



Université de Montréal

Cartographie du réseau d'interactions protéiques de la machinerie de  
transcription dans les cellules humaines

par

Natalia Rusu

Département de Biochimie

Faculté de Médecine

Mémoire présenté à la Faculté des Études Supérieures  
en vue de l'obtention du grade de Maître ès Sciences (M.Sc.)  
en biochimie

Décembre, 2010

©, Natalia Rusu, 2010

Université de Montréal  
Faculté des Études Supérieures

Ce mémoire intitulé :  
Cartographie du réseau d'interactions protéiques de la machinerie de  
transcription dans les cellules humaines

présenté par :  
Natalia Rusu

a été évalué par un jury composé des personnes suivantes :

Normand Brisson  
Président-rapporteur

Benoît Coulombe  
Directeur de recherche

Marlene Oeffinger  
Membre du jury

## RÉSUMÉ

Les protéines sont les macromolécules les plus polyvalentes de la cellule. Elles jouent un rôle fondamental dans la majorité des processus biologiques à travers la formation de complexes multi-protéiques. Durant la transcription, une multitude de facteurs sont impliquées dans le contrôle de l'activité des complexes ARN polymérase.

Notre laboratoire s'est intéressé au réseau d'interaction de la machinerie de transcription des ARN polymérase nucléaires, dans le but de mieux comprendre leurs mécanismes de régulation. Pour ce faire, une procédure protéomique comprenant la purification de complexes protéiques par affinité couplée à la spectrométrie de masse et à l'analyse bioinformatique a été développée. La méthode de purification TAP (*Tandem Affinity Purification*) a été adaptée pour permettre la purification de complexes protéiques solubles assemblés *in vivo* à partir de cellules humaines.

L'objectif de mon projet de maîtrise était de purifier le complexe de l'ARN Pol I ainsi que de poursuivre l'expansion du réseau d'interactions protéine-protéine de la machinerie de transcription de l'ARN Pol II humaine. À l'aide des protéines POLR1E, TWISTNB, POLR2E, PFDN4, MBD2, XPA, CAND1 et PDCD5 étiquetées (TAP-tag) exprimées dans des lignées cellulaires ECR-293, plusieurs complexes protéiques solubles ont été purifiés et analysés par spectrométrie de masse. Les interactions protéiques ont été triées et validées bioinformatiquement pour donner en final une liste d'interactions ayant un haut degré de confiance à partir de laquelle des réseaux d'interactions protéine-protéine ont été créés.

Le réseau créé au cours de ce projet connecte plusieurs composantes de la machinerie transcriptionnelle tels que les ARN Pol I, II et III, les complexes RPAP3/R2TP/prefoldin-like, TRiC/CCT, Mi-2/NuRD et des facteurs de transcription et de réparation de l'ADN.

Ce type d'analyse nous a permis d'identifier et de caractériser de nouveaux régulateurs de la machinerie de transcription de l'ARN Pol I et II et de mieux comprendre son fonctionnement.

*Mots clés* : Machinerie de transcription, ARN polymérase, purification TAP, partenaires d'interactions, réseau d'interactions.

## ABSTRACT

Proteins are the most versatile macromolecules of the cell. They play a fundamental role in the majority of biological processes through the formation of multiprotein complexes. During transcription, a multitude of factors are involved in the control of activity of RNA polymerases.

Our laboratory was interested in defining the nuclear RNA polymerases transcription machinery interaction network to better understand their regulatory mechanisms. To do so, a proteomic procedure that allows affinity purification of protein complexes coupled to mass spectrometry and computational data analysis was developed. The tandem affinity purification procedure has been adapted for the purification of soluble protein complexes as they likely exist in live mammalian cells.

The aim of my master project was to purify the RNA Pol I complex as well as to further pursue the expansion of the protein-protein interaction network of the human RNA Pol II transcription machinery. By using POLR1E, TWISTNB, POLR2E, PFDN4, MBD2, XPA, CAND1 and PDCD5 TAP – tagged proteins expressed in EcR-293 cell lines, multiple soluble protein complexes were purified and analyzed by mass spectrometry. Protein interactions have been sorted and validated computationally. High-confidence dataset of interactions were used to build the protein-protein interaction networks.

The network created for this project connects several components of the transcriptional machinery such as RNA Pol I, II and III, RPAP3/R2TP/prefoldin-like, TRiC/CTC, Mi-2/NuRD complexes and DNA repair and transcription factors.

This type of analysis allowed us to identify and characterize new regulators of RNA Pol I and II transcription machinery and to better understand its functioning.

*Key words:* Transcription machinery, RNA polymerase, TAP purification, interaction partners, interaction networks.

## TABLE DES MATIÈRES

RÉSUMÉ.....	iv
ABSTRACT.....	v
TABLE DES MATIÈRES.....	vi
LISTE DES TABLEAUX.....	ix
LISTE DES FIGURES.....	x
LISTE DES ABBRÉVIATIONS.....	xii
DÉDICACE.....	xiv
REMERCIEMENTS.....	xv
INTRODUCTION.....	1
1.1. La transcription : notions générales.....	2
1.2. Régulation de la transcription.....	3
1.2.1. La chromatine.....	4
1.2.2. Les modifications des histones.....	5
1.2.3. Le remodelage de la chromatine.....	5
1.3. La machinerie nucléaire de transcription.....	5
1.4. La transcription par l'ARN polymérase I.....	8
1.4.1. Le nucléole.....	8
1.4.2. L'organisation de l'ADN ribosomique.....	8
1.4.3. Le cycle de transcription de l'ARN polymérase I.....	10
1.4.3.1. L'initiation.....	10
1.4.3.2. L'élongation.....	13
1.4.3.3. La terminaison.....	14
1.5. La transcription par l'ARN polymérase II.....	15
1.5.1. Les gènes de classe II.....	15
1.5.2. Le promoteur d'un gène de classe II.....	15
1.5.3. Le cycle de transcription de l'ARN polymérase II.....	18
1.5.3.1. Le complexe de pré-initiation.....	18
1.5.3.2. L'initiation.....	22
1.5.3.3. L'élongation.....	22
1.5.3.4. La terminaison.....	23

1.6. La protéomique d'interaction à l'étude de la machinerie transcriptionnelle.....	26
1.6.1. La méthode AP-MS appliquées à la machinerie transcriptionnelle.....	26
1.7. Les chaperonne générales.....	29
1.7.1. Rôles des chaperons moléculaires dans la biogénèse des ARN Polymérase.....	30
1.8. Objectifs et stratégie.....	32
2. MATÉRIEL ET MÉTHODES.....	33
2.1. Le système de purification TAP ( <i>tandem affinity purification</i> ).....	33
2.1.1. Construction d plasmide contenant le gène d'intérêt fusionné à l'étiquette TAP.....	34
2.1.1.1.PCR.....	35
2.1.1.2.Digestion enzymatique.....	36
2.1.1.3.Ligation dans le vecteur pMZI.....	36
2.1.1.4.Transformation des bactéries compétentes.....	37
2.1.2. Création de lignées cellulaires stables contenant les protéines d'intérêt étiquetées.....	37
2.1.2.1.Transfection et sélection.....	38
2.1.2.2.Analyse d'expression inductible.....	39
2.1.2.3.Induction et préparation de l'extrait.....	39
2.1.3. Purification des complexes protéiques par la méthode TAP.....	40
2.1.3.1.La colonne d'IgG.....	40
2.1.3.2.La colonne de calmoduline.....	41
2.1.4. Identification des partenaires d'interaction par spectrométrie de masse.....	41
2.1.5. Interprétation bioinformatique et représentation graphique des interactions protéine-protéine.....	43
3. RÉSULTATS.....	45
3.1. Étude du complexe de l'ARN polymérase I.....	45
3.1.1. Purification TAP de POLR1E/hPAF53.....	45
3.1.2. Purification TAP de TWISTNB/hRPA43.....	48
3.2. Caractérisation du réseau d'interactions de la machinerie de transcription de l'ARN Pol II.....	51

3.2.1. Purification TAP de POLR2E/RPB5.....	52
3.2.2. Purification TAP de PFDN4.....	55
3.2.3. Purification TAP de MBD2 et XPA.....	58
3.2.4. Purification TAP de CAND1/TIP120A.....	63
3.2.5. Purification TAP de PDCD5/TFAR19.....	66
3.3. Représentation graphique des interactions protéine-protéine.....	69
4. DISCUSSION.....	71
4.1. Identification à haut débit des interactions protéine-protéine.....	71
4.2. Caractérisation du complexe de l'ARN Pol I.....	72
4.3. Étude protéomique de la machinerie de l'ARN Pol II.....	74
4.3.1. POLR2E/RPB5 est une sous-unité constitutive du complexe RPAP3/R2TP/prefoldin-like.....	74
4.3.2. Lien entre la méthylation des gènes et la réparation NER.....	75
4.3.3. Rôle de la Huntingtine dans le maintien de l'intégrité génétique.....	77
4.3.4. PDCD5 est un substrat physiologique de la chaperonnine TRiC/CCT...	79
5. CONCLUSION ET PERSPECTIVES.....	80
RÉFÉRENCES.....	82
ANNEXES.....	99

## LISTE DES TABLEAUX

Tableau 1-1. Les sous-unités des ARN polymérase.

Tableau 1-2. Orthologues humains de 13 des 14 sous-unités de l'ARN Pol I de *S. cerevisiae*.

Tableau 1-3. Complexes protéiques impliqués dans l'assemblage du PIC de l'ARN Pol II.

## LISTE DES FIGURES

- Figure 1-1. Schéma décrivant l'organisation des gènes ribosomiaux et leur transcrits chez l'humain.
- Figure 1-2. Facteurs principaux de l'initiation de la transcription de l'ADN ribosomique.
- Figure 1-3. Le cycle de transcription de l'ARN polymérase I.
- Figure 1-4. Structure du promoteur basal d'un gène de classe II.
- Figure 1-5. Le cycle de transcription de l'ARN polymérase II.
- Figure 2-1. Réseau mettant en évidence les interactions entre RPAPs et l'ARN Pol II.
- Figure 3-1. Schéma résumant la technique de purification TAP-tag.
- Figure 3-2. Carte du vecteur pMZI avec la version C-terminale (A) et N-terminale (B) de l'étiquette TAP.
- Figure 3-3. Schéma fonctionnel du système de régulation répondant à l'ecdysone.
- Figure 3-4. Illustration schématique de la stratégie d'identification des protéines par spectrométrie de masse.
- Figure 3-5. Schéma de validation des interactions protéiques
- Figure 4-1. Séquence d'acides aminés et représentation graphique des motifs et domaines de POLR1E.
- Figure 4-2. Analyse d'expression inductible de POLR1E.
- Figure 4-3. Purification des protéines associées à POLR1E.
- Figure 4-4. Séquence d'acides aminés et représentation graphique des motifs et domaines de TWISTNB.
- Figure 4-5. Analyse d'expression inductible de TWISTNB.
- Figure 4-6. Purification des protéines associées à TWISTNB.
- Figure 4-7. Séquence d'acides aminés et représentation graphique des motifs et domaines de POLR2E.
- Figure 4-8. Analyse d'expression inductible de POLR2E.
- Figure 4-9. Purification des protéines associées à POLR2E.
- Figure 4-10. Séquence d'acides aminés et représentation graphique des motifs et domaines de PFDN4.
- Figure 4-11. Analyse d'expression inductible de PFDN4.
- Figure 4-12. Purification des protéines associées à PFDN4.

Figure 4-13. Séquences d'acides aminés et représentations graphiques des motifs et domaines de MBD2 (A) et XPA (B).

Figure 4-14. Analyse d'expression inductible de MBD2 (A) et XPA (B).

Figure 4-15. Purification des protéines associées à MBD2 (A) et XPA (B).

Figure 4-16. Séquence et représentation graphique des motifs et domaines de CAND1.

Figure 4-17. Analyse d'expression inductible de CAND1.

Figure 4-18. Purification des protéines associées à CAND1.

Figure 4-19. Séquence et représentation graphique des motifs et domaines de PDCD5.

Figure 4-20. Analyse d'expression inductible de PDCD5.

Figure 4-21. Purification des protéines associées à PDCD5.

Figure 4-22. Réseau d'interactions connectant les ARN Pol I, II et III, les complexes RPAP3/R2TP/*prefoldin-like*, TRiC/CCT, Mi-2/NuRD et des facteurs de transcription et de réparation de l'ADN.

## LISTE DES ABBRÉVIATIONS

ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
ARNm	ARN messenger
ARN Pol	ARN polymérase
ARNr	ARN ribosomique
ARNt	ARN de transfert
Arg	Arginine
ATPase	ATP hydrolase
BRE	TFIIB recognition element
CAF-1	Chromatin assembly factor 1
CBP	CREB-binding protein et Calmodulin-binding peptide
CDK	Cyclin dependant kinase
CHD	Chromodomain SWI/SNF-like helicase/ATPase domain and DNA binding domain
ChIP	Immunoprécipitation de la chromatine
CPF	Cleavage/polyadenylation factor
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulatory factor
CTD	Carboxy-terminal domain
C-terminal	Carboxy-terminal
DCE	Downstream core element
DPE	Downstream promoter element
DSIF	DRB-sensitivity inducing factor
ELL	Eleven-nineteen lysine-rich in leukemia
ESI	Electrospray ionization
FACT	Facilitates chromatin transcription
Fcp1	TFIIF-associated CTD phosphatase
FGT	Facteurs généraux de transcription
GCN	General control of amino acid synthesis
GO	Gene ontology
HAT	Histone acétyltransférase
HDAC	Histone désacétylase
HDM	Histone déméthylase
HMG	High mobility group
HMT	Histone méthyltransférase
Htt	Huntingtine
Inr	Initiateur
ISWI	Imitation of SWI
LCR	Locus control region
MBD	Methyl-CpG binding protein
MEPCE	Methylphosphate capping enzyme
MS	Mass spectrometry
MS/MS	Double Tandem mass spectrometry
NELF	Negative elongation factor
NER	Nucleotide excision repair

N-terminal	Amino-terminal
NuRD	Nucleosome remodeling factor
PAP	Poly(A) polymérase
PCAF	p300/CBP-associated factor
PC2	Positive cofactor 2
P-TEFb	Positive transcription elongation factor b
RPAP	RNA polymerase II-associated protein
Ser	Sérine
snARN	Small nuclear RNA
snoARN	Small nucleolar RNA
snRNP	Small nuclear ribonucleoprotein
SWI	Switching mating-type
TAF	TBP-associated factor
TAP	Tandem affinity purification
TBP	TATA binding protein
TEV	Tobacco etch virus
TF	Transcription factor
TWISTNB	Twist <i>neighbor</i>
UV	Ultraviolet
XP	<i>Xeroderma pigmentosum</i>

*À ma fille, Yulianna.*

## REMERCIEMENTS

Je tiens d'abord à remercier mon directeur de recherche, Benoit Coulombe, pour son encadrement et son soutien durant ma maîtrise. Je remercie aussi les membres du jury, le Dr. Normand Brisson et le Dr. Marlene Oeffinger d'avoir accepté d'évaluer mon travail.

