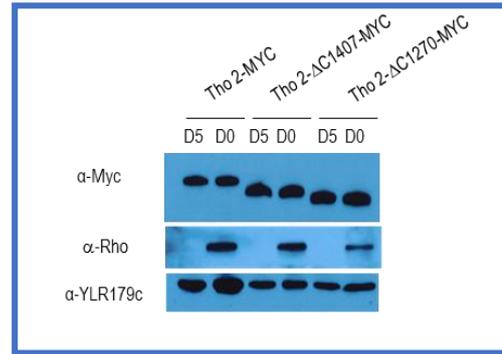
Afin de tester cette hypothèse, une série d'expériences de ChIP de Rrp6 ont été réalisées dans des souches où le domaine C-terminal de Tho2 est plus ou moins délété (Figure 4.3.4-A). Ces ChIP ont été séquencés et les résultats sont exposés en figure 4.3.4. Les délétions

du domaine C-terminal sont nommées *tho2ΔCter14* pour la forme la moins délétée et *tho2ΔCter12* pour la délétion totale du domaine C-terminal (Figure 4.3.4-B). L'analyse des résultats qui découlent de ces expériences de CHIP-seq de Rrp6 nous indique qu'en absence d'une partie du domaine C-terminal de Tho2 (*tho2ΔCter14*), le profil de recrutement de Rrp6 semble corrompu. En effet, la variation d'enrichissement de la protéine, en condition d'induction de Rho, est différente comparée à la condition WT (Figure 4.3.4-A). Nous pouvons ainsi constater des valeurs extrêmes, que ce soit dans le sens d'une diminution de l'intensité de signal ou bien à l'inverse, dans le sens d'une amplification. De surcroît le nombre de pics de Rrp6 semble être diminué, de l'ordre de ce qui était observé dans la souche *mft1Δ*. Concernant les résultats obtenus dans le cas de la pleine délétion du domaine C-terminal de Tho2 (*tho2ΔCter12*), Rrp6 semble complètement perdre sa capacité à se lier à ses cibles. Ces résultats sont comparables à ce que nous avons précédemment observé dans la souche dépourvue de la protéine Tho2 entière. Ces résultats nous indiquent que le recrutement de Rrp6, au niveau des loci correspondant aux mRNPs affectées par Rho, est intimement lié à la faculté de la protéine Tho2 à se lier aux acides nucléiques. Nous montrons donc ici, que la seule délétion du domaine C-terminal de la protéine Tho2 suffit à perturber l'accès de Rrp6 aux loci de ses ARNs cibles. L'influence de la délétion d'une portion, ou de la totalité du domaine C-terminal de Tho2, sur la stabilité de la protéine a été suivie par western blot (Figure 4.3.4-C). Ce dernier ne montre pas de variation de quantité de Tho2 ou ses variantes. De même, il ne semble pas y avoir d'effet de ces délétions de domaines sur la stabilité/quantité de protéine Rrp6 (Figure 4.3.4-D). De fait, la disparition du signal Rrp6 en +Rho, ne provient pas d'une disparition de la protéine, mais bien d'une incapacité de cette dernière à se lier à ses cibles. Des expériences de CHIP sur la protéine Tho2 plus ou moins tronquée de son domaine C-terminal ont été réalisées en parallèle mais les échantillons obtenus après immunoprécipitation ne nous permettaient pas d'amplifier de l'ADN par qPCR. Cette donnée nous indique donc que la protéine Tho2 perd bien sa capacité à se lier aux acides nucléiques lorsque le domaine C-terminal est manquant.

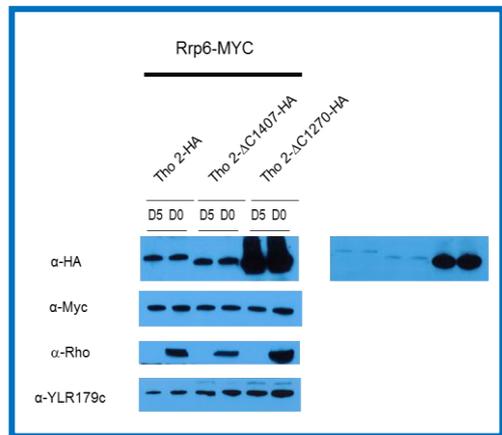
A/



C/



D/



B/

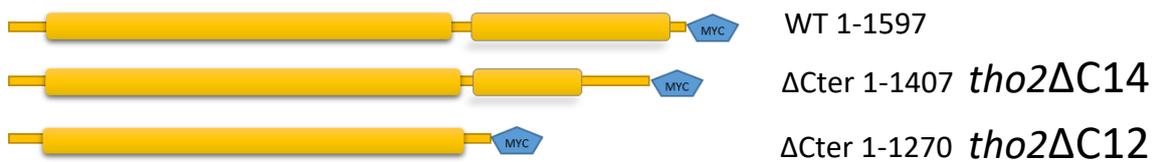


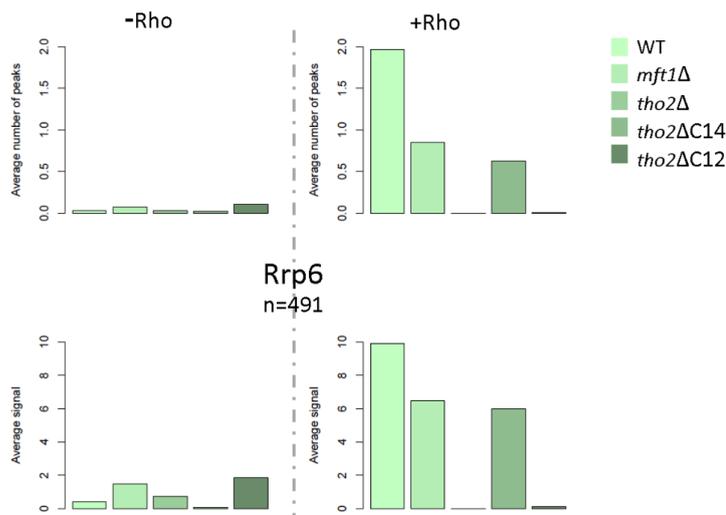
Figure 4.3.4. A/ Graphique de distribution des signaux de ChIP-seq observés pour Rrp6 dans différents mutants du complexe THO. Le code couleurs est le même que précédemment, ainsi que la signification du tableau au-dessus du graphique. B/ Schéma de la protéine Tho2 WT et des variantes tronquées plus ou moins sur le domaine C-terminal taggué HA ou MYC. C/ Western blot montrant que la délétion d'une partie ou de l'intégralité du domaine C-terminal de Tho2 n'a pas d'impact sur la quantité de protéine dans la cellule (que ce soit en condition d'induction de Rho (D0) ou non (D5)). D/ Western blot montrant que la quantité de protéine Rrp6 n'est pas impactée par la délétion du domaine C-terminal de Tho2.

Un résumé du nombre de pics ainsi que de l'intensité moyenne des signaux Rrp6 dans les différentes souches mutantes analysées, permet de voir la diminution croissante de l'affinité de Rrp6 pour ses ARNs cibles, en parallèle de la délétion croissante du domaine C-terminal de Tho2 (Figure 4.3.5-A). La moyenne du nombre de pics Rrp6 par locus passe de 2

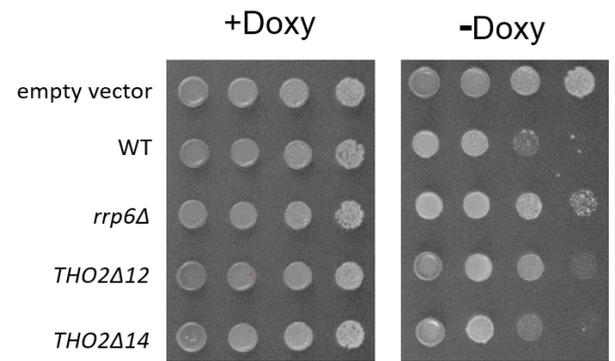
en WT à ~ 1 dans la souche *tho2ΔC14* (comparable au mutant *mft1Δ*) et tombe à 0 pour le mutant avec la délétion totale du domaine C-terminal (comparable au mutant *tho2Δ*). Comme explicité en introduction et illustré dans les deux premières parties de résultats de ce manuscrit, les levures présentent un phénotype de croissance ralentie lorsque l'expression de Rho est induite. Ce retard de croissance est la résultante de l'action de Rho sur la biogenèse des mRNPs. En effet, la dégradation des ARNs rendus défectueux par Rho, provoque un manque de certains facteurs clés pour la vie et la pousse de la cellule. Ce phénotype est atténué lorsque l'on déplete la machinerie de QC et dégradation tel Rrp6, ceci s'explique par le fait que les ARNs normalement dégradés ne le sont plus, et peuvent ainsi jouer leurs rôles (Honorine et al., 2011; Mosrin-Huaman et al., 2009; Stuparevic et al., 2013). La diminution de fixation de Rrp6 au niveau de ses cibles illustrées précédemment, a été testée d'un point de vue phénotypique (Figure 4.3.5-B) via un test de croissance en goutte. Ce test donne phénotypiquement des résultats qui corroborent les différentes analyses précédentes faites au niveau de la chromatine. Effectivement, l'impossibilité pour Rrp6 à se lier au niveau de ses cibles permet à ces dernières d'échapper à la dégradation, rendant la croissance cellulaire de nouveau possible. La figure illustre également bien les différences relevées plus tôt quant au recrutement de Rrp6 dans le mutant avec la délétion partielle du domaine C-terminal de Tho2. Plus la délétion du domaine est grande, plus le retour à un phénotype de croissance cellulaire « normale » est observé.