J'aimerais remercier Philippe Cloutier et Céline Domecq pour leur aide pratique, leurs conseils et leur support. Je remercie aussi tous les membres du laboratoire avec qui travailler a été très efficace et agréable.

Enfin, je tiens à remercier mon conjoint, Frédéric Morin, pour sa grande compréhension, sa patience et son soutien, ma famille, en particulier ma belle-mère, et mes amis.

## 1. INTRODUCTION

*« ...si le génome fournit le répertoire de chaque cellule d'un organisme, la réalité cellulaire va dépendre de l'emploi de ce répertoire en termes de choix des gènes exprimés, de niveaux d'expression des produits géniques, et de modulation de la fonctionnalité des protéines... »  
(Lescuyer et al., 2004)*

Les protéines forment la classe la plus nombreuse et la plus diversifiée des molécules biologiques. Globalement, les protéines peuvent accomplir quatre fonctions biochimiques : la fixation, la catalyse, le rôle de commutateur et celui d'élément structural. La fonction fondamentale parmi celles-ci est la fixation. Qu'il s'agisse de petites molécules ou de macromolécules, que la liaison soit stable ou dynamique, l'association protéine-protéine définit en grande partie leur fonction et est à l'origine de toutes les autres fonctions biochimiques des protéines (Petsko *et al.*, 1<sup>re</sup> éd.). L'association de plusieurs protéines en une structure physique cohérente réalisant une ou plusieurs fonctions précises forme le complexe protéique. Les complexes protéiques jouent un rôle fondamental dans la majorité des processus biologiques, depuis la formation des structures macromoléculaires et des complexes enzymatiques jusqu'à la régulation des voies métaboliques ou la transmission de signaux biologiques. À l'intérieur d'un complexe, toutes les protéines n'interagissent pas physiquement entre elles mais forme plutôt un réseau d'interactions extrêmement organisé. Pour déterminer la fonction d'une protéine, il est donc essentiel de découvrir avec quelles autres protéines celle-ci est susceptible d'interagir.

Dans le cadre de ce travail, nous nous sommes intéressés au réseau d'interaction des protéines impliquées dans la régulation de la transcription des gènes chez l'humain et à l'analyse des complexes résultant de ce dernier. Dans ce chapitre nous allons réviser les mécanismes de régulation de l'expression des gènes ainsi que les constituants et les cycles de transcription des ARN polymérases I et II.

## 1.1. La transcription : notions générales

L'information génétique nécessaire à la survie de chaque cellule est encodée dans la structure de l'ADN (acide désoxyribonucléique), molécule dont la forme native est une double hélice formée de deux chaînes antiparallèles, de séquence nucléotidique complémentaire (adénine-thymine et guanine-cytosine). Chez les eucaryotes l'ADN est localisé essentiellement dans le noyau cellulaire (ADN nucléaire) et, pour une petite fraction, dans les mitochondries et les chloroplastes (ADN extranucléaire). L'ADN nucléaire est distribué en chromosomes et condensé sous forme de chromatine (Lodish *et al.*, 3<sup>e</sup> éd.). Si l'ADN constitue le support de l'information génétique, alors les gènes représentent les unités physiques et fonctionnelles élémentaires de l'hérédité. Au niveau moléculaire, un gène correspond à un fragment d'ADN constitué d'une région transcrite sous forme d'ARN (acide ribonucléique) et d'une séquence régulatrice qui permet la transcription (Griffiths *et al.*, 3<sup>e</sup> éd.).

La transcription est la première étape de l'expression des gènes et est réalisée par les ARN polymérases ADN dépendantes à partir d'un seul brin de la double hélice d'ADN. Le mécanisme général de la transcription se décompose en trois étapes distinctes: l'initiation, l'élongation et la terminaison. Durant l'initiation l'ARN polymérase est recrutée via différents facteurs au promoteur, site de l'initiation de la transcription, permettant ainsi la mise en place du complexe de pré-initiation et l'amorçage de la transcription. L'élongation consiste en l'addition de l'ensemble des nucléotides complémentaires de la molécule d'ADN au niveau de la molécule d'ARN suivant la règle d'appariement Watson-Crick. Le mécanisme de terminaison permet l'arrêt de la transcription et la libération de l'ARN polymérase et de l'ARN néosynthétisé. Il varie en fonction du type d'enzyme et peut impliquer la présence de sites de pause en aval d'un signal de polyadénylation ou des facteurs de terminaison (Lodish *et al.*, 3<sup>e</sup> éd.; Langelier *et al.*, 2002).

## **1.2. Régulation de la transcription**

La transcription est un processus extrêmement régulé dont le bon déroulement est nécessaire au fonctionnement d'une cellule. La régulation de l'expression des gènes s'effectue non seulement au niveau de la machinerie transcriptionnelle elle-même mais également au niveau de la structure chromatinienne en modulant l'accessibilité de l'ADN à la machinerie de transcription.

### **1.2.1. La chromatine**

Dans le noyau, l'ADN eucaryote s'associe avec des histones et une série d'autres protéines pour former une structure appelée chromatine. Cette structure peut présenter différents niveaux de compaction, notamment un niveau moins dense, l'euchromatine, et un niveau plus compact, l'hétérochromatine. Seule l'euchromatine est accessible à la machinerie transcriptionnelle. L'unité de base de la chromatine est le nucléosome où 146 pb d'ADN s'enroulent 1,65 fois autour d'un octamère d'histones (deux molécules H2A, H2B, H3 et H4) formant le cœur du nucléosome. Stabilisée par l'histone H1, cette structure se répète régulièrement tout le long de la molécule d'ADN afin de former un nucléofilament d'une épaisseur de 10 nm appelé « collier de perles ». Le nucléofilament se compacte à son tour en spirale au rythme de 6 nucléosomes par hélice produisant une fibre de 30 nm appelée solénoïde. Les fibres de chromatine peuvent ensuite se replier pour former une structure de compaction ultime de 100-300 nm observée au niveau du chromosome lors de la division cellulaire.

La chromatine constitue, donc, un obstacle pour la transcription et sa structure est soumise à des modifications qui influencent l'accessibilité de l'ADN aux activateurs et aux facteurs de transcription et qui facilitent l'initiation de la transcription. Cette dynamique chromatinienne repose principalement sur deux types de mécanismes agissant en synergie : les modifications des histones et le remodelage de la chromatine.

### 1.2.2. Les modifications des histones

Les histones sont les protéines principales de la chromatine. Leurs régions amino-terminales très conservées sont riches en acides aminés basiques permettant une forte interaction avec l'ADN. Situées à l'extérieur du nucléosome, ces domaines peuvent subir de nombreuses modifications post-traductionnelles: acétylation, méthylation, phosphorylation, ubiquitinylation, ADP-ribosylation, sumoylation, glycosylation et biotinylation (Khorasanizadeh, 2004; Kouzarides, 2007). Certaines de ces modifications sont associées à l'activation de la transcription comme l'acétylation des histones H3 et H4 ou la di- ou triméthylation de la lysine 4 de l'histone H3. Par contre, les rôles respectifs des autres types de modifications ne sont pas aussi clairs. Il a été suggéré que l'ensemble des modifications post-traductionnelles des histones formerait un code, appelé le «code histone», qui permettrait d'associer à chaque combinaison de modifications un état transcriptionnel de la chromatine. Ces modifications agiraient de façon séquentielle ou combinatoire pour produire une réponse biologique unique (Strahl et Allis, 2000). De nombreuses enzymes nucléaires participent à ces modifications, par exemple, histones acétyltransférases vs déacétylases (HAT/HDAC), histones méthyltransférases vs déméthylases (HMT/HDM), kinases vs phosphatases, ubiquitine-ligases, SUMO-ligases, *etc.*

### 1.2.3. Le remodelage de la chromatine

La deuxième catégorie de modifications de la structure chromatinienne implique les complexes de remodelage. Il s'agit de complexes multi-protéiques qui ont tous en commun la présence d'une sous-unité catalytique à activité ATPase appartenant à la superfamille des ADN hélicases/ATPases *Swi2/SNF2*. Chez l'humain, ces différents complexes de remodelage sont classés en trois groupes selon la nature de domaines structuraux caractéristiques de la sous-unité catalytique: i) SWI/SNF (bromodomaine reconnaissant les lysines acétylées); ii) ISWI (domaines SANT et SLIDE liant les queues des histones et l'ADN «*linker*» respectivement); et iii) MI-2/CHD (deux chromodomains liant l'ADN et l'histone H3 méthylée). Le résultat du remodelage est une altération de la structure et/ou de la position des nucléosomes au sein de la

chromatine, par le biais de dissociation, glissement ou relocalisation des nucléosomes individuels. Les complexes SWI/SNF jouent un rôle dans l'activation transcriptionnelle, contrairement aux complexes appartenant aux deux autres groupes, ISWI et Mi-2/CHD, qui sont principalement impliqués dans la répression transcriptionnelle.

### 1.3. La machinerie nucléaire de transcription

Dans les noyaux des organismes eucaryotes la transcription est catalysée par trois ARN polymérases distinctes :

- i) l'ARN polymérase I transcrit les gènes de classe I qui codent pour les ARN ribosomiaux (ARNr) 5.8S, 18S et 28S;
- ii) l'ARN polymérase II transcrit les gènes de classe II qui codent pour l'ensemble des ARN messagers (ARNm) et certains des petits ARN nucléaires ARNsn (small nuclear);
- iii) l'ARN polymérase III transcrit les gènes de classe III qui codent pour la plupart des petits ARN : certains ARNsn, les ARN de transfert (ARNt), l'ARNr 5S et l'ARN 7SL.

Les trois polymérases diffèrent par leur localisation (l'ARN Pol I est localisée dans le nucléole, tandis que les ARN Pol II et III sont situées dans le nucléoplasme), par leur profil d'élution (les enzymes sont séparées dans l'ordre de I à III à l'aide d'un gradient croissant de concentration de sel sur colonne DEAE Sephadex) et par leur sensibilité à l' $\alpha$ -amanitine (la polymérase I est insensible à l' $\alpha$ -amanitine, la polymérase II – très sensible et la polymérase III a une sensibilité intermédiaire) (revu par Thomas et Chiang, 2006).

Malgré leurs propriétés différentes, les enzymes sont similaires du point de vue de l'activité catalytique et de leur structure. Ce sont de gros complexes multimériques chacun comportant entre 12 et 17 sous-unités ayant des masses comprises entre 500 et 600 kDa (**Tableau 1-1**). Les ARN polymérases eucaryotes les mieux caractérisées proviennent de la levure *S. cerevisiae*. Bien que leurs compositions globales soient différentes, les enzymes ont une forte homologie structurale avec l'enzyme procaryote.

Les deux plus grosses sous-unités sont en fait des homologues des sous-unités bactériennes  $\beta$  et  $\beta'$  renfermant le site catalytique. Les heterodimères AC40-AC19 de Pol I et Pol III et RPB3-RPB11 de Pol II sont des homologues fonctionnels des sous-unités bactériennes  $\alpha 1$  et  $\alpha 2$ , tandis que la sous-unité RPB6 (ABC23), commune aux trois polymérases, est un homologue structurel et fonctionnel de la sous-unité  $\omega$ . En plus de leurs sous-unités centrales apparentées aux sous-unités de l'ARN polymérase procaryote, les trois ARN polymérases eucaryotes contiennent quatre sous-unités supplémentaires qui leur sont communes (RPB5, RPB8, RPB10 et RPB12) (revu par (Langelier *et al.*, 2002; Cramer *et al.*, 2008; Kwapisz *et al.*, 2008)).

De plus, chaque polymérase eucaryote possède plusieurs sous-unités spécifiques. Les sous-unités A12.2, RPB9, et C11 des polymérases I, II, et III, respectivement, impliquées dans l'activité de clivage de l'ARN (Awrey *et al.*, 1997; Chedin *et al.*, 1998), complètent le noyau structurel des enzymes. Les ARN polymérases I, II et III contiennent chacune un hétérodimère (Rpb4/Rpb7, A14/A43 et C17/C25, respectivement) partageant caractéristiques génétiques, biochimiques et structurelles, ainsi qu'homologie de séquence entre RPB7, A43 et C25. Il y a, aussi, deux sous-unités spécifiques supplémentaires pour l'ARN Pol I (A49 et A34.5) et cinq sous-unités spécifiques pour l'ARN Pol III (C37, C53, C82, C34 et C31). Les hétérodimères A49/A34.5 dans l'ARN Pol I et C37/C53 dans l'ARN Pol III montrent une homologie partielle avec les sous-unités Rap74 et Rap30 du facteur d'élongation TFIIF, impliqué dans l'élongation et le recyclage de l'ARN Pol II. Les sous-unités C82, C34 et C31 de l'ARN Pol III forment un trimère et semblent être importantes dans l'initiation de la transcription (Werner *et al.*, 1992, 2009). Contrairement à l'ARN Pol III où toutes les sous-unités sont essentielles, les deux autres ARN polymérases eucaryotes possèdent quelques sous-unités non essentielles à la croissance: Rpa49, Rpa34, Rpa14 et Rpa12 pour l'ARN Pol I ; Rpb4 et Rpb9 dans le cas de l'ARN Pol II (Bischler *et al.*, 2002; Hull *et al.*, 1995).

**Tableau 1-1. Les sous-unités des ARN polymérase.** (Adapté de Cramer *et al.*, 2008)

ARN polymérase	<i>E. coli</i>	<i>S. cerevisiae</i>			
		I	II	III	type de sous-unité
Noyau structurel	$\beta'$	A190	Rpb1	C160	homologue
	$\beta$	A135	Rpb2	C128	homologue
	$\alpha 1$	AC40	Rpb3	AC40	homologue
	$\alpha 2$	AC19	Rpb11	AC19	homologue
	$\omega$	A12.2	Rpb9	C11	homologue
		Rpb5 (ABC27)	Rpb5	Rpb5	commune
		Rpb6 (ABC23)	Rpb6	Rpb6	commune
		Rpb8 (ABC14.5)	Rpb8	Rpb8	commune
		Rpb10 (ABC10 $\alpha$ )	Rpb10	Rpb10	commune
		Rpb12 (ABC10 $\beta$ )	Rpb12	Rpb12	commune
Sous-complexe Rpb4/7		A14	Rpb4	C17	paralogue
		A43	Rpb7	C25	paralogue
Sous-complexe TFIIIF-like		A49	(Tfg1/Rap74)	C37	spécifique
		A34.5	(Tfg2/Rap30)	C53	spécifique
Sous-complexe spécifique de Pol III		–	–	C82	spécifique
		–	–	C34	spécifique
		–	–	C31	spécifique
Nombre de sous-unités	5	14	12	17	–

Les ARN polymérase sont les facteurs permettant la transcription, cependant, à elles seules ces enzymes sont peu efficaces. En effet, les polymérase n'ont qu'une petite, voir aucune capacité intrinsèque de lier l'ADN. Pour fonctionner, elles ont besoin de l'aide et de la coopération de facteurs protéiques supplémentaires, les facteurs de transcription. Ils permettent la reconnaissance du site d'initiation, mais agissent aussi sur l'élongation. Chaque polymérase nécessite un ensemble de facteurs de transcription propre pour initier la transcription à leurs promoteurs respectifs. Dans le cadre de ce mémoire, nous nous sommes intéressés plus particulièrement au fonctionnement des ARN polymérase I et II humaines, moins bien connues que leur contrepartie chez la levure.

## **1.4. La transcription par l'ARN polymérase I**

La synthèse de l'ARNr par l'ARN polymérase I est la première étape dans la biogenèse du ribosome et a un rôle central dans la régulation de la croissance cellulaire chez les eucaryotes. Dans une cellule, l'ARNr compte 50% de tous les ARN naissants et l'activité de l'ARN polymérase I compte pour 60% de l'ensemble de la transcription nucléaire (Grummt, 2003; Moss *et al.*, 2007; Russell et Zomerdijk, 2005).

### **1.4.1. Le nucléole**

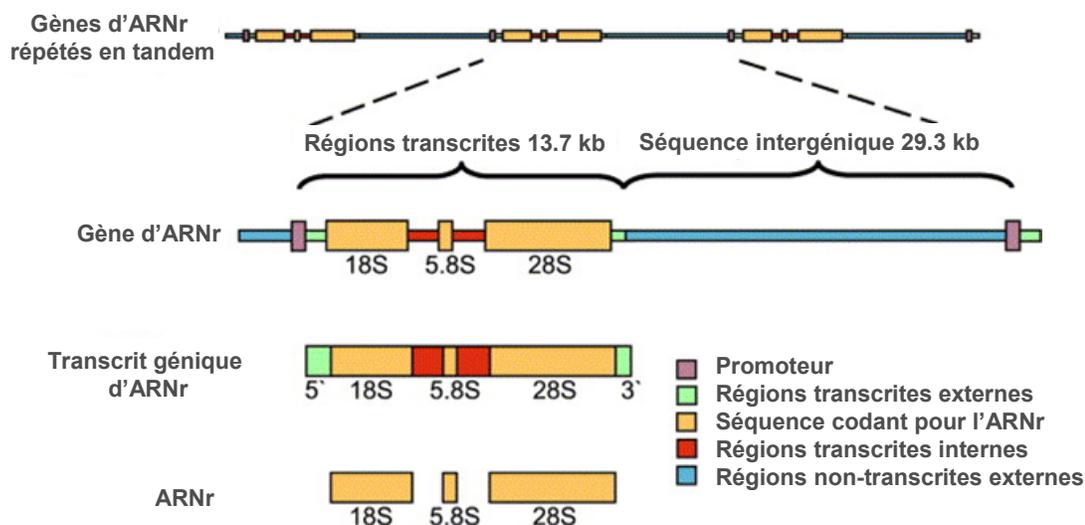
La biogenèse des ribosomes se déroule au sein d'une structure nucléaire transitoire, le nucléole. Ce dernier regroupe les gènes actifs codant pour les ARNr ainsi que des particules contenant les ARNr à différents stades de maturation et d'assemblage avec les protéines ribosomiques. Le nucléole est organisé en trois composants majeurs visibles en microscopie électronique: le centre fibrillaire (FC), le composant fibrillaire dense (DFC) et le composant granulaire (GC). Chacun de ces compartiments se voit attribuer un rôle dans la transcription ou la maturation des ARNr. Le FC correspond aux régions promotrices des gènes d'ADNr et est le lieu de stockage de l'ARN polymérase I sous forme inactive. À l'interface entre le FC et le DFC se déroule l'initiation de la transcription et l'élongation de l'ARNr, tandis que la maturation des ARNr est déclenchée dans le DFC et se poursuit durant la migration vers le GC. Enfin, le GC sert de lieu de pré-assemblage des sous-unités ribosomiques, principalement pour la sous-unité 60S. Par contre, aucun domaine particulier dans ce compartiment ne correspond à la sous-unité 40S, probablement dû au fait que la dernière étape d'assemblage se produit dans le cytoplasme. (Hernandez-Verdun et Louvet, 2004; Sirri *et al.*, 2008).

### **1.4.2. L'organisation de l'ADN ribosomique**

Chez les eucaryotes supérieurs, on retrouve plusieurs centaines de gènes codant pour l'ARNr répartis sur plusieurs chromosomes. Le nombre de copies et leur activité transcriptionnelle est modulée en fonction des conditions de croissance. Les gènes actifs en transcription peuvent être visualisés par microscopie électronique sur des étalements

de Miller (Miller et Beatty, 1969), l'organisation dite en «arbre de Noël» de la synthèse des ARNr. Ces gènes sont regroupés en unités répétée en tandem, appelés régions organisatrices du nucléole ou NOR (*Nucleolar Organizer Regions*). Une unité d'ADNr mesure 43 kb et contient une séquence codante (l'unité transcriptionnelle) de 13,7 kb et une séquence intergénique de 29,3 kb. Chaque unité transcriptionnelle est constituée de gènes correspondant aux ARN 18S, 28S et 5,8S et de séquences régulatrices qui sont transcrits par l'ARN polymérase I sous forme d'un ARNr précurseur 47S (**Figure 1-1**). Ces unités sont séparées par des régions intergéniques non transcrites contenant les séquences régulatrices en cis de la transcription : le promoteur proximal, le promoteur distal, les amplificateurs (enhancer-like sequences) et les séquences de terminaison de la transcription.

Le promoteur proximal permet la synthèse du précurseur de l'ARNr. Il est constitué d'un domaine appelé Core ou CPE (*Core Promoter Element*) qui contient le site d'initiation de la transcription et d'une région dite UCE (*Upstream Control Element*) dispensable à la transcription *in vitro*, mais qui permet de stabiliser l'interaction des facteurs de transcription avec le CPE *in vivo* (Russell et Zomerdijk, 2005). Ces éléments sont situés chez l'humain entre les résidus -45 et +18 et -156 et -107 respectivement.



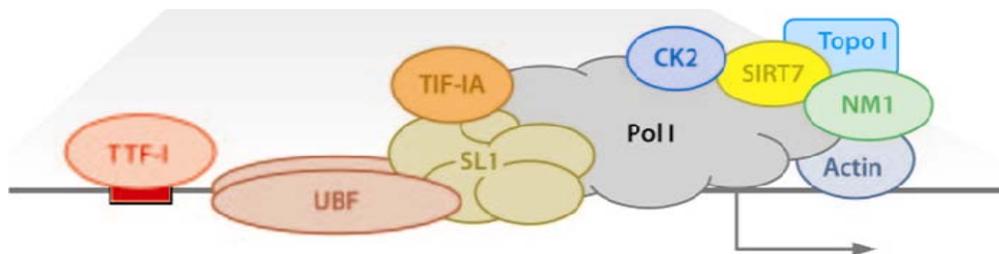
**Figure 1-1.** Schéma décrivant l'organisation des gènes ribosomiaux et leurs transcrits chez l'humain. (Raska *et al.*, 2004)

### 1.4.3. Le cycle de transcription de l'ARN polymérase I

#### 1.4.3.1. L'initiation

L'initiation de transcription de l'ADNr requiert l'assemblage d'un complexe multi-protéique spécifique, le complexe de pré-initiation (*PIC : Pre-Initiation Complex*), qui permet le recrutement de l'ARN Pol I au niveau du promoteur. Chez les mammifères, l'action synergique de deux facteurs de transcription est requise pour l'initiation de la transcription : UBF (*Upstream Binding Factor*), le facteur liant en amont, et SL1 (*Selectivity factor 1*), le facteur de sélectivité-1, ou TIF-IB chez la souris (**Figure 1-2**).

UBF intervient dans la première étape de l'assemblage du complexe d'initiation de la transcription de l'ARN polymérase I. Cette protéine comporte six boîtes HMG (*High mobility group*) qui vont lui permettre de lier l'ADN. UBF lie sous forme d'homodimère le sillon mineur de l'ADN pour former une boucle de 140 pb qui se replie sur elle-même, rapprochant de cette façon les éléments Core et UCE (Grummt, 2003). UBF existe sous deux formes nommées UBF1 et UBF2. La mise en place du complexe transcriptionnel sur le promoteur est initiée par la phosphorylation en C-terminal d'UBF. UBF1 phosphorylé recrute SL1, qui reconnaît le CPE. SL1 est un complexe multiprotéique constitué d'une protéine TBP (*TATA Binding Protein*) et cinq facteurs associés TAF<sub>1</sub>s (*TATA binding protein-Associated Factors*) : TAFI110/95, TAFI63, TAFI48, TAFI35, and TAFI12 (Denissov *et al.*, 2007; Gorski *et al.*, 2007). UBF stabilise SL1 sur le promoteur *via* des interactions entre son extrémité C-terminale et TAFI48 d'une part, et TBP, d'autre part. SL1 recrute à son tour l'ARN Pol I sur le promoteur grâce à des interactions entre TAFI68, TAFI110 et le facteur RRN3/TIF-IA. UBF lie aussi l'ARN Pol I via les sous-unités CAST et PAF53. L'acétylation de l'UBF est importante pour le recrutement de l'ARN Pol I, car elle stimule l'interaction avec PAF53.



**Figure 1-2. Facteurs principaux de l'initiation de la transcription de l'ADN ribosomique.** (Adapté de Drygin *et al.*, 2010)

D'autres facteurs interviennent à l'étape de l'initiation comme le facteur de terminaison TTF-1, qui est capable d'activer la transcription des gènes ribosomique dans un contexte chromatinien, via sa capacité à déstabiliser le nucléosome dans la région promotrice, tout en augmentant l'accessibilité des facteurs de transcription (Sirri *et al.*, 2008). La **Figure 1-3** illustre un schéma des étapes de transcription par l'ARN Pol I.

L'ARN Pol I est la composante la plus complexe du PIC. Chez la levure, elle comprend 14 sous-unités et il y a des homologues mammifères pour toutes les sous-unités sauf pour RPA14 (Russell et Zomerdijk, 2005; Panov *et al.*, 2006) (**Tableau 1-2**). Chez les mammifères, l'ARN Pol I existe sous deux formes distinctes,  $\alpha$  et  $\beta$  dans une proportion d'environ 9:1, respectivement. Les deux formes sont actives du point de vue catalytique, mais seulement l'ARN Pol I  $\beta$  peut s'assembler dans des complexes d'initiation de transcription productifs. Le facteur RRN3 est essentiel à la formation du complexe productif, mais beaucoup d'autres protéines, y compris la protéine kinase CK2, les complexes de remodelage de la chromatine PCAF et G9a, l'actine et la myosine (NM1) nucléaires, ainsi que des facteurs de réparation et de réplication de l'ADN, comme les topo-isomérase I et IIa, Ku70/80, PCNA, TFIIH, et CSB, ont été trouvées en association avec l'ARN Pol I  $\beta$  (Drygin *et al.*, 2010). Ces données sont compatibles avec un mécanisme par lequel l'ARN Pol I est recrutée sur le promoteur des gènes ribosomiques en tant que complexe multiprotéique massif pré-assemblé en une holoenzyme, contenant les protéines nécessaires à la synthèse et maturation de l'ARNr, modification de la chromatine et réparation de l'ADN. Néanmoins, ce concept a été récemment remis en question par une étude décrivant la dynamique du complexe de

l'ARN Pol I *in vivo* via la technique de FRAP (*Fluorescence Recovery After Photobleaching*). Les observations suggèrent que l'assemblage de l'ARN Pol I et des complexes d'initiation productifs peut découler de manière séquentielle via des intermédiaires métastables, dont la stabilité s'accroît au fur et à mesure que les sous-unités s'ajoutent (Dundr *et al.*, 2002).

**Tableau 1-2. Orthologues humains de 13 des 14 sous-unités de l'ARN Pol I de *S. cerevisiae*** (Adapté de Panov *et al.*, 2006)

Sous-unités de l'ARN Polymérase I			
<i>S. cerevisiae</i>	PM (kDa)	Unique ou partagé	Orthologue humain
RPA190 (A190)	186.4	I	hRPA190 (A190 or A194)
RPA135 (A135)	135.7	I	hRPA135 (A127)
RPA49 (A49)	46.7	I	hRPA49 (hPAF53/POLR1E)
RPA43 (A43)	36.2	I	hRPA43 (A43/TWIST neighbor)
RPA40 (AC40)	37.7	I, III	hRPA40 (AC40/hRPA5)
RPA34.5 (A34.5)	26.9	I	CAST (ASE-1/ hPAF49/ CD3EAP)
RPB5 (ABC27)	25.1	I, II, III	hRPB5
RPB6 (ABC23)	17.9	I, II, III	hRPB6
RPA19 (AC19)	16.2	I, III	hRPA19 (AC19)
RPB8 (ABC14.5)	16.5	I, II, III	hRPB8
RPA14 (A14)	14.6	I	NI
RPA12.2 (A12.2)	13.7	I	hRPA12.2 (ZNRD1)
RPB10 (ABC10 $\beta$ )	8.3	I, II, III	hRPB10
RPB12 (ABC10 $\alpha$ )	7.7	I, II, III	hRPB12

NI – non identifié

Le recrutement de l'ARN Pol I par RRN3 sur le PIC conduit à l'ouverture du double brin d'ADN au niveau du promoteur suivie par la formation des premières liaisons phosphodiester et l'initiation de la transcription. Lorsque la taille de l'ARN néosynthétisé atteint une dizaine de nucléotides, l'ARN Pol I se détache du PIC et quitte le promoteur pour s'engager dans l'étape d'élongation. Cette sous-étape, appelée dégagement du promoteur, constitue l'étape limitante de la transcription des gènes de classe I.

### 1.4.3.2. L'élongation

L'élongation par l'ARN Pol I est un processus extrêmement efficace. En fait, lorsque l'ARN Pol I se détache et quitte le promoteur pour entamer l'élongation, le facteur RRN3/TIF-IA est relâché, tandis qu'UBF et SL1 restent liés au promoteur, prêts à recruter le prochain complexe d'ARN Pol I $\beta$  et à ré-initier la transcription à partir du même promoteur, soutenant ainsi plusieurs tours de transcription.