Ces données mises en commun nous confirment que Tho2 doit se lier aux acides nucléiques (ADN ou ARN) des cibles rendues aberrantes par Rho, pour que Rrp6 puisse les dégrader. Les prises de vues IGV illustrent bien l'impact des différentes délétions au niveau du recrutement de Rrp6 (Figure 4.3.5.-C). En présence de Rho dans un contexte génétique WT, un sur-recrutement de Tho2, accompagné d'un sur-recrutement de Rrp6 est observable. La déplétion de Mft1, modifie le profil de recrutement de Rrp6 et semble déstabiliser sa liaison au loci analysés. En revanche, la déplétion de Tho2, supprime tous signaux Rrp6. La délétion courte du domaine C-terminal de Tho2, montre bien une perte du nombre de signaux Rrp6 observables. Enfin, la délétion totale du domaine C-terminal s'accompagne d'une perte de signal Rrp6, à l'image de la déplétion totale de Tho2.

A/



B/



C/

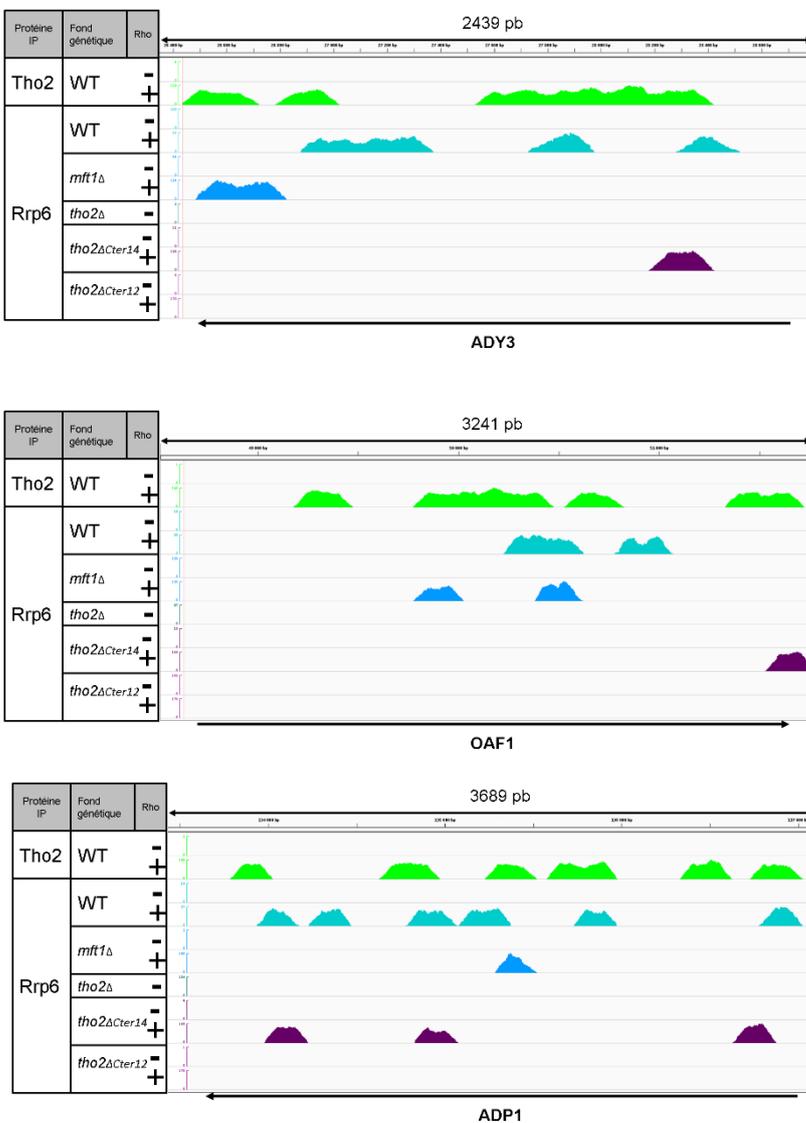


Figure 4.3.5. A/Histogramme du nombre moyen de pics Rrp6 par locus (haut) et de l'intensité moyenne du signal Rrp6 (bas) obtenus dans les différents fonds génétiques étudiés. Ces données sont extraites au niveau des loci correspondants aux ARNm sauvés par la déplétion de Rrp6. B/ Test de pousse en gouttes des différentes souches en absence (+Doxy) ou présence (-Doxy) de Rho. C/ Snapshots IGV, montrant l'impact des différents mutants sur les signaux de ChIP de Rrp6 (Des exemples supplémentaires seront fournis en figures supplémentaires lors de la soumission de ce travail pour publication).

Afin de jauger l'impact direct des mutants du complexe THO étudiés sur certains ARNs sensibles à Rho et dégradés par Rrp6, des expériences de RT-qPCR ont été menées pour déterminer le niveau de ces derniers dans les différentes conditions. Ces données nous indiquent que la déplétion de Tho2 donne des résultats similaires, bien que légèrement moindres en moyenne, à la déplétion de Rrp6 du point de vue du niveau des ARNs en question. De manière intéressante, la grande délétion du domaine C-terminal de Tho2 donne des résultats de « sauvetage » d'ARN plus élevé que la délétion de la protéine totale. Ceci peut être expliqué par le fait que dans le cas du domaine délété, la protéine Tho2 tronquée est toujours présente dans les cellules, et va potentiellement pouvoir jouer d'autres rôles via ses autres domaines.

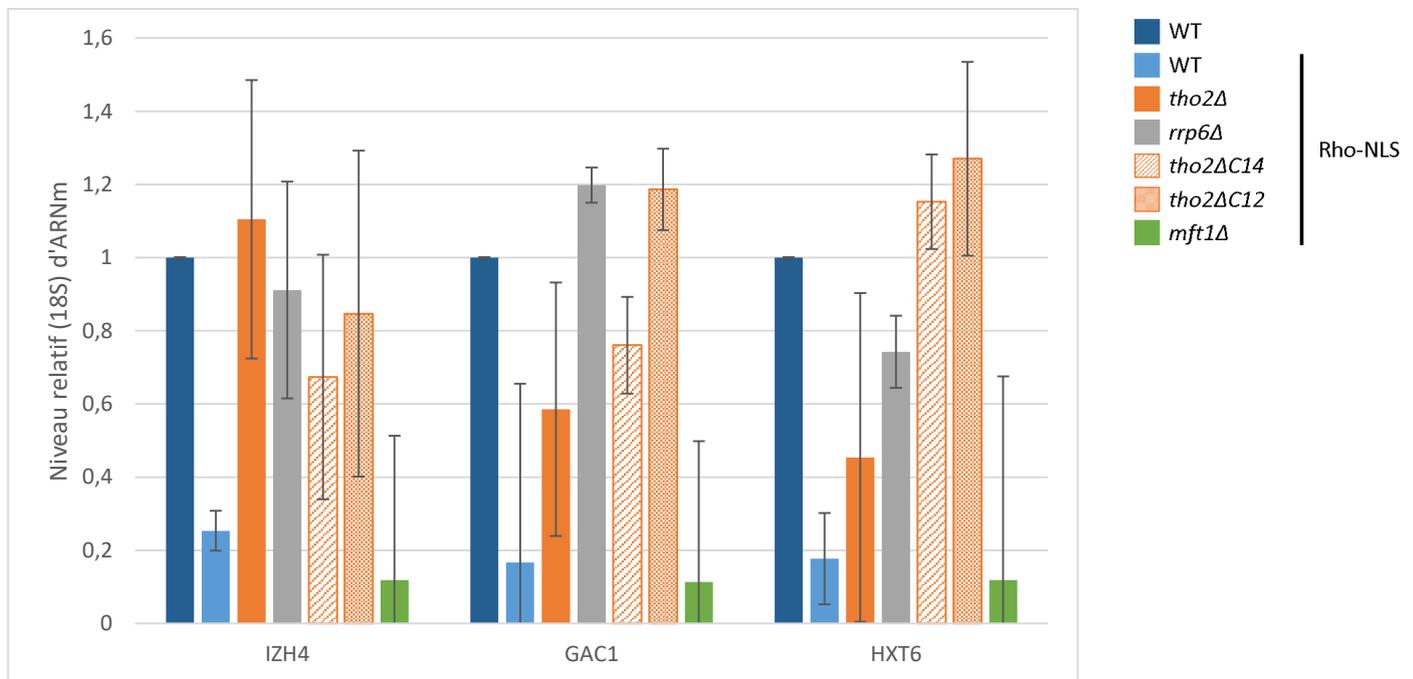


Figure 4.3.6. Impact des délétions des membres du complexe THO sur le niveau de certains ARNs. RT-qPCR sur 3 ARNm choisis aléatoirement parmi les 491 transcrits affectés par Rho et sauvés par la déplétion de Rrp6. Le niveau de chaque ARN est normalisé par rapport au niveau de l'ARNr 18S. Le niveau « normal » de chaque transcrit est fixé à 1 dans la condition WT sans Rho. Chaque niveau, dans chaque souche mutante en présence de Rho est relatif au niveau fixé à 1.