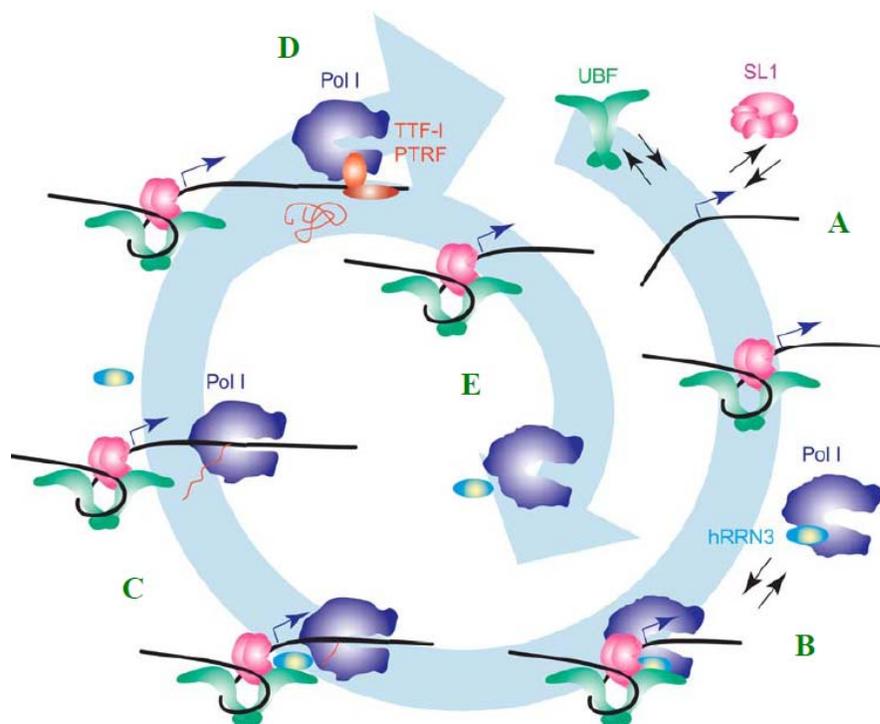
*In vitro*, l'élongation sur une matrice d'ADN dénudée découle généralement sans aucune entrave. En fait, il a été démontré que les deux sous-unités spécifiques non essentielles de l'ARN Pol I, RPA34 et RPA49, forment un hétérodimère au sein de l'enzyme qui sert de facteur d'élongation intégré similaire au facteur TFIIF de l'ARN Pol II (Beckouet *et al.*, 2008; Kuhn *et al.*, 2007).

*In vivo*, la vitesse de l'ARN Pol I en cours d'élongation peut être modulée par différents facteurs liés à l'ADNr qui influent sur la structure de la chromatine (Moss *et al.*, 2007). Plusieurs études démontrent l'implication de divers facteurs agissant sur le remodelage de la chromatine dans l'élongation par l'ARN Pol I. UBF1, qui lie de manière non spécifique l'ADNr sur toute la longueur, module l'élongation par le biais de sa phosphorylation (Stefanovsky *et al.*, 2006). La topo-isomérase I, dont l'activité catalytique consiste à relaxer la chromatine, agit en éliminant les surenroulements positifs et négatifs autour du complexe d'élongation (Rose *et al.*, 1988). Le complexe FACT (*FAcilitates Chromatin Transcription*), qui agit en tant que chaperonne d'histones, facilite l'élongation en déstabilisant la structure du nucléosome (Birch *et al.*, 2009).

Des études récentes suggèrent aussi l'implication des facteurs d'élongation de l'ARN Pol II dans l'élongation par l'ARN Pol I comme le complexe Paf1 (Zhang *et al.*, 2009), les facteurs Spt4/5 (DSIF chez l'humain) (Schneider *et al.*, 2006), TFIIF et CSB (Bradsher *et al.*, 2002; Iben *et al.*, 2002; Grummt, 2007).

### 1.4.3.3. La terminaison

Les sites de terminaison de la transcription T1 à T10 sont localisés en aval de la région transcrite, entre les promoteurs distal et proximal. TTF-I reconnaît le site de terminaison riche en thymine, bloque physiquement l'élongation par l'ARN Pol I et coopère avec PTRF (*Pol I and transcript release factor*) pour dissocier l'ARN Pol I et le transcrit de l'ADNr (Jansa et Grummt, 1999; Russell et Zomerdijk, 2005). Se référer à la **Figure 1-3** pour un schéma récapitulatif des étapes de transcription par l'ARN Pol



**Figure 1-3. Le cycle de transcription de l'ARN polymérase I.** (A) L'initiation de la transcription par l'ARN Pol I débute avec la reconnaissance du promoteur par les facteurs UBF et SL1/TIF-IB. (B) RRN3/TIF-IA amène l'ARN Pol I au niveau du site d'initiation. (C) Le départ de RRN3/TIF-IA est nécessaire pour que l'ARN Pol I puisse entrer en phase d'élongation. (D) TTF-I et PTRF sont des facteurs de terminaison qui permettent le recyclage de l'ARN Pol I. (E) Pendant l'élongation, UBF et SL1 restent liés au promoteur ce qui permet d'initier un nouveau cycle de transcription. (Adapté de Russell et Zomerdijk, 2005)

## 1.5. La transcription par l'ARN polymérase II

Trois classes d'ARN participent à la synthèse protéique : les ARN ribosomiaux (ARNr), les ARN de transfert (ARNt) et les ARN messagers (ARNm). Cependant, seuls les ARNm renferment l'information nécessaire à la synthèse des protéines.

### 1.5.1. Les gènes de classe II

Les ARN messenger (ARNm) sont transcrits par l'ARN polymérase II à partir des gènes de classe II. Chaque gène comprend des séquences codantes qui seront présentes dans l'ARNm mature (exons), des régions qui en seront exclues (introns), ainsi que des séquences régulatrices. Ces dernières, se caractérisent par la présence de différents éléments, dont la plupart est localisée à proximité du site où commence la transcription, le promoteur, mais également à plusieurs centaines de kb, en amont ou en aval de celui-ci, les séquences régulatrices distales (Maston *et al.*, 2006). Ces éléments, appelés facteurs cis-régulateurs, contiennent les sites de reconnaissance pour des facteurs de transcription, ou facteurs trans-régulateurs, qui lient l'ADN et contrôlent la production du messenger en fonction du type cellulaire ou du type d'environnement nécessitant cette transcription.

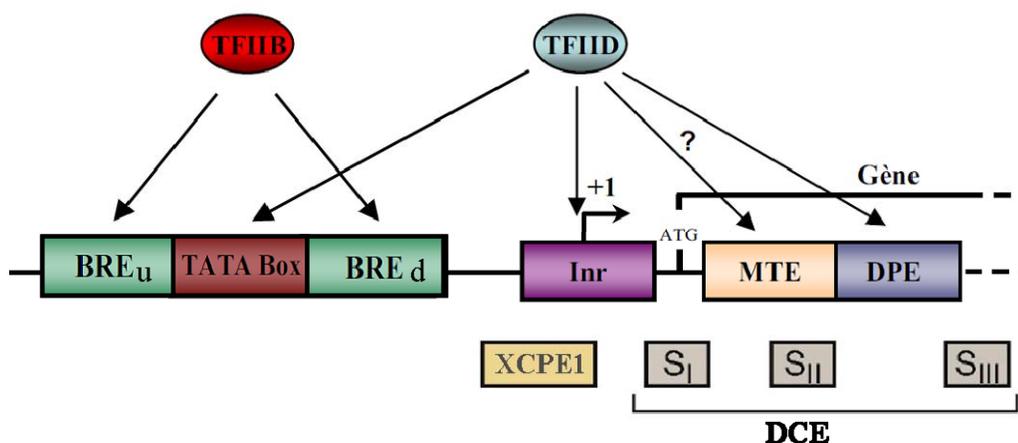
### 1.5.2. Le promoteur d'un gène de classe II

Chez les eucaryotes, le promoteur d'un gène de classe II se décompose en deux parties principales, le promoteur basal et le promoteur proximal. Le promoteur basal sert de site d'ancrage pour la machinerie transcriptionnelle de base formant le complexe de pré-initiation et définit la position du site d'initiation de la transcription (SIT) (site +1), de même que, la direction de la transcription. Les différents éléments du promoteur basal connus à ce jour incluent la boîte TATA, l'initiateur (Inr), le DPE (*Downstream core Promoter Element*), le BRE (*TFIIB Recognition Element*), le MTE (*Motif Ten Element*), le DCE (*Downstream Core Element*) et le XCPE1 (*X Core Promoter Element 1*) (**Figure 1-4**).

La boîte de TATA, généralement localisée entre -28 et -33 pb, est l'élément du promoteur le mieux connu (Carninci *et al.*, 2006). Sa séquence consensus, TATA[T/A]AA[G/A], est reconnue et liée par la sous-unité TBP (*TATA Binding Protein*) du facteur général de transcription TFIID, qui fait partie du complexe de pré-initiation. La boîte TATA est un composant autonome du promoteur et suffisant pour initier la transcription. Elle est souvent associée avec des promoteurs forts tissu-spécifiques et peut agir de concert avec Inr pour initier la transcription (Baumann *et al.*, 2010).

L'initiateur (Inr) est localisé de part et d'autre du site d'initiation. La séquence consensus est PyPyA<sub>(+1)</sub>N(T/A)PyPy où A<sub>(+1)</sub> correspond au premier nucléotide transcrit par l'ARN Pol II (Carninci *et al.*, 2006). Inr est reconnu et lié spécifiquement par le complexe TFIID via les sous-unités TAF1 et TAF2 (TAF : *TBP Associated Factors*). Il peut permettre l'initiation de la transcription seul ou en association avec la boîte TATA ou DPE (Smale et Kadonaga, 2003)

DPE est présent uniquement dans les promoteurs sans boîte TATA mais fonctionne en coopération avec Inr. Situé en aval du site +1, entre +28 et +32 pb, sa séquence consensus [A/G]G[A/T][C/T][G/A/C] permet le recrutement de TFIID via les sous-unités TAF6 et TAF9 (Juven-Gershon *et al.*, 2008). Le régulateur de transcription NC2 est également capable de reconnaître DPE. NC2 est un régulateur négatif de la transcription lorsqu'il se lie à l'élément TATA en l'absence d'Inr ; en revanche, lié au DPE, il active l'initiation de la transcription (Smale et Kadonaga, 2003).



**Figure 1-4. Structure du promoteur basal d'un gène de classe II.**

MTE (séquence consensus C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C), est un élément du promoteur basal qui a été découvert récemment et auquel TFIID pourrait aussi se lier. Localisé en aval du SIT, entre +18 et +29, il agit conjointement avec Inr afin d'augmenter le taux basal de transcription (Thomas et Chiang, 2006). L'addition d'un MTE peut compenser la perte d'activité de DPE et/ou de la boîte TATA. De plus, MTE peut agir en synergie avec ces deux éléments et ce d'une manière Inr-dépendante. Il a récemment été démontré que cette synergie entre les éléments TATA, Inr, MTE et DPE pouvait former un super promoteur (*Super Core Promoter*) pouvant atteindre un taux de transcription plus élevé que celui du cytomegalovirus IE1 (CMVIE1) ou de l'adénovirus majeur tardif (AdML) (Juven-Gershon *et al.*, 2008).

DCE contient trois sous-éléments discontinus : SI, SII, et SIII, avec une séquence promotrice de CTTC, CTGT, et AGC, respectivement, s'étalant de +6 à + 34. Comme DPE, DCE agit préférentiellement lorsque le promoteur comporte un élément Inr et est aussi reconnu et lié par TFIID. Par contre, DPE et DCE ne sont pas reconnus par les mêmes TAFs (TAF1 dans le cas de DCE). De plus, la présence de DCE et DPE semble être mutuellement exclusive (Lee *et al.*, 2005; Baumann *et al.*, 2010).

XCPE1 a été récemment mis en évidence dans le promoteur du gène X du virus de l'hépatite B et dans les promoteurs eucaryotes. Ce motif, localisé de -8 à +2 par rapport au SIT et ayant la séquence consensus (G/A/T)(G/C)G(T/C)GG(G/A)A(G/C<sub>+1</sub>)(A/C), ne se retrouve que dans environ 1% des promoteurs humains, la plupart ne comportant pas de boîte TATA. Cet élément aurait une faible activité transcriptionnelle constitutive, mais une activité beaucoup plus grande lorsque combiné à d'autres activateurs comme NRF1, NF-1, et Sp1. Son activité pourrait se faire suivant deux modes: soit par TFIID via les sous-unités TAF, soit, par la sous-unité TBP seule (Juven-Gershon *et al.*, 2008).

BRE est le seul élément connu du promoteur basal qui est reconnu par un facteur autre que TFIID, soit TFIIB. Initialement, BRE a été identifié comme une séquence liant TFIIB située immédiatement en amont de la boîte TATA (BREu). Plus tard, la présence d'un second BRE, localisé cette fois-ci en aval de la boîte TATA (BREd), a été mise en évidence. Les séquences consensus de BREu et BREd sont, respectivement, (G/C)(G/C)(G/A)CGCC et (G/A)T(T/G/A)(T/G)(G/T)(T/G)(T/G). Contrairement à BREd, dont la liaison au TFIIB est dépendante de TBP, BREu peut lier le facteur TFIIB

de manière indépendante de TBP et son rôle est d'orienter le complexe de pré-initiation sur le promoteur. Dépendamment du contexte du promoteur, BREu et BREd peuvent avoir un effet positif ou négatif sur la transcription (Smale et Kadonaga, 2003; Baumann *et al.*, 2010).

Les éléments du promoteur proximal incluent la boîte CCAAT et la boîte GC. La boîte CCAAT se situe en amont de la boîte TATA, entre la position -120 et -80 du site de démarrage de la transcription. Plusieurs facteurs peuvent interagir avec, comme CTF (*CCAAT binding Transcription Factor*), C/EBP (*CCAAT/Enhancer Binding Protein*) ou encore NF-Y (*Nuclear Factor-Y*) aussi appelé CBF (*CCAAT-Binding Factor*) (Mantovani, 1999). Quelque fois, une boîte GC, riche en guanine et cytosine, est retrouvée également. Elle est appelée aussi la *Sp1-box*, car elle permet la fixation du facteur de transcription Sp1. La boîte GC présente une localisation moins bien définie, située généralement entre la boîte CCAAT et la boîte TATA. Elle forme 6 répétitions en tandem et sa séquence est sensible à la méthylation au niveau des cytosines. Sp1 contribue au maintien de la nature hypométhylée de ces îlots CpG (Baumann *et al.*, 2010). Les éléments CCAAT et GC sont facultatifs et fonctionnent plus comme des régions régulatrices qui modulent l'activité transcriptionnelle du promoteur de base.

### **1.5.3. Le cycle de transcription de l'ARN polymérase II**

Le cycle transcriptionnel peut être divisé en plusieurs étapes : la formation du complexe de pré-initiation, l'initiation, l'élongation et la terminaison.

#### **1.5.3.1. Le Complexe de Pré-Initiation**

La transcription des gènes de classe II requiert la présence de l'ARN polymérase II au niveau du promoteur. Le processus de recrutement de l'enzyme fait intervenir une multitude de facteurs protéiques, appelés facteurs généraux de transcription (GTF). Ils sont nécessaires à la synthèse de tous les ARNm et interviennent dans la reconnaissance des séquences du promoteur en réponse aux facteurs régulateurs de la transcription (activateurs et répresseurs). En effet, en s'associant à la région promotrice d'un gène de

classe II, ces GTFs échafaudent un complexe de pré-initiation (*PIC : Pre-Initiation Complex*) qui assure la reconnaissance et la liaison de ce promoteur par l'ARN Pol II. L'ARN Pol II et les FGT TFIID, TFIIA, TFIIB, TFIIIE, TFIIF et TFIIH sont suffisants pour la transcription basale dans un système reconstitué *in vitro*, cependant des facteurs supplémentaires sont requis pour permettre l'expression des gènes en réponse à des activateurs. Le PIC fonctionnel *in vivo* est un grand assemblage coopératif de plus de 40 protéines différentes avec une masse totale de plus de 2 MDa. À l'exception de TFIIB, qui est un monomère, tous les autres FGT ainsi que l'ARNPII sont des complexes stables d'environ 2 à 15 protéines différentes (Baumann *et al.*, 2010).