La présence de la protéine Tho2 semble être une condition indispensable au recrutement de Rrp6 au niveau de ses cibles. Reste à savoir si les deux protéines interagissent physiquement l'une avec l'autre. Afin de répondre à cette question, des expériences de co-immunoprécipitations sont en cours et les résultats, bien que préliminaires, sont encourageants.

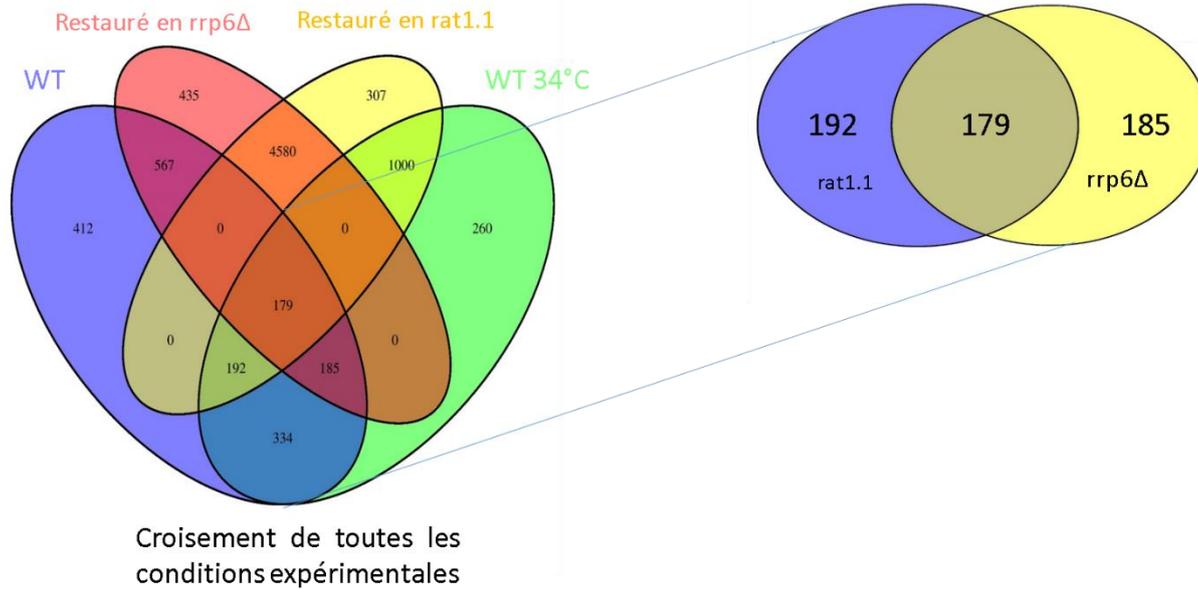
En résumé, nous montrons ici que la protéine Tho2, sous-unité du complexe THO, est capable, en dehors de son complexe canonique, de se fixer, entre autre, aux loci des ARNs rendus aberrants par l'action de l'hélicase translocase bactérienne Rho. De plus, Tho2 semble posséder une fonction de guide pour Rrp6, dans le sens où cette dernière se fixe principalement au niveau de loci, correspondant à des transcrits aberrants, où Tho2 est présente. Dans la veine de cette supposition, Tho2 doit pouvoir se lier aux acides nucléiques pour jouer son rôle de guide de Rrp6. Lors d'une délétion du domaine de liaison aux acides nucléiques de Tho2, on ne retrouve plus Rrp6 sur ses cibles, ni même sur la chromatine (absence de signal) lorsque la délétion du domaine C-terminal de Tho2 est totale. Ces résultats de localisations génomiques sont appuyés par des données phénotypiques, qui montrent que la délétion croissante du domaine C-terminal de Tho2 tend à un retour de croissance normale des levures en présence de Rho. Ce phénomène, aux lumières des précédents résultats de l'équipe avec ce genre de test, va dans le sens d'une implication de la protéine (et par extension de sa capacité à se lier aux acides nucléiques) dans des mécanismes de contrôle qualité. De surcroît, les analyses qPCR du niveau de certains transcrits cibles du système de QC s'inscrivent également dans ce faisceau d'informations concernant l'implication de la protéine Tho2 dans des mécanismes de surveillance de la biogenèse des mRNPs. En effet, les ARNs analysés montrent des niveaux relatifs diminués sous l'action de Rho, et dans les souches délétées d'une partie ou de la totalité du domaine C-terminal de Tho2 ces niveaux se voient restaurés.

Nous montrons donc, qu'en plus du rôle primordial dans l'assemblage co-transcriptionnel de la mRNP et le couplage de l'élongation de la transcription avec la maturation de l'extrémité 3' de l'ARN et son export vers le cytoplasme, la protéine Tho2 du complexe THO participe activement à la surveillance de la conformité de la molécule ribonucléoprotéique. L'étape suivante serait de voir si Tho2 et Rrp6 ont la capacité d'interagir physiquement l'une avec l'autre, et ainsi apporter la preuve que Tho2 peut diriger Rrp6 vers les cibles à dégrader.

5. Travaux en cours

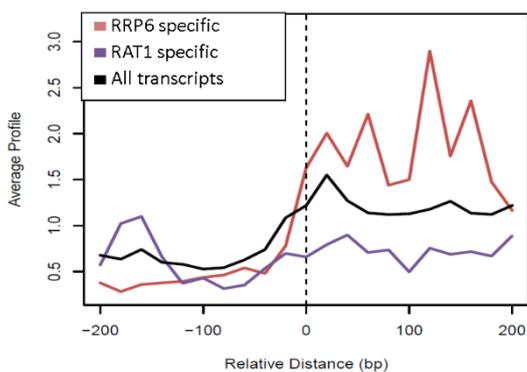
Les analyses bio-informatique de cette partie ont été réalisés par mes soins en se basant sur des données de séquençage haut débit publiées par d'autres groupes de recherche. Les données de RNA-seq (diagramme de Venn) proviennent de mes propres manipulations (culture des différentes souches de levures et extraction d'ARN). Les résultats présentés dans cette courte section sont préliminaires et sont des pistes de réflexions prometteuses dont l'analyse est en cours. Ces résultats sont discutés dans la section 6 « Discussion et Perspectives ». Les ARNs dont les niveaux baissent suite à la perturbation de leurs assemblages en mRNPs, ne voient pas tous leurs niveaux restaurés dans des souches déplétés de Rrp6. Cette observation n'est pas incohérente avec les précédents travaux de l'équipe qui montrent qu'une seconde voie de dégradation des ARNs aberrants coexiste avec la voie Rrp6-dépendante. Cette seconde voie passe par l'exonucléase 5'-3' Rat1 (Mosrin-Huaman et al., 2016). J'ai réalisé une série d'expériences dans une souche mutante pour Rat1 (mutation ponctuelle présentant une termosensibilité à 34°C, *rat1.1*) qui confirme ces observations. Les résultats obtenus montrent que les ARNm qui ne sont pas sauvés par la déplétion de Rrp6 (*rrp47Δ* (Stuparevic et al., 2013)) sont en partie sauvés dans ce mutant ponctuel de Rat1 (Figure 5.1). Les résultats montrent, entre autres, une différence entre les deux lots d'ARNs (dégradés par le 3' ou par le 5') du point de vue du « pausing » de polII en début de transcription (résultats obtenus en utilisant les données de NET-seq de Churchman et al. (Churchman and Weissman, 2011) en filtrant les données avec nos lots d'ARNs). La transcription des ARNs sensibles à la dégradation par Rat1 ne montre pas de pause en début de gène. Ces résultats sont appuyés par l'observation d'une hypophosphorylation de la Ser5 dans la même fenêtre d'observation pour les ARNs dégradés par Rat1 (données provenant de (Mayer et al., 2010) en filtrant avec nos lots d'ARNs). Egalement, un pic de recrutement de Nrd1 en début de transcription sur les ARNs sensible à une dégradation Rrp6-dépendante 3'-5' (Creamer et al., 2011) est remarquable comparé à l'autre lot d'ARN. Des expériences d'immunoprécipitations du complexe de liaison à la coiffe (CBC) ont également été réalisées et nous montrent que les ARNs dégradés par Rat1 présentent moins de CBC que les ARNs Rrp6-dépendant. Nous avons donc des résultats en cours d'analyses qui semblent converger vers un modèle dans lequel deux machinerie de dégradation des ARNs aberrants coexistent et où les mécanismes de reconnaissance des défauts de biogenèse semblent différents.