L'ARNPII, elle-même, représente un complexe de masse supérieure à 500 kDa, composé de 12 sous-unités différentes (voir la section 1.3.). Une caractéristique particulière de sa plus grande sous-unité, Rpb1, est la présence d'un domaine carboxy-terminal (CTD) qui joue un rôle capital dans l'expression des gènes en servant de plateforme d'assemblage pour différents facteurs intervenant au cours de la transcription. Ce domaine est constitué de multiples répétitions d'un heptapeptide de séquence consensus YSPTSPS, dont le nombre varie selon les espèces et augmente avec la complexité du génome (52 chez l'humain). Localisé à l'extérieur du complexe de l'ARN Pol II, ce polypeptide contient cinq sites potentiels de phosphorylation faisant l'objet de régulation de la transcription. Selon l'état de phosphorylation du CTD, *in vivo*, l'ARN Pol II existerait principalement sous deux formes, une forme hypo-phosphorylée nommée IIA, présente lors de la formation du complexe de pré-initiation, et une forme hyper-phosphorylée appelée IIO, se trouvant dans le complexe d'élongation. Des CTD kinases et CTD phosphatases régulent son niveau de phosphorylation (Thomas et Chiang, 2006). Malgré sa complexité, l'ARN Pol II eucaryote dépend des facteurs accessoires requis pour la reconnaissance du promoteur et l'initiation de la transcription. Voir le **Tableau 1-3** pour un résumé de leurs composition et fonctions.

L'assemblage du complexe de pré-initiation au niveau du promoteur peut se produire en plusieurs étapes successives. TFIID serait le premier élément à se fixer sur la matrice d'ADN au niveau de la boîte TATA par sa sous-unité TBP ou à d'autres éléments comme l'Inr, le DPE ou le DCE via les facteurs associés TAF. TFIIA et TFIIB

viendraient ensuite s'ajouter pour permettre de stabiliser la liaison de TFIID à l'ADN. Le complexe promoteur-TFIID-TFIIA-TFIIB permet le recrutement ultérieur de TFIIF pré-associé à l'ARNPII. L'incorporation supplémentaire de TFIIIE et TFIIH est nécessaire pour finaliser le PIC, conduisant à une initiation de transcription efficace. Cette manière ordonnée de recrutement découle des expérimentations *in vitro* avec des facteurs purifiés et est connue comme le modèle d'assemblage séquentiel (Buratowski *et al.*, 1989; Thomas et Chiang, 2006; Baumann *et al.*, 2010).

Il existe un second modèle d'initiation de la transcription appelé holoenzyme (Thompson et Young, 1995; Greenblatt, 1997). Ce modèle d'assemblage a été proposé lorsque différents complexes pré-assemblés de la forme hypo-phosphorylée de l'ARNPII non liée au promoteur ont été purifiés dans plusieurs laboratoires. La composition de l'holoenzyme varie en fonction des conditions de purification, mais on y trouve généralement TFIIF, TFIIIE, TFIIH, des composants du complexe Médiateur et d'autres protéines impliquées dans le remodelage de la chromatine, la réparation de l'ADN et la maturation des ARNm (Baumann *et al.*, 2010) (**Tableau 1-3**). Il est fort probable que les deux voies d'assemblage coexistent *in vivo* et qu'en fonction des molécules de signalisation impliquées et du contexte du promoteur, une voie ou l'autre soit employée pour répondre aux signaux environnementaux.

L'ensemble minimal de facteurs basaux ne répond pas aux activateurs et n'est pas suffisant pour la transcription dans un contexte chromatinien, une conclusion qui a mené à la découverte du complexe Médiateur (Kim, *et al.*, 1994). Le Médiateur, initialement isolé comme un co-activateur, est un adaptateur qui relie les régulateurs transcriptionnels (activateurs/répresseurs) et la machinerie basale pour influencer (en induisant ou en inhibant) la formation d'un PIC au promoteur associé, et, par conséquent, le niveau de transcription. Il sert donc à intégrer les différents signaux correspondant à la fixation de ces régulateurs à leurs séquences régulatrices distales.

**Tableau 1-3. Complexes protéiques impliqués dans l'assemblage du PIC de l'ARN Pol II.** (Adapté de Sikorski and Buratowski, 2009)

Complexe protéique	sous-unités	Fonctions
RNA Pol II	12	Catalyse la transcription des ARNm et d'un sous-ensemble d'ARN non-codant comme snoARN et miARN.
TFIIA	$\alpha$ $\beta$ $\gamma$	Agit comme co-activateur et interagit avec les composants de la machine d'initiation basale ; empêche la répression par des cofacteurs négatifs.
TFIIB		Stabilise la liaison TFIID-promoteur; aide au recrutement de TFIIF/ARN Pol II sur le promoteur ; dirige la sélection du site d'initiation.
TFIID	TBP & 14 TAF	Initie l'assemblage du PIC par la liaison de TBP à la boîte TATA ou la liaison des TAF à d'autres éléments; activité de co-activateur via l'interaction directe entre les TAFs et les activateurs spécifiques.
TFIIE	$\alpha$ $\beta$	Recrute TFIIH aux promoteurs ; stimule les activités hélicase et kinase de TFIIH; lie l'ADNsb et est essentiel à l'ouverture du promoteur.
TFIIF	RAP30 RAP74	Recrute l'ARN Pol II et améliore son affinité pour le complexe TBP-TFIIB-promoteur; induit l'enroulement du promoteur autour du PIC; rôle dans la sélection du site d'initiation et dégagement du promoteur.
TFIIH	XPB, XPD, p62, p52, p44, cycline H, p34, MAT1, Cdk7 et p8.	Activités ATPase/hélicase/kinase; ouverture et dégagement du promoteur ; transcription couplée à la réparation de l'ADN; phosphorylation du CTD; facilite la transition entre l'initiation et l'élongation.
Médiateur	Au moins 24 en fonction du type cellulaire	Interface entre les activateurs et les GTFs; stimule les transcriptions par des activateurs et de base; requis pour la transcription de la plupart des promoteurs ARN Pol II-dépendants.
SAGA	Au moins 20	Interagit avec TBP, les activateurs et l'histone H3; activité histone acétyltransférase et dé-ubiquitinase.
Trf2		Paralogue de TBP; ne peut pas lier la boîte TATA; interaction avec TFIIA et TFIIB; nécessaire à l'expression spécifique de gènes durant l'embryogenèse.
Trf3		Paralogue de TBP; peut lier la boîte TATA; interaction avec TFIIA et TFIIB; important pour la différenciation des cellules musculaires.
TFIIS		Stimule l'activité de clivage de l'ARN Pol II et permet la sortie des sites d'arrêt de la transcription; stimule l'assemblage du PIC à certains promoteurs.
NC2	$\alpha$ $\beta$	Lie le complexe TBP-ADN ou le DPE et bloque l'assemblage du PIC; effets positifs ou négatifs sur la transcription.
BTAF1		Induit la dissociation du complexe TBP-ADN de manière ATP-dépendante; effets positifs ou négatifs sur la transcription.

### 1.5.3.2.L'initiation

Une fois le PIC assemblé au niveau du promoteur, TFIID permet la formation de la bulle de transcription entre les positions -9 et +2, ce qui rend possible l'appariement des ribonucléotides (rNTP) libres au brin matrice (Forget *et al.*, 1997; Robert *et al.*, 1998). Après l'addition du quatrième rNTP, le complexe pré-transcriptionnel subit un changement de conformation qui conduit à la stabilisation du complexe ouvert. La transcription débute par un cycle d'initiations avortées menant à la synthèse de plusieurs courts transcrits de 2 à 10 nt. Dès que l'ARN Pol II atteint la position +11, la région du promoteur entre les positions -9 et +2 se referme et empêche le complexe de ré-initier la transcription. Le CTD de l'ARN Pol II est par la suite phosphorylé et l'enzyme dégage le promoteur, avant d'entrer dans l'étape d'élongation (Coulombe et Burton, 1999; Langelier *et al.*, 2002). Après le détachement de la polymérase, un PIC partiel pré-assemblé peut être maintenu sur le promoteur par un activateur transcriptionnel afin de permettre une ré-initiation plus rapide dans le prochain cycle de transcription.

### 1.5.3.3. L'élongation

Après la libération des facteurs d'initiation, excepté TFIIF, différents facteurs d'élongation sont recrutés au niveau du CTD de l'ARN Pol II. Ces facteurs modulent l'activité d'élongation de l'ARN Pol II par différents mécanismes. Ils peuvent agir de façon positive ou négative directement sur la machinerie transcriptionnelle ou indirectement au niveau de la structure chromatinienne. Divers signaux ou facteurs peuvent bloquer l'activité d'élongation de l'ARN Pol II, entraînant soit une pause transitoire, soit un arrêt du complexe avec son décrochage de la matrice d'ADN et terminaison précoce de la transcription. Les facteurs modulant l'activité de l'enzyme incluent le système P-TEFb/NELF/DSIF, TFIIF, les Elongins, la famille ELL et TFIIIS (revu par Shilatifard *et al.*, 2003). Ainsi, suite à l'étape d'initiation, les complexes négatifs d'élongation DSIF/Spt4-Spt5 (*DRB Sensitivity-Inducing Factor*) et NELF (*Negative Elongation Factor*) induisent une pause de la transcription, permettant le recrutement des facteurs de maturation des ARNm par le CTD de l'ARN Pol II (Mandal *et al.*, 2004). Le facteur d'élongation P-TEFb/CDK9-cycline T1/T2 (*Positive-*

*Transcription Elongation Factor b*), permet la sortie de l'état de pause en phosphorylant les facteurs d'élongation négatifs DSIF et NELF ainsi que le CTD de l'ARN Pol II et, de ce fait, l'entrée en phase d'élongation productive (Cheng et Price, 2007). TFIIF, les Élongines et les protéines de la famille ELL (pour *11-19 Lysine-rich Leukemia*) stimulent l'activité d'élongation de l'ARN Pol II par la suppression des pauses transitoires. Si l'état de pause est un processus réversible, les arrêts transcriptionnels, eux, sont irréversibles à moins de l'intervention de TFIIIS qui est nécessaire au clivage du transcrit naissant permettant ainsi à l'ARN Pol II de poursuivre la transcription à partir d'une nouvelle extrémité 3' de l'ARNm (Kettenberger, 2003).

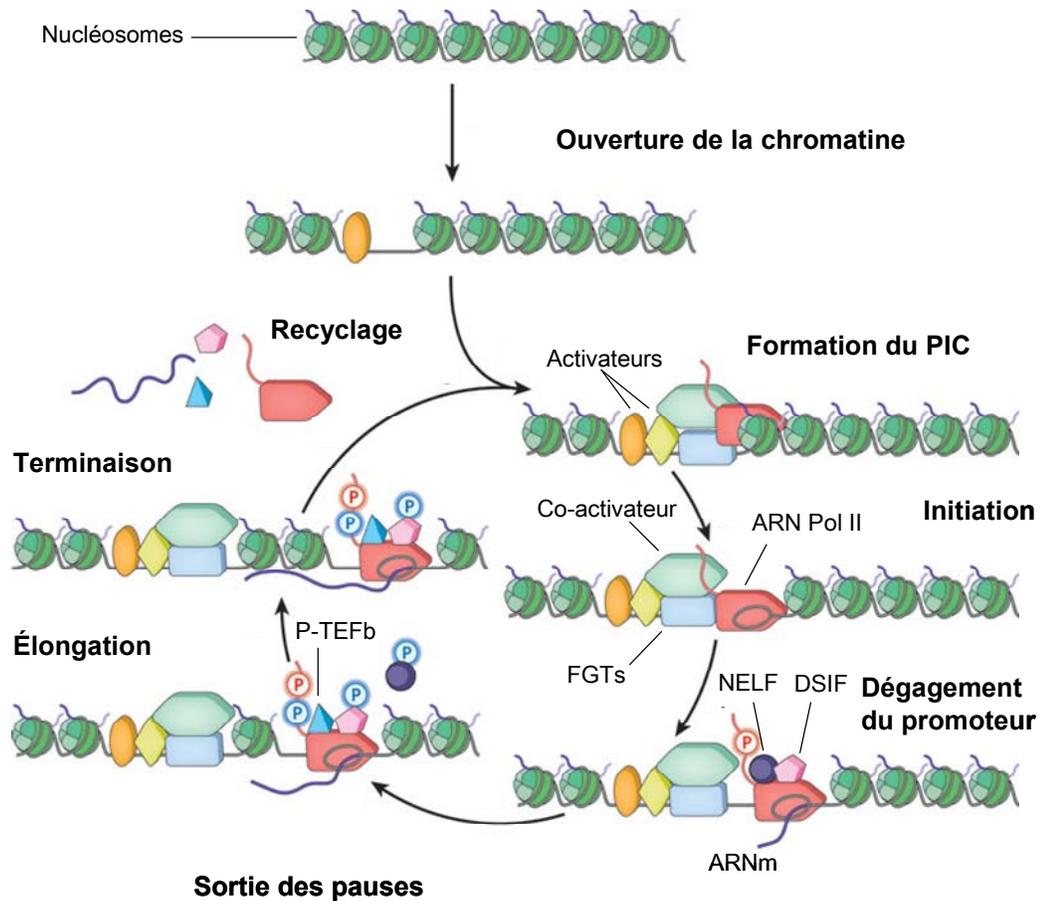
Dans le contexte chromatinien, la poursuite de la phase d'élongation est favorisée par des facteurs remodelant la chromatine de manière ATP- dépendante (les familles ISWI, SWI/SNF et CHD (section 1.2.2.)), des facteurs déstabilisant et rétablissant la structure du cœur du nucléosome (FACT/Spt16-SSRP1 (*FAcilitates Chromatin Transcription*) et SPT6) et des facteurs agissant au niveau des modifications post-traductionnelles des histones du cœur du nucléosome (PAF et Elongator) (revu par Shilatifard *et al.*, 2003).

#### **1.5.3.4. La terminaison**

La terminaison est une étape essentielle du cycle de transcription qui empêche la transcription des séquences en aval d'un gène, phénomène pouvant mener à un effet dévastateur sur les processus cellulaires. La terminaison implique l'arrêt de la synthèse d'ARN, le relâchement du transcrit et le recyclage de la polymérase. Elle peut avoir lieu 0,5-2 kb en aval de l'extrémité 3' de l'ARNm et nécessite la présence d'un signal de terminaison. Ce signal est constitué par la séquence 5'-AAUAAA-3', appelée signal de polyadénylation ou site d'addition du poly(A), et est reconnu par un large complexe multi-protéique de clivage/polyadénylation. La transcription s'arrête 10 à 30 nt après ce signal, qui induit un état de pause du complexe d'élongation. Le complexe de clivage et polyadénylation, comprenant les facteurs CPSF (*Cleavage and Polyadenylation Specificity Factor*), CSF (*Cleavage Stimulation Factor*), CFI (*Cleavage Factor I*), CFII

(*Cleavage Factor II*) et la poly(A) polymérase (PAP), reconnaît et lie le signal poly(A) ainsi qu'une séquence riche en U/GU moins bien conservée.

Le transcrit est d'abord clivé par l'endonucléase CPSF-73 via son domaine metallo- $\beta$ -lactamase, générant une nouvelle extrémité 3'-OH. C'est à partir de cette extrémité que la PAP procède à la synthèse de la queue de poly(A). PABP (*poly A binding protein*) lie la queue poly(A) et protège ainsi le pré-ARNm de la dégradation par les exonucléases. Le clivage du transcrit couplé à sa polyadénylation est suivi par la libération de l'ARNm et le détachement de la polymérase (Richard et Manley, 2009). Par la suite, des phosphatases vont déphosphoryler le CTD de l'ARN Pol II, permettant la régénération de l'enzyme native et son recyclage. La **Figure 1-5** illustre un schéma récapitulatif des étapes de transcription par l'ARN Pol II.



### Figure 1-5. Le cycle de transcription de l'ARN polymérase II.

Suivant l'ouverture de la chromatine par des activateurs et des co-activateurs, l'ARN Pol II et les facteurs généraux de transcription s'assemblent d'abord sur le promoteur pour former le complexe de pré-initiation. Ensuite, la double hélice d'ADN est déroulée au niveau du site d'initiation de la transcription, permettant la formation du complexe ouvert. L'ARN Pol II dégage le promoteur et initie la synthèse du transcrit. DSIF et NELF induisent une pause de la transcription. Après la phosphorylation du CTD et le recrutement de P-TEFb, l'ARN Pol II entre en mode d'élongation productive. La transcription se termine par le relâchement de L'ARN Pol II et du transcrit de la matrice d'ADN. L'enzyme libre peut entreprendre un nouveau cycle transcriptionnel. (Adapté de Fuda *et al.*, 2009).

## 1.6. La protéomique d'interaction à l'étude de la machinerie transcriptionnelle

Comme on a pu le constater, une multitude de facteurs sont impliqués dans le contrôle de l'activité des ARN polymérases durant la transcription. Malgré cela, de nombreux facteurs régulateurs restent encore à découvrir. Dans notre laboratoire nous sommes intéressés surtout aux éléments de contrôle qui déterminent le sort de ces enzymes avant ou après la transcription. Et plus précisément : comment se déroule le processus de biogenèse des trois ARN polymérases, où et de quel façon a lieu l'assemblage des sous-unités, quels sont les facteurs impliqués et comment l'enzyme est-elle importée au noyau? Pour trouver les réponses à ces questions, nous nous sommes tournés vers l'analyse protéomique.

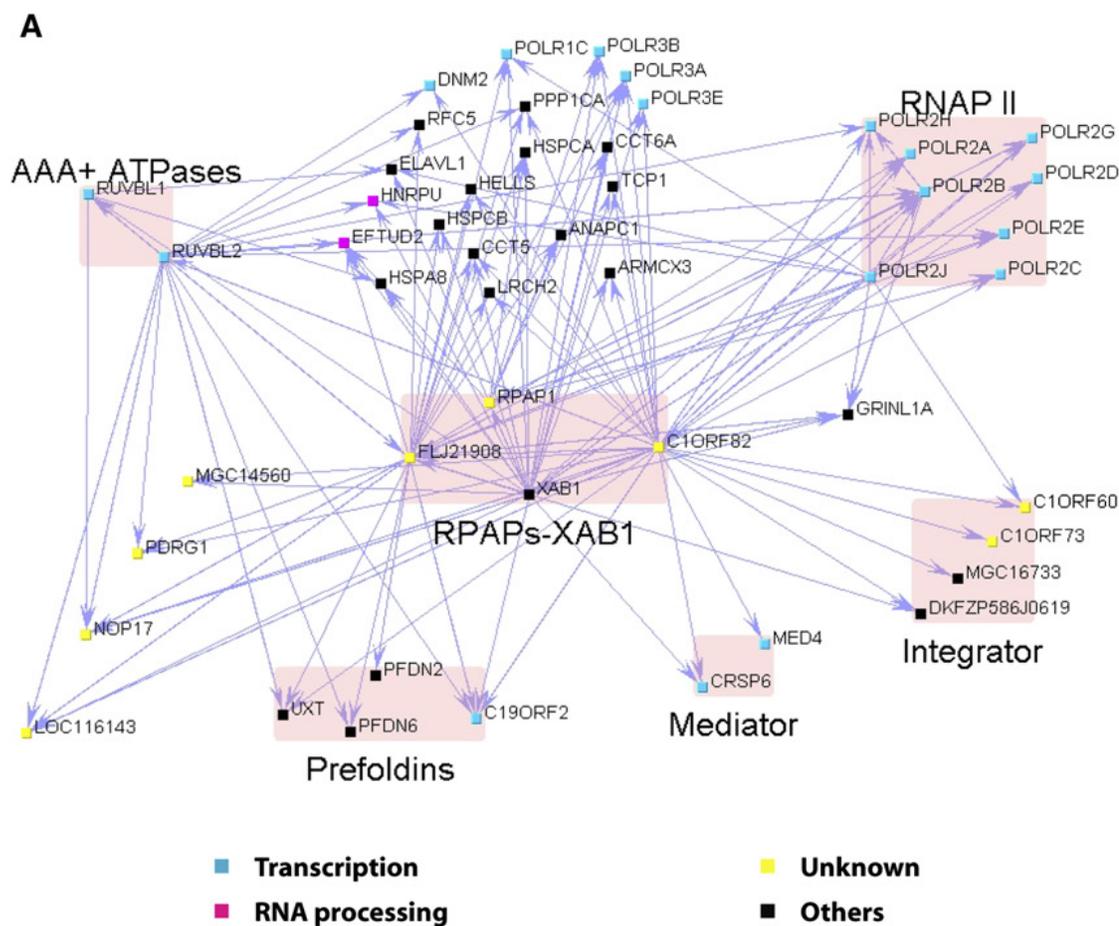
Développés surtout durant la dernière décennie, les approches de la protéomique fonctionnelle se sont avérées très utiles dans l'identification systématique des partenaires protéiques au sein des complexes. Plus précisément, la protéomique d'interaction, qui vise à mettre en évidence les interactions physiques entre les protéines cellulaires, et les éventuelles variations de la composition de ces complexes dans différentes situations biologiques, repose sur plusieurs approches méthodologiques : le système double hybride, le FRET (*Fluorescence Resonance Energy Transfer*), l'analyse de complexes après purification (par des méthodes biochimiques conventionnelles ou par la méthode TAP (Tandem Affinity Purification) ou l'utilisation de sondes chimiques multifonctionnelles (Lescuyer, *et al.*, 2004; Porquet, 2005). Plus récemment, le couplage entre la purification par chromatographie d'affinité et l'identification de polypeptides par spectrométrie de masse (AP-MS) a considérablement fait progresser notre compréhension de la composition des complexes protéiques (Gingras *et al.*, 2007).

### 1.6.1. La méthode AP-MS appliquée à la machinerie transcriptionnelle

La purification par double chromatographie d'affinité à l'aide de protéines étiquetées (TAP-tag) est une technique qui permet de caractériser les réseaux d'interaction protéiques à l'échelle de cellules entières. Initialement développé pour isoler des complexes protéiques chez la levure (Rigaut *et al.*, 1999; Puig *et al.*, 2001), ce

système est très flexible et a pu être adapté à divers organismes. Par exemple, dans notre laboratoire le système de purification TAP a été mis au point afin de purifier des complexes protéiques dans les cellules humaines (Coulombe *et al.*, 2004; Cloutier *et al.*, 2009). Le projet a commencé avec la purification des composants solubles de la machinerie transcriptionnelle de l'ARN Pol II, étant donné qu'il y avait plus de chances de trouver les complexes protéiques impliqués dans l'assemblage, le repliement et le transport nucléaire de l'ARN Pol II dans cette fraction, plutôt que dans la fraction insoluble, contenant de la chromatine et des molécules d'ARN polymérase engagées dans le processus transcriptionnel. Ainsi, 32 polypeptides cibles (appâts) purifiés selon la méthode TAP (voir la section Matériel et Méthodes) ont permis de générer un réseau de 805 interactions protéiques validées à l'aide d'outils bioinformatiques spécialisés. Cette base de données a permis l'identification de nouveaux facteurs qui régulent l'activité de P-TEFb, soit MEPCE (*Methylphosphate Capping Enzyme*) et LARP7 (*Lar-related protein 7*) (Jeronimo *et al.*, 2007; Krueger *et al.*, 2008), ainsi que d'un ensemble de protéines associées à l'ARN Pol II nommées RPAPs (*RNA polymerase II-associated proteins*): RPAP1, RPAP2/C1ORF82, RPAP3/FLJ21908 et RPAP4/XAB1/GPN1 (Jeronimo *et al.*, 2004; Jeronimo *et al.*, 2007) (**Figure 2-1**).

Nous avons montré que ces protéines RPAP formaient une interface entre l'ARN Pol II et des protéines chaperonnes et d'échafaudage et actuellement nous poursuivons la caractérisation de ces facteurs et des mécanismes par lesquels ils accomplissent leurs fonctions. Dernièrement, à partir d'un réseau d'interactions plus élaboré, nous avons montré qu'un de ces RPAPs, RPAP3, faisait partie d'un complexe multi-protéique nommé RPAP3/R2TP/*prefoldin-like*, dont plusieurs sous-unités étaient connues comme étant des préfoldines et des protéines de type préfoldines (Gstaiger *et al.* 2003; Cloutier *et al.* 2009).



**Figure 2-1. Réseau mettant en évidence les interactions entre RPAPs et l'ARN Pol II.**

Les protéines RPAP sont physiquement et fonctionnellement associées à l'ARN Pol II, formant une interface entre l'enzyme et des protéines chaperonnes et d'échafaudage (Jeronimo *et al.*, 2007).

## 1.7. Les chaperonnes générales

Les chaperonnes moléculaires forment une famille de protéines ayant pour rôle de soutenir le repliement et l'assemblage de polypeptides dans des structures oligomériques, et de prévenir la formation de structures protéiques incorrectes (Ellis *et al.*, 1989). Cette famille comporte plusieurs classes de protéines indépendantes ayant des fonctions différentes, comme, les Hsp40 (la famille DnaJ), les Hsp60 (les chaperonines GroEL-GroES et TRiC/CCT), les Hsp70 (la famille DnaK), les Hsp90 (liaison aux récepteurs des hormones stéroïdes), les nucléoplasmines (assemblage du nucléosome), Bip et calnexine (repliement des protéines dans la lumière du réticulum endoplasmique) et la superfamille de AAA+ ATPases. Ces chaperonnes sont dites générales car elles sont en quantité très abondante dans les cellules et partagent la propriété de reconnaître des séquences d'acides aminés plutôt qu'une protéine précise.

Le repliement des protéines se produit principalement dans trois compartiments intracellulaires : cytosol, réticulum endoplasmique et mitochondries. Au niveau de chacun de ces compartiments, il existe différents systèmes de chaperonnes fonctionnellement et structurellement distinctes, certains communes à tous les polypeptides, d'autres limitées à des substrats spécifiques (Melnikov et Rotanova, 2010; Anelli et Sitia, 2008). Dans certains cas, plusieurs systèmes de chaperonnes agissent de concert afin d'aboutir à la formation d'un complexe protéique fonctionnel. On peut mentionner la coopération entre les complexes préfoldines/GimC (chaperonne liant les polypeptides naissants) et TRiC/CCT (chaperonine de classe II) dans le repliement de l'actine et de la tubuline (Vainberg *et al.*, 1998) ou Hsp70 et TRiC/CCT dans l'assemblage du complexe VBC (von Hippel-Lindau protein –elongin BC) (Melville *et al.*, 2003).

### 1.7.1. Rôles des chaperonnes moléculaires dans la biogénèse des ARN polymérase

Les sous-unités des ARN polymérase doivent probablement être assistées par des systèmes de chaperonnes dans leur repliement et assemblage. Le complexe RPAP3/R2TP/*prefoldin-like* que notre laboratoire a purifié, se compose de plusieurs préfoldines classiques et protéines de type préfoldines (« *prefoldin-like* »), notamment PFDN2, PFDN6, UXT, PDRG1 et URI.

PFDN2 et PFDN6 sont connues comme des sous-unités du complexe Prefoldin (PFD)/GimC (gènes impliqués dans la biogénèse des microtubules). Les membres de la famille des préfoldines sont composés de structures de spires enroulées,  $\alpha$  - hélicoïdales, en N- et C-terminal, connectées par un (prefoldine  $\beta$ ) ou deux (prefoldine  $\alpha$ ) motifs en épingles à cheveux  $\beta$ . Prefoldin/GimC est un complexe cytosolique de 90 kDa formé de deux sous-unités  $\alpha$  (VBP1/PFDN3 et PFDN5) et quatre sous-unités  $\beta$  (PFDN1, PFDN2, PFDN4 et PFDN6/HKE2). La structure du complexe PFD ressemble à celle d'une méduse, avec six « tentacules » sous forme d'hélices  $\alpha$ -hélicoïdales émanant du corps cylindrique d'un tonneau  $\beta$  (Siegert *et al.*, 2000). Le complexe PFD reconnaît et lie les protéines nouvellement synthétisées indépendamment de l'ATP et coopère avec la chaperonine TRiC/CCT pour le repliement correct de l'actine et de la tubuline (Vainberg *et al.*, 1998; Seigers *et al.*, 2003).

UXT, PDRG1 et URI sont des protéines de type préfoldine. UXT (*ubiquitously-expressed transcript*) est une préfoldine de type  $\alpha$ , appelée aussi STAP1 (pour SKP2-associated alpha PFD 1). Elle interagit avec le récepteur des androgènes et joue un rôle dans l'activation transcriptionnelle induite par des récepteurs nucléaires (Marcus *et al.*, 2002). UXT est également susceptible d'être impliquée dans le processus de tumorigénèse étant donné qu'elle est surexprimée dans le tissu tumoral (Zhao *et al.*, 2005).

PDRG1 (*p53 and DNA damage regulated 1*) est une protéine dont l'expression est régulée de façon positive par les rayons UV et de façon négative par le suppresseur de tumeur p53. Étant donné que sa surexpression rendrait les cellules plus sensibles aux dommages induits par les rayonnements UV, PDRG1 aurait un rôle potentiel dans l'apoptose (Luo *et al.*, 2003). L'analyse de sa séquence protéique indique une grande

similitude avec la famille des préfoldines  $\beta$ . En effet, dans nos expériences de TAP-tag, PDRG1 est la seule protéine de type préfoldine qui a été co-purifiée avec les autres sous-unités du complexe PFD, à l'exception de PFDN4. Donc, PDGR1 pourrait former un nouveau complexe PFD dans lequel elle substituerait PFDN4 (Cloutier et Coulombe, 2010).