A/



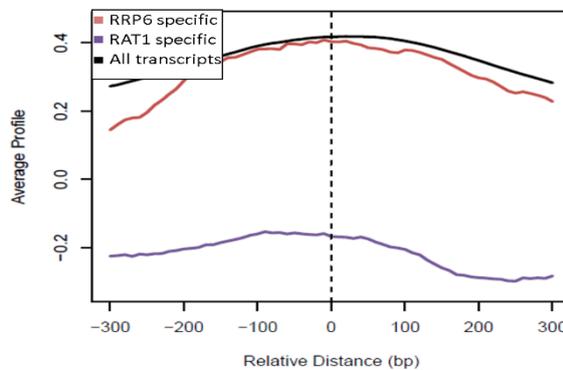
B/

Profil de NET-seq



C/

Profil de Ser5 phosphorylée par CHIP-seq



D/

Profil de PAR-CLIP de Nrd1

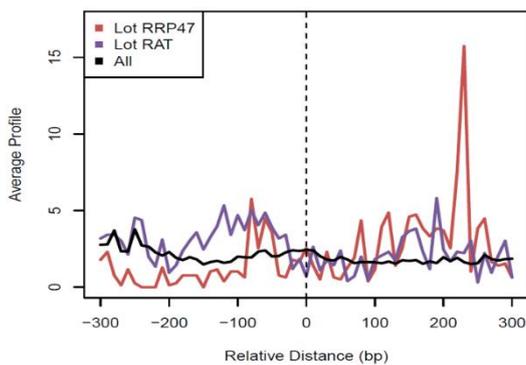


Figure 5.1. Comparaison des conditions expérimentales entre les expériences en *rrp47Δ* et *rat1.1*. A/ Diagramme de Venn croisant toutes les données obtenues en RNA-seq. Le diagramme de droite est un résumé des conditions pour lesquelles les ARNs sont sauvées par l'altération de l'une, l'autre, ou les deux machineries de dégradations. B/ Profil de NET-seq pour les deux lots d'ARNs spécifiques (rouge pour les ARNs dégradés par Rrp6, bleue pour les ARNs dégradés par Rat1, et noire pour tous les ARNm de la levure) (Churchman and Weissman, 2011). C/ Profil de phosphorylation de la Ser5 de polII obtenu en CHIP-seq (Mayer et al., 2010). D/ Profil de PAR-CLIP de Nrd1 (Creamer et al., 2011).

6. Discussion et Perspectives

Les travaux présentés dans cette thèse ont été initiés par le besoin que connaît la Biologie moderne de voir les choses dans leurs globalités. En effet, les observations faites par l'équipe, depuis plusieurs années, devaient être étendues à une étude génomique et transcriptomique afin de pouvoir tracer un modèle global de la réponse des cellules de levure à la perturbation de la biogenèse de leurs mRNPs.

Les premières observations à l'échelle transcriptomique de l'impact de Rho sur le niveau des ARNm, nous a permis de voir que toutes les mRNPs ne voyaient pas leur biogenèse touchée. Notre système nous permet donc d'avoir des témoins interne. En effet, les ARNs dont le niveau est invariant peuvent servir de contrôle d'une expérience à l'autre. Qui plus est, les ARNs dont les niveaux baissent suite à la perturbation de leurs assemblages en mRNPs, peuvent être sauvés par la déplétion de l'une ou l'autre des machineries de QC et dégradation. Les résultats présentés dans la dernière partie de ce manuscrit concernant les différentes caractéristiques des ARNs dont la dégradation est spécifique de l'une ou l'autre des voies sont préliminaires et demandent de plus amples analyses. En effet, en utilisant les données provenant de différents travaux de plusieurs équipes, nous pouvons commencer à décrire les différences qui existent entre les deux populations d'ARNs. Ces descriptions nous donnent des pistes pour comprendre l'impact de la coexistence de deux machineries au sein du noyau des levures sans avoir la variable Rho dans l'équation.

La dégradation des ARNs par l'exonucléase 5'-3' Rat1, s'effectue sur des ARNs dont la transcription semble posséder des caractéristiques différentes des ARNs dégradés par l'exonucléase 3'-5' Rrp6 (Figure 24). En effet, polII n'a pas l'air de présenter de pause vraiment marquée en début de transcription. De plus, une hypophosphorylation de la Ser5 du CTD de polII est observable pour ces ARNs dont la dégradation est spécifique de Rat1. Enfin, les ARNs sensibles à la dégradation par Rrp6 présentent un pic de Nrd1 de manière basal en début de transcription, qui n'est pas retrouvé pour la population d'ARNs dont la dégradation est assurée par Rat1. Ces caractéristiques différentes entre les transcrits dégradés par l'une ou l'autre des exonucléases ne nous indiquent pas encore clairement les tenants et les aboutissants des mécanismes de détections de la malformation en mRNP qui conduiraient à l'une ou l'autre des voies. Une possible hypothèse quant à la coexistence de ces deux voies de dégradation des ARNs pourrait émerger d'une localisation des gènes différente dans le noyau. Une des

deux populations pourrait se retrouver en périphérie du noyau et nécessiterait une surveillance accrue, ainsi qu'une dégradation rapide en cas de mauvaise formation afin d'éviter l'export d'une mRNP non conforme vers le cytoplasme.

L'outil de perturbation Rho a été un grand avantage pour déterminer les deux voies de dégradations différentes et les ARNs soumis à l'une ou l'autre de ces voies. Afin de définir les caractéristiques physiologiques qui différencient ces ARNs, il est maintenant nécessaire de s'intéresser aux différents travaux déjà publiés sur les facteurs généraux de transcriptions et de biogénèse des mRNPs à travers le prisme de nos deux groupes d'ARNs. De nombreuses études transcriptomiques prennent en compte l'ensemble des ARNs et par voie de conséquence moyennent les observations, en écrasant les possibles différences qui existent dans l'ensemble du transcriptome de l'organisme étudié. J'ai commencé à réaliser cette ré-analyse de données en utilisant différents jeux de données issues de travaux publiés et dont certains sont présentés en partie 5 de mes résultats. Ceci ne représente que le début d'un travail qui va nécessiter de plus amples analyses, et potentiellement de nouvelles expérimentations de notre part. Une difficulté sous-jacente à ce genre d'analyse, et qu'il est important de souligner, est liée à l'hétérogénéité des jeux de données publique. En effet, l'analyse bio-informatique est un outil puissant permettant de mener des analyses de pointes et précises sur différents aspects de la biologie des acides nucléiques dans mon cas. En revanche, l'utilisation des données publiques nécessite une rigueur de ré-analyse poussée afin de ne pas faussement interpréter des résultats issus de données quelque fois décrites succinctement. L'apparition de nombreux outils, comme le CHIP-Atlas (Oki *et al.*, 2018), vont nous permettre de s'affranchir de plus en plus de ces limites et, je l'espère, faire entrer l'analyse bio-informatique dans une ère d'accessibilité au plus grand nombre.

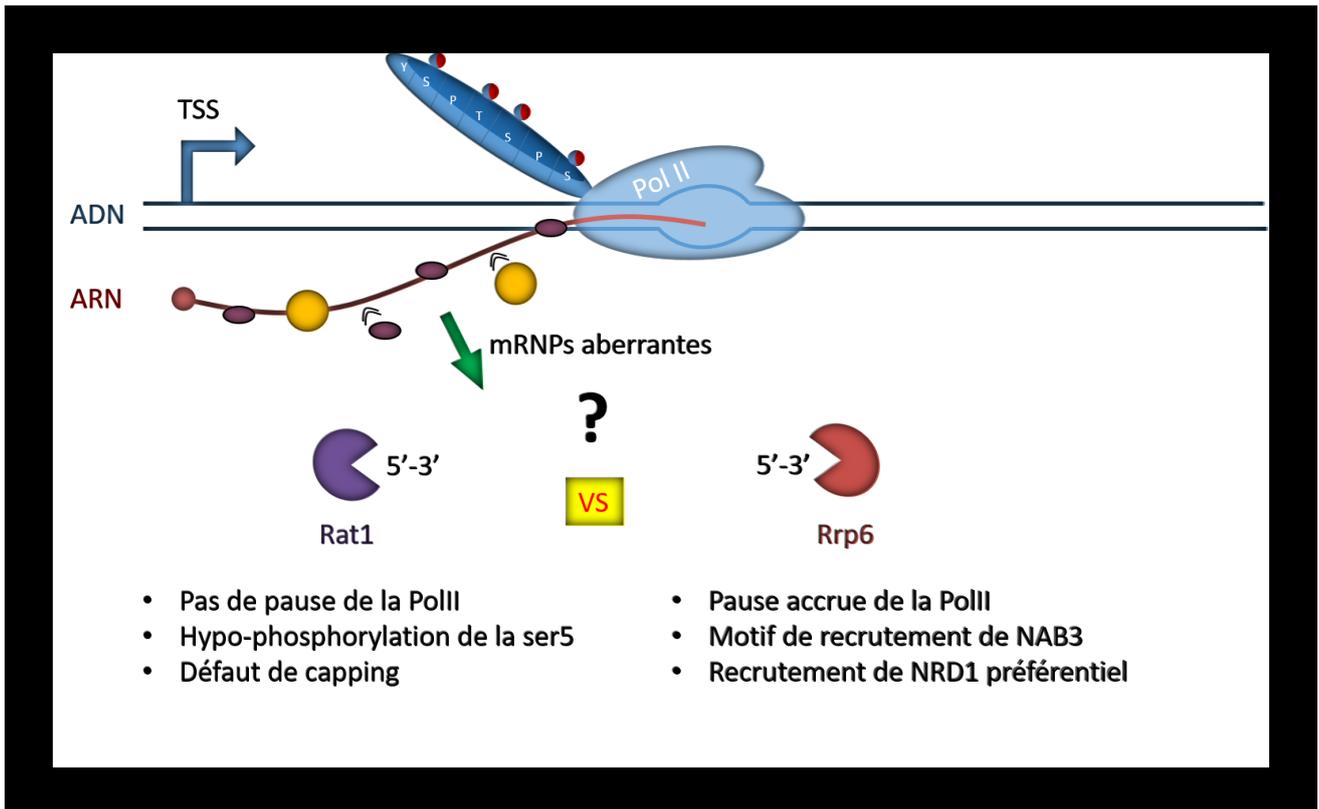


Figure 24 Schéma du modèle à deux machineries de dégradation des ARNs aberrants

Concernant les résultats de l'article n°2, le système de QC peut être vu comme « occupé » par la surveillance des transcrits non codants en permanence. Cette machinerie ne va se remobiliser au niveau des ARNm que lorsqu'un défaut de formation en mRNPs apparaît. Les signaux de CHIP-seq concernant Nrd1, Nab3, Trf4 et Rrp6 nous indiquent que ces 4 protéines sont proches spatialement au niveau des loci correspondant aux ARNs dont la biogenèse est perturbée et qui sont sauvés lorsque Rrp6 est déplétée. Ces résultats nous apprennent que les protéines du QC et dégradation agissent en cours de transcription et sont présentes ensemble bien avant la terminaison de ce procédé.

Pour pouvoir obtenir un aperçu complet du modèle de recrutement de ces composants, des expériences de PAR-CLIP pourraient être menées afin de déterminer de manière spatio-temporelle la manière dont le système de contrôle qualité bascule sur l'ARN, et va le préparer à être dégradé. De même, il serait intéressant dans ce modèle,

d'immunoprécipiter certaines de ces protéines dans un contexte de perturbation par Rho, et de regarder par spectrométrie de masse les protéines présentes sur ces ARNs. Ces expériences nous donneraient une vue d'ensemble des acteurs de la surveillance des procédés de la biogenèse des mRNPs.

Les expériences concernant le complexe THO viennent compléter notre compréhension des mécanismes de détection des mRNPs aberrantes. Le fait que le complexe soit désassemblé en présence de Rho et que la sous-unité Tho2 se voit sur-recrutée, entre autre, au niveau des loci correspondants aux 491 ARNm sauvés par la déplétion de Rrp6 peut avoir plusieurs explications : 1/ Rho désassemble le complexe par son activité et seule Tho2, qui possède un domaine de liaison aux acides nucléiques, reste présente. 2/ la perturbation globale de la biogenèse par Rho est un signal pour que le complexe, excepté Tho2, se sépare des ARNs cibles, afin d'empêcher son export futur, et la sous-unité Tho2 sert de marqueur de contrôle qualité pour le recrutement du reste du système de surveillance y compris Rrp6. Cette dernière hypothèse est en cours de vérification au laboratoire à travers des expériences de co-immunoprécipitation entre Tho2 et Rrp6. Dans la même optique, des expériences de surexpression du domaine C-terminal de Tho2 dans des souches de levures exprimant Rho, pourraient permettre de renforcer nos observations sur l'importance de ce domaine permettant la liaison aux acides nucléiques. En effet, si la surexpression du domaine permet d'entrer en compétition avec la protéine endogène sur ces cibles, elle inhiberait le recrutement d'Rrp6. Par voie de conséquence, nous verrions le phénotype de retard de croissance induit par l'action de Rho rétablit à une croissance « normale ». De même le niveau des ARNm spécifiques de ce ciblage et de cette voie de dégradation se verrait rétablit.

Il est intéressant de noter qu'un des premiers complexes protéiques qui vient s'apposer sur l'ARN naissant afin de l'empaqueter et de le préparer à l'export via des interactions avec d'autres protéines peut également être acteur du contrôle qualité. Sa place de choix sur l'ARN naissant en fait un senseur parfait de son intégrité. Tho2, en plus de jouer ce rôle dans la surveillance, pourrait bien être une des pierres angulaires du recrutement de Rrp6 sur ses cibles.

Toutefois, les résultats présentés sur le complexe THO peuvent être critiqués. En effet, afin d'affirmer l'implication de Tho2 dans le QC, les analyses bio-informatiques sont à poursuivre sur l'ensemble du génome de la levure, et non plus sur les 491 loci spécifiques. J'ai

pu observer que le profil de Tho2 n'était pas exclusif à ces derniers. Il va donc falloir tester si le recrutement global de Tho2 sur la chromatine influence le recrutement de Rrp6 en dehors de ses cibles. Des analyses préliminaires et donc à confirmer, semblent indiquer que le recrutement de Rrp6 n'est pas limité aux 491 loci étudiés, mais ne semble pas être aussi intense en dehors de ceux-ci. L'hypothèse de guidage de Rrp6 par Tho2 peut donc être nuancée et d'autres facteurs déterminants doivent exister et jouer un rôle important dans ce mécanisme.

La « disparition » des autres sous unités du complexe THO de la chromatine des ARNs perturbés, en plus de la présence accrue de Tho2 et de Rrp6 ainsi que celle de Nrd1, Nab3 et Trf4, soulève des questions quant aux connaissances sur les complexes protéiques définis et canoniques. En effet, sans remettre en question les fonctions connues de ces complexes sur l'ARN ainsi que les nombreuses études qui ont montré la place importante de ces complexes dans la biogenèse, il semblerait qu'une dynamique d'interaction protéine-protéine nouvelle existe. Cette dernière pourrait être spécifique à des cas de figures où la formation des mRNPs conformes est compromise (Figure 25).

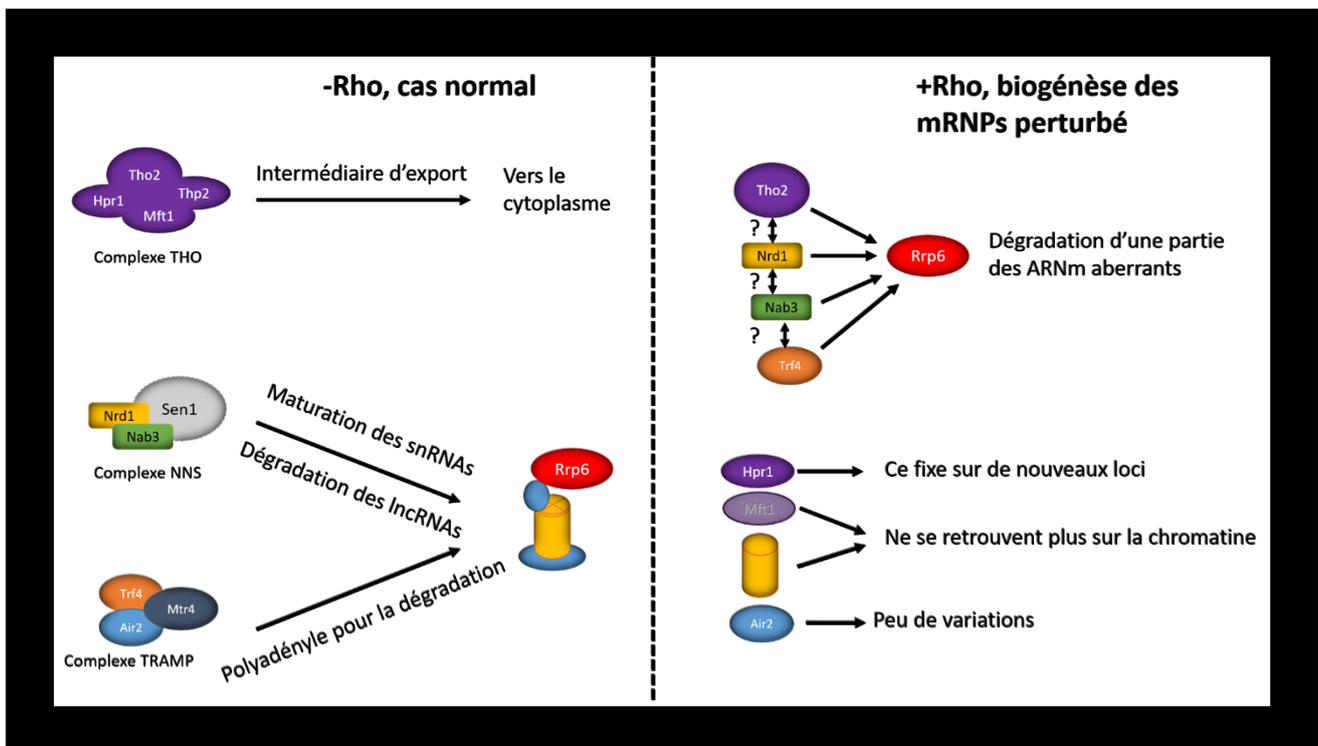


Figure 25 Résumé de la dynamique de remobilisation des différentes sous-unités des complexes étudiés

Bibliographie

Abou Elela, S., and Ares, M., Jr. (1998). Depletion of yeast RNase III blocks correct U2 3' end formation and results in polyadenylated but functional U2 snRNA. *EMBO J* 17, 3738-3746.