C19ORF2 (RMP/URI) est une protéine connue pour son interaction avec POLR2E (RBP5-mediating protein ou RMP) (Dorjsuren *et al.*, 1998) ainsi que pour son rôle conservé dans la voie de signalisation TOR (Delgermaa *et al.*, 2004; Gstaiger *et al.*, 2003). URI (Unconventional prefoldin RPB5 interactor) a été décrit comme un membre peu conventionnel de la famille des préfoldines  $\alpha$  (à cause de sa grande taille) qui existe dans les cellules humaines sous forme d'un complexe multi-protéique de  $\sim 1$  MD (Gstaiger *et al.* 2003). Il semble y exercer un rôle de protéine d'échafaudage, capable d'assembler un complexe « *prefoldin-like* » qui contient d'autres préfoldines ainsi que des protéines ayant des rôles dans la transcription et l'ubiquitination.

En plus des préfoldines décrites ci-dessus, le complexe RPAP3/R2TP/*prefoldin-like* contient un module R2TP (formé par RPAP3, RUVBL1, RUVBL2 et PIH1D1), WDR92 et possiblement POLR2E. Dans notre réseau d'interactions ce complexe apparaît accompagné d'autres systèmes de chaperonnes comme les complexes Prefoldin, TRiC/CCT et les HSP90.

## 1.8. Objectifs et stratégie

Le but général de mon projet de maîtrise était de poursuivre la cartographie du réseau d'interactions protéiques de la machinerie de transcription humaine.

Le premier objectif était d'appliquer la méthode de purification TAP à l'étude du complexe de l'ARN Pol I humaine afin d'établir le réseau d'interactions des protéine-protéine impliquées dans la transcription des ARNr. Nous avons choisi les sous-unités spécifiques de l'ARN Pol I, POLR1E et TWISTNB, comme protéines-appât.

Le deuxième objectif découle d'un autre aspect déterminant de notre procédure qui s'appuie sur l'étiquetage réciproque des partenaires d'interactions (proies) identifiées dans nos expériences antérieures. Cette étape est importante pour confirmer certaines interactions ce qui contribue à accroître la confiance dans notre réseau. Ainsi, suite à la purification des complexes protéiques de la machinerie de transcription de l'ARN Pol II humaine, POLR2E et PFDN4, MBD2 et XPA, CAND1 et PDCD5 ont été identifiés comme des partenaires d'interactions possibles de RPAP3, RPAP4/XAB1 et CCT5 (sous-unité du complexe TRiC), respectivement. Par conséquent, ces protéines ont été choisies comme appâts dans nos expériences de purification TAP et d'analyse protéomique subséquente. Pour ce faire, j'ai procédé, d'abord, à la création des lignées cellulaires stables EcR-293 exprimant les protéines d'intérêt à un niveau physiologique. Ensuite j'ai purifié par double chromatographie d'affinité les protéines étiquetées ainsi que leurs partenaires d'interaction à partir de la fraction soluble des lysats cellulaires. Les protéines présentes dans les éluats ont été séparées sur gel, extraites et identifiées par spectrométrie de masse. Finalement, à l'aide d'outils bioinformatiques, les résultats ont été analysés et intégrés dans un réseau d'interactions.

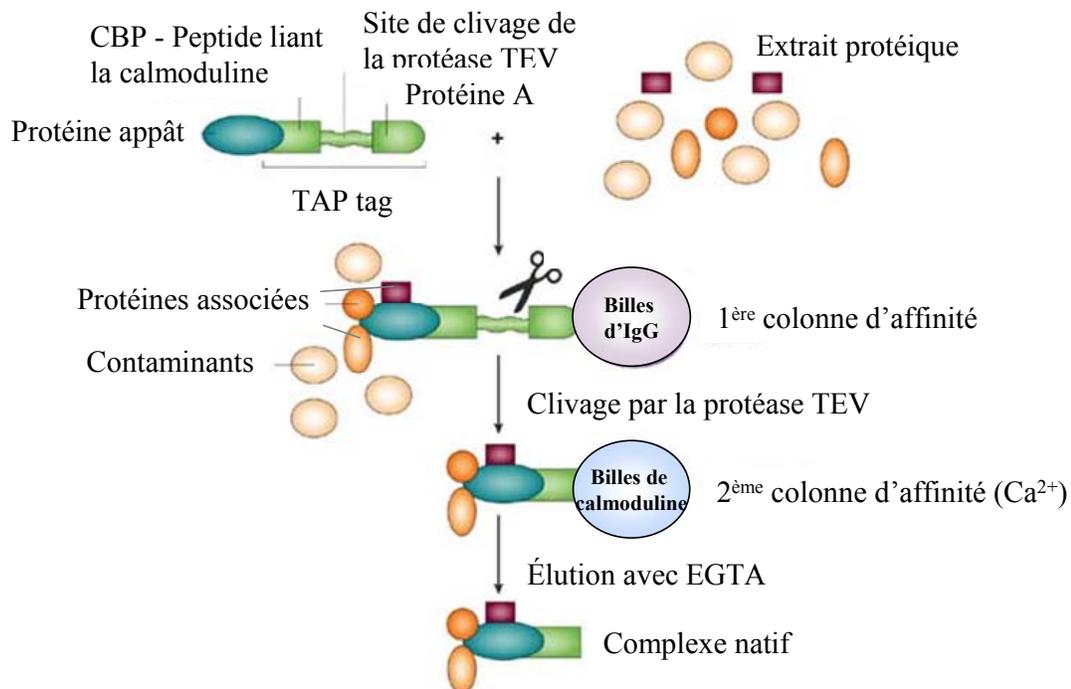
Les résultats obtenus dans cette étude protéomique ont permis d'identifier la composition de l'ARN Pol I assemblée *in vivo* dans la fraction soluble des cellules humaines, ainsi que de mieux définir les composantes des complexes protéiques associés à la machinerie transcriptionnelle de l'ARN Pol II.

## 2. MATÉRIEL ET MÉTHODES

### 2.1. Le système de purification TAP (*tandem affinity purification*)

Cette méthode permet de purifier des complexes protéiques en conditions natives à l'aide de protéines étiquetées (appâts) exprimées à des niveaux physiologiques. La méthode TAP permet une purification rapide des complexes à partir d'un nombre relativement faible de cellules, et ce, sans la connaissance préalable de la composition, l'activité ou la fonction des complexes. L'étiquette TAP est constituée de trois éléments fusionnés, permettant une double purification d'affinité et diminuant ainsi les interactions non spécifiques. Il s'agit d'une fusion en tandem d'un peptide de liaison à la calmoduline (CBP : *calmodulin-binding peptide*), d'un site de clivage pour la protéase TEV (*Tobacco Etch Virus*) et d'un double site de liaison pour les immunoglobulines G (IgG) de la Protéine A de *Staphylococcus aureus* (**Figure 3-1**). Il est possible de réaliser le marquage soit à l'extrémité C-terminale, soit à l'extrémité N-terminale, afin de minimiser son impact sur l'activité de la protéine. En N-terminal, l'agencement des trois domaines est inversé afin de toujours garder la Protéine A à l'extrémité N-terminale de la protéine étiquetée. Une fois exprimée dans la cellule hôte, la protéine de fusion ainsi que l'ensemble des protéines du (des) complexe(s) peuvent être récupérés par une double chromatographie d'affinité. La première en présence de billes d'IgG suivie d'une élution après clivage à la protéase TEV et la deuxième avec des billes de calmoduline, élution en conditions non dénaturantes (Rigaut *et al.*, 1999; Puig *et al.*, 2001).

Grâce à cette méthode de purification, un procédé d'analyse systématique a été mis au point dans notre laboratoire, afin de détecter et de caractériser les interactions protéiques provenant des complexes solubles assemblés *in vivo* dans les cellules humaines (Jeronimo *et al.*, 2004; Jeronimo *et al.*, 2007). Cette stratégie se décompose en plusieurs étapes et comprend : la construction du plasmide contenant le gène d'intérêt fusionné à l'étiquette TAP; la création de lignées cellulaires stables contenant les protéines d'intérêt; la purification des complexes protéiques par la méthode TAP; l'identification des partenaires d'interaction par spectrométrie de masse; l'analyse bioinformatique et la représentation graphique des interactions.



**Figure 3-1. Schéma résumant la technique de purification TAP-tag.**

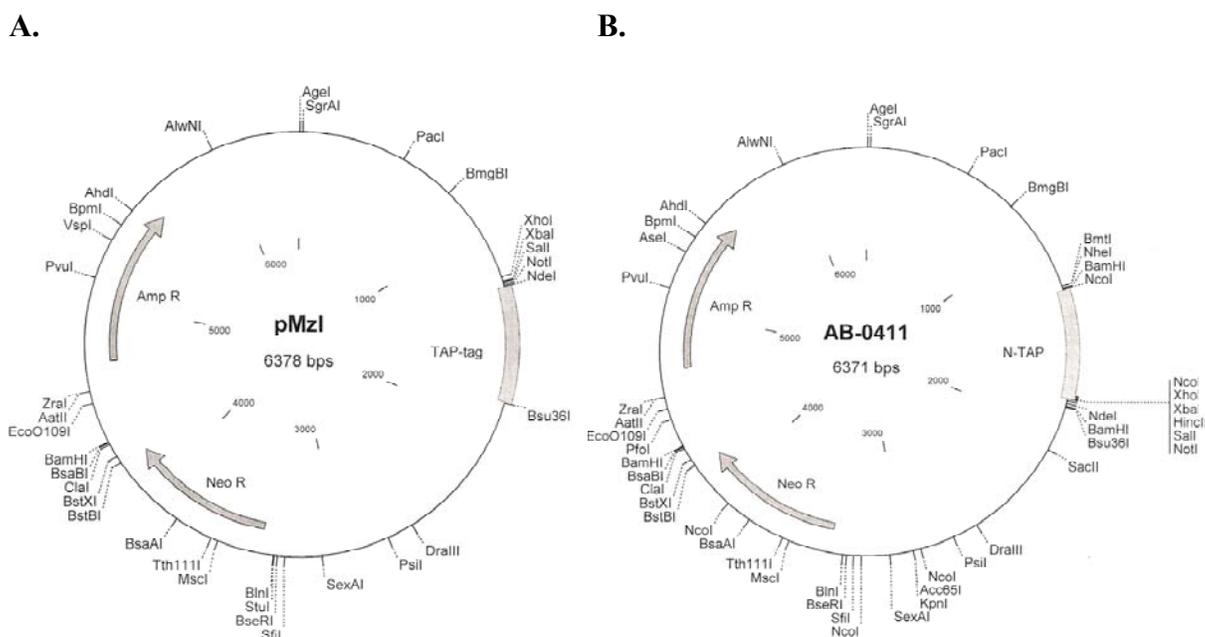
L'étiquette TAP (tandem-affinity-purification) est composée de trois éléments : un peptide de liaison à la calmoduline (CBP : *calmodulin-binding peptide*), un site de clivage pour la protéase TEV (Tobacco Etch Virus) et un double site de liaison pour les immunoglobulines G (IgG) de la Protéine A de *Staphylococcus aureus*. Des lignées cellulaires contenant des protéines d'intérêt étiquetées (TAP-tag) et exprimées à des niveaux physiologiques sont d'abord créées. Ensuite, les protéines de fusion et les composants associés sont récupérés à partir des extraits protéiques solubles en deux étapes successives de purification grâce aux billes d'IgG et de calmoduline. Figure modifié à partir de (Huber, 2003).

### 2.1.1. Construction du plasmide contenant le gène d'intérêt fusionné à l'étiquette TAP

Pour obtenir un taux d'expression proche du niveau physiologique et contrevenir ainsi à des possibles effets néfastes de la surexpression de la protéine étiquetée, un system d'expression inducible a été utilisé. L'ADNc des protéines d'intérêts ont été sous-cloné dans le vecteur pMZI qui permet l'expression des protéines recombinantes dans des cellules de mammifères sous le contrôle du promoteur EcRE inducible à l'ecdysone (section 3.1.2.). Ce plasmide se compose de deux cassettes de résistance,

NeoR (néomycine) et AmpR (ampicilline), un site de clonage multiples et la région codante de l'étiquette TAP en N- ou en C-terminal (**Figure 3-2**).

Dans le cadre de ce projet de recherche les plasmides contenant l'ADNc humain des gènes d'intérêt fusionné à l'étiquette TAP ont été créés en plusieurs étapes.



**Figure 3-2. Carte du vecteur pMZI avec la version C-terminale (A) et N-terminale (B) de l'étiquette TAP.**

#### 2.1.1.1.PCR

Les clones contenant l'ADNc d'intérêt proviennent de la compagnie Open Biosystems. L'amplification par PCR a été réalisée avec des amorces contenant des sites de restriction compatibles avec le vecteur pMZI en 5' et en 3' du gène, de sorte que ce dernier se trouve dans le même cadre de lecture avec l'étiquette TAP. Les sites de restriction insérées sont XhoI et NotI pour XPA, POLR1E et TWISTNB; XbaI et SalI pour PFDN4 et PDCD5; XhoI et NdeI pour MBD2; SalI et NotI pour CAND1 et finalement XhoI et XbaI pour POLR2E.

Les réactions PCR sont réalisées en utilisant de l'ADN polymérase Phusion Hot Start (Finnzymes), selon les spécifications du fabricant. La réaction de PCR contient 1 ng de matrice d'ADN; 1,75 µl de chaque amorce (10 µM); 0,75 µl de dNTPs (10 mM); 0,5 µl de l'ADN Polymérase Phusion Hot Start (1U) et 10 µl de tampon Phusion (5X) (7,5 mM MgCl<sub>2</sub>) pour un volume total de 50 µl. Le programme des cycles de température de la réaction PCR comprend les étapes suivantes : 98°C / 30 sec; 98°C / 10 sec, (T<sub>m</sub>-5°C) / 30 sec, 72°C / (30 sec par kb), les trois dernières étapes sont répétées en boucle 35 fois et finalement 72°C / 10 min.

#### 2.1.1.2. Digestion enzymatique

Les produits PCR ainsi que le vecteur pMZI sont séparés sur gel d'agarose 2 % (1 µg/mL bromure d'éthidium) dans le tampon d'électrophorèse TAE 1X, purifiés avec le kit de purification de Qiagen et digérés par les enzymes de restriction respectifs (New England Biolabs) pendant 3 h (ou durant la nuit) à 37°C, dans le tampon recommandé par le fabricant. Les produits de digestion sont de nouveau séparés sur gel d'agarose 1% et purifiés avec le kit de purification de Qiagen. L'ADN est ensuite quantifié par spectrométrie en utilisant une longueur d'onde de 260 nm.