Abruzzi, K.C., Lacadie, S., and Rosbash, M. (2004). Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *EMBO J* 23, 2620-2631.

Aguilera, A. (2005). Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. *Curr Opin Cell Biol* 17, 242-250.

Aguilera, A., and Garcia-Muse, T. ((2012)). R loops: from transcription byproducts to threats to genome stability. *Mol Cell* 46, 115-124.

Andersson, R., Refsing Andersen, P., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T., and Sandelin, A. (2014). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun* 5.

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

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

Avery, O.T., Macleod, C.M., and McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *The Journal of experimental medicine* 79, 137-158.

Bacikova, V., Pasulka, J., Kubicek, K., and Stefl, R. (2014). Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res* 42, 8024-8038.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* 15, 163-175.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., *et al.* (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol* 25, 833-840.

Bonnet, A., Bretes, H., and Palancade, B. (2015). Nuclear pore components affect distinct stages of intron-containing gene expression. *Nucleic Acids Research* 43, 4249-4261.

Boudvillain, M., Figueroa-Bossi, N., and Bossi, L. (2013). Terminator still moving forward: expanding roles for Rho factor. *Current opinion in microbiology* 16, 118-124.

Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., *et al.* (2018). Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* 28, 420-435.

Brownlee, G.G., Sanger, F., and Barrell, B.G. (1968). The sequence of 5 s ribosomal ribonucleic acid. *J Mol Biol* 34, 379-412.

Buratowski, S. (2003). The CTD code. *Nat Struct Biol* 10, 679-680.

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.

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.

Cassart, C., Drogat, J., Migeot, V., and Hermand, D. (2012). Distinct requirement of RNA polymerase II CTD phosphorylations in budding and fission yeast. *Transcription* 3, 231-234.

Castelnuovo, M., Zaugg, J.B., Guffanti, E., Maffioletti, A., Camblong, J., Xu, Z., Clauder-Munster, S., Steinmetz, L.M., Luscombe, N.M., and Stutz, F. (2014). Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Res* 42, 4348-4362.

Chanarat, S., Burkert-Kautzsch, C., Meinel, D.M., and Strasser, K. (2012). Prp19C and TREX: interacting to promote transcription elongation and mRNA export. *Transcription* 3, 8-12.

Chanfreau, G., Rotondo, G., Legrain, P., and Jacquier, A. (1998). Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J* *17*, 3726-3737.

Chavez, S., Beilharz, T., Rondon, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera, A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *EMBO J* *19*, 5824-5834.

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*. *Eukaryotic cell* *11*, 417-429.

Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* *11*, 3319-3326.

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

Ciampi, M.S. (2006). Rho-dependent terminators and transcription termination. *Microbiology* *152*, 2515-2528.

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.

Costello, J.L., Stead, J.A., Feigenbutz, M., Jones, R.M., and Mitchell, P. (2011). The C-terminal region of the exosome-associated protein Rrp47 is specifically required for box C/D small nucleolar RNA 3'-maturation. *J Biol Chem* *286*, 4535-4543.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* *292*, 1863-1876.

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.

D'Heygere, F., Rabhi, M., and Boudvillain, M. (2013). Phyletic distribution and conservation of the bacterial transcription termination factor Rho. *Microbiology* *159*, 1423-1436.

Davis, C.A., and Ares, M., Jr. (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.

Decatur, W.A., and Fournier, M.J. (2002). rRNA modifications and ribosome function. *Trends Biochem Sci* *27*, 344-351.

Delan-Forino, C., Schneider, C., and Tollervey, D. (2017). Transcriptome-wide analysis of alternative routes for RNA substrates into the exosome complex. *PLoS Genet* *13*, e1006699.

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.

Egecioglu, D.E., Henras, A.K., and Chanfreau, G.F. (2006). Contributions of Trf4p- and Trf5p-dependent polyadenylation to the processing and degradative functions of the yeast nuclear exosome. *RNA* *12*, 26-32.

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

Epshtein, V., Dutta, D., Wade, J., and Nudler, E. (2010). An allosteric mechanism of Rho-dependent transcription termination. *Nature* *463*, 245-249.

Eshleman, N., Liu, G., McGrath, K., Parker, R., and Buchan, J.R. (2016). Defects in THO/TREX-2 function cause accumulation of novel cytoplasmic mRNP granules that can be cleared by autophagy. *RNA* *22*, 1200-1214.

Falk, S., Bonneau, F., Ebert, J., Kogel, A., and Conti, E. (2017). Mpp6 Incorporation in the Nuclear Exosome Contributes to RNA Channeling through the Mtr4 Helicase. *Cell Rep* *20*, 2279-2286.

Fasken, M.B., Laribee, R.N., and Corbett, A.H. (2015). Nab3 facilitates the function of the TRAMP complex in RNA processing via recruitment of Rrp6 independent of Nrd1. *PLoS Genet* *11*, e1005044.

Fasken, M.B., Leung, S.W., Banerjee, A., Kodani, M.O., Chavez, R., Bowman, E.A., Purohit, M.K., Rubinson, M.E., Rubinson, E.H., and Corbett, A.H. (2011). Air1 zinc knuckles 4 and 5 and a conserved IWRXY motif are critical for the function and integrity of the Trf4/5-Air1/2-Mtr4 polyadenylation (TRAMP) RNA quality control complex. *J Biol Chem* *286*, 37429-37445.

Faza, M.B., Kemmler, S., Jimeno, S., Gonzalez-Aguilera, C., Aguilera, A., Hurt, E., and Panse, V.G. (2009). Sem1 is a functional component of the nuclear pore complex-associated messenger RNA export machinery. *The Journal of cell biology* *184*, 833-846.

Geiselman, J., Yager, T.D., Gill, S.C., Calmettes, P., and von Hippel, P.H. (1992). Physical properties of the Escherichia coli transcription termination factor rho. 1. Association states and geometry of the rho hexamer. *Biochemistry* *31*, 111-121.

Geisler, S., and Collier, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* *14*, 699-712.

Gocheva, V., Le Gall, A., Boudvillain, M., Margeat, E., and Nollmann, M. (2015). Direct observation of the translocation mechanism of transcription termination factor Rho. *Nucleic Acids Res* *43*, 2367-2377.

Gomez-Gonzalez, B., and Aguilera, A. (2009). R-loops do not accumulate in transcription-defective hpr1-101 mutants: implications for the functional role of THO/TREX. *Nucleic Acids Res* *37*, 4315-4321.

Gomez-Gonzalez, B., Garcia-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marin, A., Foiani, M., and Aguilera, A. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J* *30*, 3106-3119.

Gonzalez-Aguilera, C., Tous, C., Gomez-Gonzalez, B., Huertas, P., Luna, R., and Aguilera, A. (2008). The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol Biol Cell* *19*, 4310-4318.

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., Seraphin, 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.

Gwizdek, C., Iglesias, N., Rodriguez, M.S., Ossareh-Nazari, B., Hobeika, M., Divita, G., Stutz, F., and Dargemont, C. (2006). Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. *Proc Natl Acad Sci U S A* *103*, 16376-16381.

Hackmann, A., Wu, H., Schneider, U.-M., Meyer, K., Jung, K., and Krebber, H. (2014). Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. *Nat Commun* *5*.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* *11*, 394-403.

Hahn, S., and Young, E.T. (2011). Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* *189*, 705-736.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep* *15*, 2147-2158.

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.

Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* *18*, 637-650.

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.

Honorine, R., Mosrin-Huaman, C., Hervouet-Coste, N., Libri, D., and Rahmouni, A.R. (2011). Nuclear mRNA quality control in yeast is mediated by Nrd1 co-transcriptional recruitment, as revealed by the targeting of Rho-induced aberrant transcripts. *Nucleic Acids Res* 39, 2809-2820.

Huber, F., Bunina, D., Gupta, I., Khmelinskii, A., Meurer, M., Theer, P., Steinmetz, Lars M., and Knop, M. (2016). Protein Abundance Control by Non-coding Antisense Transcription. *Cell Reports*.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* 12, 711-721.

Huertas, P., Garcia-Rubio, M.L., Wellinger, R.E., Luna, R., and Aguilera, A. (2006). An hpr1 point mutation that impairs transcription and mRNP biogenesis without increasing recombination. *Mol Cell Biol* 26, 7451-7465.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3, 318-356.

Jamonnak, N., Creamer, T.J., Darby, M.M., Schaughency, P., Wheelan, S.J., and Corden, J.L. (2011). Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA* 17, 2011-2025.

Jia, H., Wang, X., Liu, F., Guenther, U.P., Srinivasan, S., Anderson, J.T., and Jankowsky, E. (2011). The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. *Cell* 145, 890-901.

Jimeno, S., Rondon, A.G., Luna, R., and Aguilera, A. (2002). The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *EMBO J* 21, 3526-3535.

- Kadaba, S., Krueger, A., Trice, T., Krecic, A.M., Hinnebusch, A.G., and Anderson, J. (2004). Nuclear surveillance and degradation of hypomodified initiator tRNAMet in *S. cerevisiae*. *Genes Dev* *18*, 1227-1240.
- Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science* *301*, 1096-1099.
- Kaplan, D.L., and O'Donnell, M. (2003). Rho factor: transcription termination in four steps. *Current biology : CB* *13*, R714-716.
- Katahira, J. (2015). Nuclear export of messenger RNA. *Genes* *6*, 163-184.
- Kilchert, C., Wittmann, S., and Vasiljeva, L. (2016). The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol* *17*, 227-239.
- Kiss, A.M., Jady, B.E., Darzacq, X., Verheggen, C., Bertrand, E., and Kiss, T. (2002). A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic Acids Res* *30*, 4643-4649.
- Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol* *8*, 761-773.
- Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature methods* *11*, 319-324.
- Kyriakou, D., Stavrou, E., Demosthenous, P., Angelidou, G., San Luis, B.J., Boone, C., Promponas, V.J., and Kirmizis, A. (2016). Functional characterisation of long intergenic non-coding RNAs through genetic interaction profiling in *Saccharomyces cerevisiae*. *BMC biology* *14*, 106.
- 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.

Libri, D., Dower, K., Boulay, J., Thomsen, R., Rosbash, M., and Jensen, T.H. (2002). Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. *Mol Cell Biol* 22, 8254-8266.

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

Loya, T.J., O'Rourke, T.W., Degtyareva, N., and Reines, D. (2013). A Network of Interdependent Molecular Interactions Describes a Higher Order Nrd1-Nab3 Complex Involved in Yeast Transcription Termination. *Journal of Biological Chemistry* 288, 34158-34167.

Loya, T.J., O'Rourke, T.W., and Reines, D. (2017). The hnRNP-like Nab3 termination factor can employ heterologous prion-like domains in place of its own essential low complexity domain. *PLoS One* 12, e0186187.

Loya, T.J., and Reines, D. (2016). Recent advances in understanding transcription termination by RNA polymerase II. *F1000Research* 5.

Luna, R., Gaillard, H., Gonzalez-Aguilera, C., and Aguilera, A. (2008). Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. *Chromosoma* 117, 319-331.

Luna, R., Rondon, A.G., and Aguilera, A. (2012). New clues to understand the role of THO and other functionally related factors in mRNP biogenesis. *Biochim Biophys Acta* 1819, 514-520.

Lunde, B.M., Horner, M., and Meinhart, A. (2011). Structural insights into cis element recognition of non-polyadenylated RNAs by the Nab3-RRM. *Nucleic Acids Res* 39, 337-346.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 17, 1195-1201.

Luo, W., Johnson, A.W., and Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev* 20, 954-965.

Massenet, S., Bertrand, E., and Verheggen, C. (2017). Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol* 14, 680-692.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 17, 1272-1278.

Meinel, D.M., Burkert-Kautzsch, C., Kieser, A., O'Duibhir, E., Siebert, M., Mayer, A., Cramer, P., Soding, J., Holstege, F.C., and Strasser, K. (2013). Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II. *PLoS Genet* 9, e1003914.

Meselson, M., and Stahl, F.W. (1958). The Replication of DNA in Escherichia Coli. *Proc Natl Acad Sci U S A* 44, 671-682.

Minocha, R., Popova, V., Kopytova, D., Misiak, D., Huttelmaier, S., Georgieva, S., and Strasser, K. (2018). Mud2 functions in transcription by recruiting the Prp19 and TREX complexes to transcribed genes. *Nucleic Acids Res* 46, 9890.

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

Milligan, L., Decourty, L., Saveanu, C., Rappsilber, J., Ceulemans, H., Jacquier, A., and Tollervey, D. (2008). A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. *Mol Cell Biol* 28, 5446-5457.

Milligan, L., Huynh-Thu, V.A., Delan-Forino, C., Tuck, A., Petfalski, E., Lombraña, R., Sanguinetti, G., Kudla, G., and Tollervey, D. (2016). Strand-specific, high-resolution mapping of modified RNA polymerase II. *Molecular systems biology* 12.

Mitchell, P. (2014). Exosome substrate targeting: the long and short of it. *Biochemical Society transactions* 42, 1129-1134.

Mitchell, P., Petfalski, E., Houalla, R., Podtelejnikov, A., Mann, M., and Tollervey, D. (2003). Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. *Mol Cell Biol* 23, 6982-6992.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. *Cell* *91*, 457-466.

Mitchell, S.F., and Parker, R. (2014). Principles and Properties of Eukaryotic mRNPs. *Molecular Cell* *54*, 547-558.

Mor, A., Suliman, S., Ben-Yishay, R., Yunger, S., Brody, Y., and Shav-Tal, Y. (2010). Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. *Nature cell biology* *12*, 543-552.

Mosrin-Huaman, C., Hervouet-Coste, N., and Rahmouni, A.R. (2016). Co-transcriptional degradation by the 5'-3' exonuclease Rat1p mediates quality control of HXK1 mRNP biogenesis in *S. cerevisiae*. *RNA Biology* *13*, 582-592.

Mosrin-Huaman, C., Honorine, R., and Rahmouni, A.R. (2009). Expression of bacterial Rho factor in yeast identifies new factors involved in the functional interplay between transcription and mRNP biogenesis. *Mol Cell Biol* *29*, 4033-4044.

Mouaikel, J., Causse, S.Z., Rougemaille, M., Daubenton-Carafa, Y., Blugeon, C., Lemoine, S., Devaux, F., Darzacq, X., and Libri, D. (2013). High-frequency promoter firing links THO complex function to heavy chromatin formation. *Cell Rep* *5*, 1082-1094.

Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalishman, Y., Liu, X., Gibbons, B.J., *et al.* (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* *342*, 1238724.

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.

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.

Nevers, A., Doyen, A., Malabat, C., Neron, B., Kergrohen, T., Jacquier, A., and Badis, G. (2018). Antisense transcriptional interference mediates condition-specific gene repression in budding yeast. *Nucleic Acids Res* *46*, 6009-6025.

O'Rourke, T.W., Loya, T.J., Head, P.E., Horton, J.R., and Reines, D. (2015). Amyloid-like assembly of the low complexity domain of yeast Nab3. *Prion* *9*, 34-47.

O'Rourke, T.W., and Reines, D. (2016). Determinants of Amyloid Formation for the Yeast Termination Factor Nab3. *PLoS One* *11*, e0150865.

Ogami, K., Chen, Y., and Manley, J.L. (2018). RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Non-coding RNA* *4*.

Palancade, B., Zuccolo, M., Loeillet, S., Nicolas, A., and Doye, V. (2005). Pml39, a novel protein of the nuclear periphery required for nuclear retention of improper messenger ribonucleoparticles. *Mol Biol Cell* *16*, 5258-5268.

Park, J.S., and Roberts, J.W. (2006). Role of DNA bubble rewinding in enzymatic transcription termination. *Proc Natl Acad Sci U S A* *103*, 4870-4875.

Paul, B., and Montpetit, B. (2016). Altered RNA processing and export lead to retention of mRNAs near transcription sites and nuclear pore complexes or within the nucleolus. *Mol Biol Cell* *27*, 2742-2756.

Pena, A., Gewartowski, K., Mroczek, S., Cuellar, J., Szykowska, A., Prokop, A., Czarnocki-Cieciura, M., Piwowarski, J., Tous, C., Aguilera, A., *et al.* (2012). Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. *EMBO J* *31*, 1605-1616.

Piruat, J.I., and Aguilera, A. (1998). A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *EMBO J* *17*, 4859-4872.

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.

- Porrua, O., and Libri, D. (2013). 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.
- Poulsen, J.B., Sanderson, L.E., Agerschou, E.D., Dedic, E., Boesen, T., and Brodersen, D.E. (2014). Structural characterization of the *Saccharomyces cerevisiae* THO complex by small-angle X-ray scattering. *PLoS One* 9, e103470.
- Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. *Genes Dev* 23, 1247-1269.
- Richardson, J.P. (2002). Rho-dependent termination and ATPases in transcript termination. *Biochim Biophys Acta* 1577, 251-260.
- Rougemaille, M., Dieppois, G., Kisseleva-Romanova, E., Gudipati, R.K., Lemoine, S., Blugeon, C., Boulay, J., Jensen, T.H., Stutz, F., Devaux, F., *et al.* (2008). THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association. *Cell* 135, 308-321.
- Rougemaille, M., Gudipati, R.K., Olesen, J.R., Thomsen, R., Seraphin, B., Libri, D., and Jensen, T.H. (2007). Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. *EMBO J* 26, 2317-2326.
- Saguez, C., and Jensen, T.H. (2009). Assembly of export-competent mRNP: it's all about being connected. *Mol Cell* 33, 139-140.
- Saguez, C., Schmid, M., Olesen, J.R., Ghazy, M.A., Qu, X., Poulsen, M.B., Nasser, T., Moore, C., and Jensen, T.H. (2008). Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. *Mol Cell* 31, 91-103.
- Samarsky, D.A., Fournier, M.J., Singer, R.H., and Bertrand, E. (1998). The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J* 17, 3747-3757.

- Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-Wide Mapping of Yeast RNA Polymerase II Termination. *PLOS Genetics* *10*, e1004632.
- Schmid, M., and Jensen, T.H. (2008). Quality control of mRNP in the nucleus. *Chromosoma* *117*, 419-429.
- Schneider, C., Kudla, G., Wlotzka, W., Tuck, A., and Tollervey, D. (2012). Transcriptome-wide Analysis of Exosome Targets. *Mol Cell* *48*, 422-433.
- 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.
- Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* *155*, 1075-1087.
- Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B.J. (2016). Global Mapping of Human RNA-RNA Interactions. *Mol Cell* *62*, 618-626.
- Skordalakes, E., and Berger, J.M. (2003). Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* *114*, 135-146.
- Skordalakes, E., and Berger, J.M. (2006). Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor. *Cell* *127*, 553-564.
- Skordalakes, E., Brogan, A.P., Park, B.S., Kohn, H., and Berger, J.M. (2005). Structural Mechanism of Inhibition of the Rho Transcription Termination Factor by the Antibiotic Bicyclomycin. *Structure* *13*, 99-109.
- Sloan, K.E., Schneider, C., and Watkins, N.J. (2012). Comparison of the yeast and human nuclear exosome complexes. *Biochemical Society transactions* *40*, 850-855.
- Soares, E., Schwartz, A., Nollmann, M., Margeat, E., and Boudvillain, M. (2014). The RNA-mediated, asymmetric ring regulatory mechanism of the transcription termination Rho

helicase decrypted by time-resolved nucleotide analog interference probing (trNAIP). *Nucleic Acids Res* 42, 9270-9284.

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. (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., 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.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R., *et al.* (2002). TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* 417, 304-308.

Stuparevic, I., Mosrin-Huaman, C., Hervouet-Coste, N., Remenaric, M., and Rahmouni, A.R. (2013). Cotranscriptional recruitment of RNA exosome cofactors Rrp47p and Mpp6p and two distinct Trf-Air-Mtr4 polyadenylation (TRAMP) complexes assists the exonuclease Rrp6p in the targeting and degradation of an aberrant messenger ribonucleoprotein particle (mRNP) in yeast. *J Biol Chem* 288, 31816-31829.

Suh, M.H., Meyer, P.A., Gu, M., Ye, P., Zhang, M., Kaplan, C.D., Lima, C.D., and Fu, J. (2010). A dual interface determines the recognition of RNA polymerase II by RNA capping enzyme. *J Biol Chem* 285, 34027-34038.

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.

Thomsen, N.D., and Berger, J.M. (2009). Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* 139, 523-534.

Tollervey, D., and Hurt, E.C. (1990). The role of small nucleolar ribonucleoproteins in ribosome synthesis. *Molecular biology reports* *14*, 103-106.

Tuck, A.C., and Tollervey, D. (2013). A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell* *154*, 996-1009.

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.

Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS 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.

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.

Wasmuth, E.V., Zinder, J.C., Zattas, D., Das, M., and Lima, C.D. (2017). Structure and reconstitution of yeast Mpp6-nuclear exosome complexes reveals that Mpp6 stimulates RNA decay and recruits the Mtr4 helicase. *eLife* *6*, e29062.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* *171*, 737-738.

Webb, S., Hector, R.D., Kudla, G., and Granneman, S. (2014a). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome biology* *15*, R8.

Webb, S., Hector, R.D., Kudla, G., and Granneman, S. (2014b). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome biology* 15, R8.

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*. *The Journal of cell biology* 127, 1173-1184.

Winey, M., and Culbertson, M.R. (1988). Mutations affecting the tRNA-splicing endonuclease activity of *Saccharomyces cerevisiae*. *Genetics* 118, 609-617.

Wu, J., Delneri, D., and O'Keefe, R.T. (2012). Non-coding RNAs in *Saccharomyces cerevisiae*: what is the function? *Biochemical Society transactions* 40, 907-911.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., *et al.* (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121, 725-737.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, 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.

Yu, X., Horiguchi, T., Shigesada, K., and Egelman, E.H. (2000). Three-dimensional reconstruction of transcription termination factor rho: orientation of the N-terminal domain and visualization of an RNA-binding site. *J Mol Biol* 299, 1279-1287.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural basis of transcription initiation. *Science* 338, 1076-1080.

Zinder, J.C., and Lima, C.D. (2017). Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes & Development* 31, 88-100.

Zwiefka, A., Kohn, H., and Widger, W.R. (1993). Transcription termination factor rho: the site of bicyclomycin inhibition in *Escherichia coli*. *Biochemistry* 32, 3564-3570.

Kévin MOREAU

Etude génomique des mécanismes nucléaires de contrôle qualité et dégradation de l'ARN

La transcription des ARN messagers (ARNm), chez les eucaryotes est un processus complexe nécessitant de nombreuses étapes. En parallèle de ce mécanisme fondamentale, de nombreuses protéines vont se fixer à l'ARNm naissant afin de le maturer et de l'emballer pour former une particule ribonucléoprotéique (mRNP) compétente pour l'export vers le cytoplasme, où aura lieu la traduction. Les étapes de biogénèse de cette mRNP sont soumises à la surveillance par un système de contrôle qualité (QC) qui détectera tout évènement conduisant à une mauvaise formation de la particule. Les transcrits aberrants sont retenus au noyau pour y être dégradés par l'exosome. Afin de provoquer l'apparition massive d'ARNm aberrants dans le noyau de notre modèle d'étude, la levure *S. cerevisiae*, nous utilisons le facteur bactérien Rho. L'hélicase translocase Rho, une fois exprimée, va venir perturber l'assemblage co-transcriptionnel des mRNPs et ainsi générer des transcrits aberrants qui seront substrats de la machinerie de QC et dégradation. La présente étude étend les précédentes observations faites par l'équipe quant à l'implication de certaines protéines dans le système de QC par des approches globales (RNA-seq, ChIP-seq). De plus, l'étude du complexe THO, impliqué dans l'emballage et le transport de la mRNP, révèle l'implication d'une de ses sous-unités dans le marquage des ARNm aberrants ainsi qu'un lien de première importance dans le recrutement de l'exonucléase Rrp6 sur ces cibles à dégrader. Enfin, nous montrons, de manière globale l'existence, au sein du noyau des levures, d'une seconde voie de dégradation des ARNs aberrants distincte de la voie canonique passant par Rrp6.

Mots clés : biogénèse des mRNPs, transcription, RNA-seq, ChIP-seq, système de contrôle qualité nucléaire, Rho

Genomic study of nuclear RNA quality control and degradation machineries

Eukaryotic transcription of messenger RNAs (mRNAs) is a complex multistep process. In parallel with this fundamental mechanism, many proteins will bind to the nascent mRNA in order to process and package it to form an export competent ribonucleoprotein particle (mRNP). These mRNP biogenesis steps are under the surveillance of a quality control system (QC) that will detect all the faulty events that can lead to the formation of an aberrant particle. Aberrant transcripts will be retained in the nucleus and degraded. To study the QC mechanisms, we previously implemented a powerful assay based on the global perturbation of mRNP biogenesis by the bacterial Rho factor. When expressed in the yeast nucleus, Rho will interfere with co-transcriptional mRNP assembly and generates aberrant transcripts which will be substrates for the QC and degradation system. This study extend the previous observations made by the team about implication of some proteins in the QC pathway by genome-wide methods (RNA-seq, ChIP-seq). Moreover, study of the THO complex, which is a packaging and export factor, shows that the Tho2 subunit is involved in the tagging of aberrant transcripts and in recruitment of the exonuclease Rrp6 on its targets. Finally, we are giving insights about the presence, in yeast, of a second degradation pathway for aberrant mRNPs different from the canonical pathway involving Rrp6.

Keywords : transcription, mRNP biogenesis, RNA-seq, ChIP-seq, nuclear RNA quality control, Rho



**Centre de Biophysique Moléculaire,
CNRS UPR4301 groupe « biologie de l'ARN et
ARN thérapeutiques »
Rue Charles Sadron Orléans Cedex 2**