#### 2.1.1.3. Ligation dans le vecteur pMZI

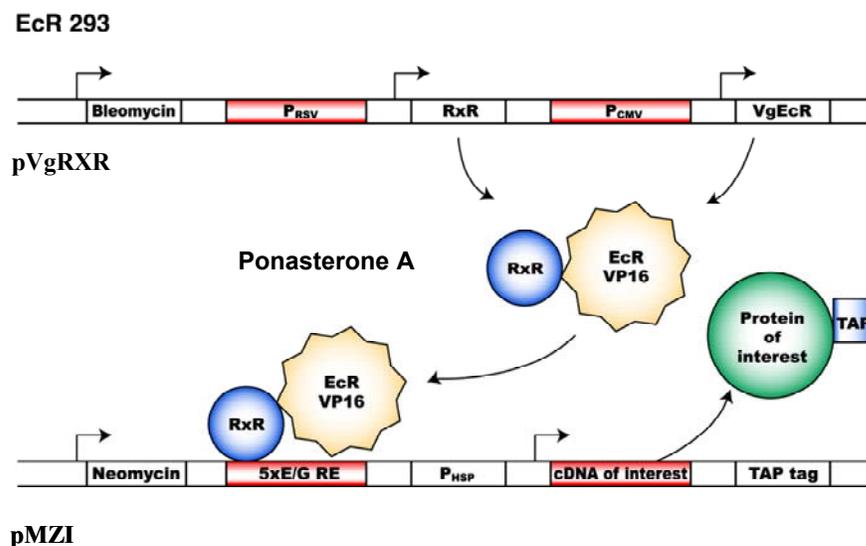
L'introduction de l'insert digéré dans le vecteur pMZI linéarisé a lieu en présence de la T4 ADN ligase (NEB) pendant 1 h à température ambiante. Le volume réactionnel de 20 µL comprend 250 fmol d'ADN de l'insert (gène d'intérêt), 5 fmol d'ADN du vecteur (pMZI contenant l'étiquette TAP), 1 unité de T4 ADN ligase et 2 µL de tampon ligase 5X (66 mM Tris-HCl; 110 mM MgCl<sub>2</sub>; 1 mM dithiothreitol; 1 mM ATP; 6% Polyethylene glycol (PEG 6000); pH 7.6 à 25°C).

#### 2.1.1.4. Transformation des bactéries compétentes

Les bactéries de la souche XL-1 compétentes ont été préparées au laboratoire selon la méthode standard. La transformation bactérienne des produits de ligation est effectuée par choc thermique. 10  $\mu$ L du produit de ligation sont ajoutés à 100  $\mu$ L de bactéries compétentes et le mélange est incubé 30 min sur glace. Un choc thermique de 90 sec à 42°C est appliqué suivi de 2 min d'incubation sur glace. Après l'ajout de 900  $\mu$ L de milieu LB, les bactéries sont incubées 1 h à 37°C avec agitation. Les clones bactériens sont cultivés sur boîtes de Petri (100 mm) milieu LB additionnés d'ampicilline (100  $\mu$ g/mL) dans un incubateur à 37°C durant 18 à 24 h. Plusieurs colonies sont prélevées et cultivées individuellement dans 5 mL LB ampicilline. Après purification de l'ADN plasmidique (kit Qiagen), la présence de l'insert est vérifiée soit par PCR ou par digestion enzymatique. Les clones positifs sont ensuite séquencés et produits en grande quantité. Des stocks glycérol (40 %) des clones bactériens sélectionnés sont réalisés.

#### 2.1.2. Création de lignées cellulaires stables contenant les protéines d'intérêt étiquetées

Les constructions plasmidiques obtenues sont transfectées dans les cellules humaines immortalisées EcR-293 (ATCC). Ces cellules ont été choisies comme système d'expression car elles permettent d'induire de manière spécifique l'expression des protéines d'intérêt. Les cellules EcR-293 contiennent le plasmide pVgR $\times$ R (Invitrogen) codant le récepteur de l'ecdysone (VgEcR dérivé de la Drosophile et modifié pour contenir le domaine de transactivation VP16) et le récepteur X rétinoïde (RXR dérivé des cellules mammifères). Les deux récepteurs exprimés VgEcR et RXR, hétérodimérisent et activent le promoteur contenant la séquence EcRE du pMZI en présence de l'hormone ecdysone ou d'un analogue (Ponastérone A) (**Figure 3-3**). Ce système permet l'expression de manière dose-dépendante de la protéine d'intérêt et, donc, en modulant la concentration de l'inducteur, il est possible d'ajuster l'expression à un niveau physiologique.



**Figure 3-3. Schéma fonctionnel du système de régulation répondant à l'ecdysone.**

#### 2.1.2.1. Transfection et sélection

La transfection stable des cellules EcR-293 a été réalisée par la méthode utilisant du phosphate de calcium. L'ADN plasmidique (8 µg) est dissout dans 500 µL de 250 mM CaCl<sub>2</sub> stérile et ajouté goutte à goutte à 500 µL de 2 X HeBS (0.5 mM Na<sub>2</sub>HPO<sub>4</sub>.7H<sub>2</sub>O, 274 mM NaCl, 54.5 mM HEPES, pH 7.0), tout en effectuant l'agitation par émulsion. Après incubation pendant 10 min à température ambiante, le précipité d'ADN est versé dans une boîte de Petri (100 mm) contenant de cellules EcR-293 cultivées jusqu'à un taux de 50 à 70% de confluence (milieu DMEM complétement avec 10 % FBS (sérum de veau fœtal), 2 mM glutamine, 100 µg/ml pénicilline, 100 µg/ml streptomycine et 30 µg/ml de bleocine, incubateur à 37°C et 5% CO<sub>2</sub>). Après 12 à 14 h d'incubation, des lignées stables sont sélectionnées et maintenues dans le même milieu additionné de l'agent sélectif G418 (Invitrogen) à une concentration de 300 µg/ml. À partir de ce pool, des clones individuels sont isolés sur plaque de 96 puits et cultivés en grandes quantités.

### 2.1.2.2. Analyse d'expression inductible

Lors de la transfection, le plasmide pMZI contenant la protéine de fusion est inséré de manière aléatoire dans le génome. Pour cette raison, les cellules expriment la protéine recombinante différemment selon le site d'intégration du plasmide. Dans certains cas, il arrive que les cellules expriment la protéine recombinante en l'absence d'induction (*leaky expression*), or un contrôle strict de l'expression avant induction est indispensable à l'identification des contaminants après la purification TAP.

L'expression de la protéine de fusion est vérifiée par immunodétection de la protéine recombinante par immunobuvardage de type western avant et après induction à la Ponastérone A (3  $\mu$ M) au niveau des plaques de 12 puits. Après environ 24 h d'induction, les cellules sont lysées dans un tampon de lyse (0,1 % Triton x-100, 0,1 M Tris-HCl pH8.0) et les protéines sont séparées sur un gel 10 % de polyacrylamide 30 : 0,8 en conditions dénaturantes (SDS-PAGE). Après l'électrophorèse (125V / 1h,) les protéines du gel sont transférées sur une membrane de nitrocellulose. Les incubations de 1 h pour chaque anticorps, primaires (1 : 1000) et secondaire (1 : 1000), sont suivies d'une révélation par chimioluminescence (ECL Plus, Amersham) sur film autoradiographique (Hyperfilm ECL, Amersham). Les anticorps primaire et secondaire utilisés sont, respectivement, l'anticorps polyclonal de lapin (IgG) (C-21 : sc-900 Santa Cruz Biotechnology) dirigé contre le domaine CTD de la plus grande sous-unité de l'ARN Pol II et l'anticorps anti-IgG de lapin couplé à la peroxydase de raifort (GE Healthcare). La détection de RPB1 par C-21 sert de référence pour juger de la quantité d'extrait par puits.

### 2.1.2.3. Induction et préparation de l'extrait

Les clones qui expriment correctement la protéine de fusion sont cultivés pour atteindre 40 boîtes de Petris de 150 mm. La synthèse des protéines est induite à une concentration de 3  $\mu$ M de Ponasterone A pour une durée de 42 h dans la moitié du volume de culture cellulaire, l'autre moitié servant de contrôle négatif. Les cellules sont détachées, en tapant modérément sur les boîtes, lavées avec du PBS 1X glacial et

récoltées par centrifugation à 3000 rpm. Les étapes qui suivent sont réalisées parallèlement pour les échantillons induits et non-induits.

L'extraction des protéines cellulaires est effectuée à l'aide d'un homogénéisateur Dounce (Wheaton) dans les tampons A et B. D'abord le tampon A (10 mM Hepes, pH 7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.5 mM DTT, 0.5 mM AEBSF et 1 tablette de cocktail d'inhibiteurs de protéases sans EDTA (Roche Applied Sciences)) est ajouté au culot cellulaire dans une proportion de 4/3 (ml/g de cellules). Après 10 cycles de compression lente, le tampon B (50 mM Hepes, pH 7,9, 1,5 mM MgCl<sub>2</sub>, 0,5 mM DTT, 0,5 mM AEBSF, 1,26 M CH<sub>3</sub>COOK, glycérol 75 %) est ajouté à l'homogénat dans une proportion de 1 : 1 (ml/g de cellules). Les membranes cellulaires sont complètement brisées après 10 cycles de compression supplémentaires. Le lysat est ensuite ultracentrifugé à 4°C à une vitesse de 37 000 rpm pendant 3 h et le surnageant est dialysé durant 18 h à 4°C dans 3 L de tampon de dialyse (10 mM Hepes, pH 7.9, 0.1 mM EDTA, 0.1 mM DTT, 0.1 M K acetate, 10% glycerol). L'extrait protéique est récupéré après une dernière centrifugation effectuée à 4°C pendant 30 min à 14 000 rpm.

### **2.1.3. Purification des complexes protéiques par la méthode TAP**

#### **2.1.3.1. La colonne d'IgG**

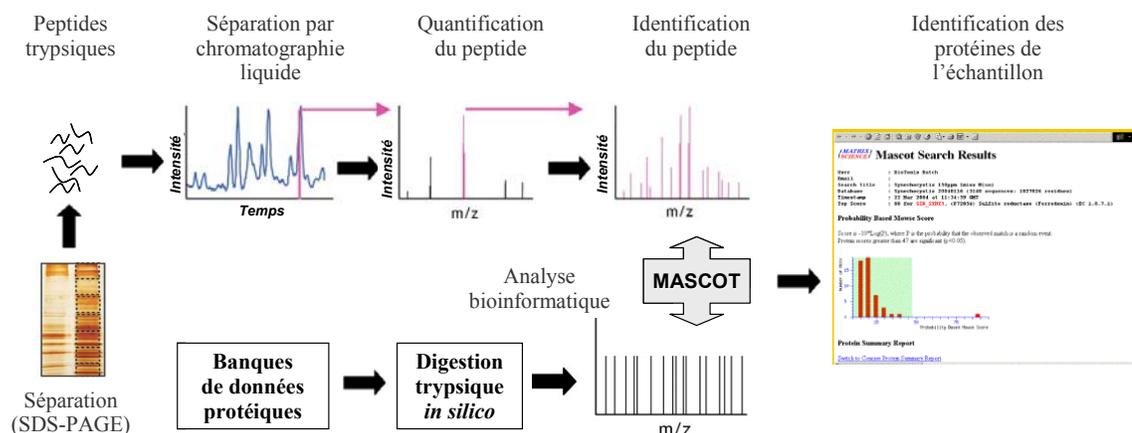
L'extrait protéique est mis en présence des billes d'IgG sépharose 6 Fast Flow (GE Healthcare) (50 µL/g de culot cellulaire) qui ont été équilibrées au préalable avec le tampon IPP (10 mM Tris pH 8; 500 mM NaCl; 0,1% Triton X100; 10 % glycérol) et incubé durant 1 h à 4°C avec agitation. Les billes sont ensuite lavées 2 fois avec le tampon IPP, transférées sur une colonne de chromatographie Bio-spin (Bio-Rad) et équilibrées avec le tampon TEV (10 mM Tris-HCl, pH 8, 100 mM NaCl, 0.1% Triton X-100, 0.5 mM EDTA, 10% glycerol, 1 mM DTT). Les billes sont traitées avec la protéase TEV (30 U/g de culot cellulaire) dans le tampon TEV (200 µL/g de culot cellulaire) à 4°C durant 16 h avec agitation. Le clivage libère les protéines appâts liées aux billes d'IgG et expose le domaine de liaison à la calmoduline afin de permettre la deuxième étape de purification.

### 2.1.3.2. La colonne de calmoduline

L'éluât de la première colonne d'affinité est récupéré et mis en présence des billes de calmoduline sépharose 4B (GE Healthcare) (50 µL/g de culot cellulaire), préalablement équilibrées avec le tampon CBB (10 mM Tris-HCl, pH 8; 100 mM NaCl; 1 mM Imidazole; 1 mM MgAc; 2 mM CaCl<sub>2</sub>; 10 mM β-mercaptoéthanol; 0,1 % Triton X100; 10 % glycérol). Les billes sont incubées 2 h à 4°C en présence de CaCl<sub>2</sub> (2 mM). Elles sont ensuite lavées deux fois avec le tampon CBB et transférées sur une deuxième colonne Bio-spin. L'éluat est réalisé en deux fois (100 et 150 µL/g de culot cellulaire) avec le tampon CEB (10 mM Tris pH 8; 500 mM NaCl; 1 mM Imidazole; 1 mM MgAc; 2 mM CaCl<sub>2</sub>; 10 mM β-mercaptoéthanol; 0,1 % Triton X100; 10 % glycérol). L'éluât a été congelé dans de l'azote liquide et son volume a été réduit par centrifugation sous vacuum à environ 30 µL.

### 2.1.4. Identification des partenaires d'interaction par spectrométrie de masse

Les protéines sont chargées sur un gel à gradient NuPAGE 4-12% Bis-Tris 1.0 mm (Invitrogen) et séparées par électrophorèse à 200 V pendant 40 min dans un système de mini-cellules Novex (Invitrogen). Le gel est coloré au nitrate d'argent puis découpé en 18 à 20 bandes pour chacun des deux échantillons, induit et non-induit. Les protéines sont réduites et alkylées puis digérées à la trypsine dans le gel (Havlis *et al.*, 2003; Cloutier *et al.*, 2009). Les peptides issues de cette digestion sont analysés par LC-MS/MS afin d'identifier les protéines correspondantes. L'analyse est réalisée avec un système nanoLC-2D (Eksigent Technologies) couplé au spectromètre de masse LTQ Orbitrap de Thermo Fisher Scientific. Les peptides sont séparés sur une colonne de phase inversée Jupiter C18 (100 mm x 75 µm, 5 µm, 300 Å) (Phenomenex) avec une phase mobile A (0.2 % acide formique) et B (100 % acétonitrile / 0.2 % acide formique) (voir le schéma à la **Figure 3-4**) (Cloutier *et al.*, 2009). En sortie de colonne, une analyse des peptides dans un spectromètre de masse en mode ESI positif (ElectroSpray Ionization) permet de caractériser la masse de ces peptides puis leur composition en acides aminés par fragmentation en mode MS/MS.



**Figure 3-4. Illustration schématique de la stratégie d'identification des protéines par spectrométrie de masse.** Les protéines de l'échantillon purifié sont séparées sur gel SDS-PAGE. Les bandes sont découpées et les protéines sont digérées par la trypsine *in situ* dans le gel. Les peptides obtenus sont séparés par chromatographie liquide, ionisés et analysés en mode MS, afin d'en déduire leur masse. Après fragmentation et analyse en mode MS/MS, les spectres de masse obtenus sont comparés à l'aide du logiciel MASCOT aux spectres de masse virtuels des protéines répertoriées dans les banques de données et les protéines sont finalement identifiées par homologie de séquences.  $m/z$  (ratio masse/charge). Figure modifiée à partir de (Gingras *et al.*, 2005).

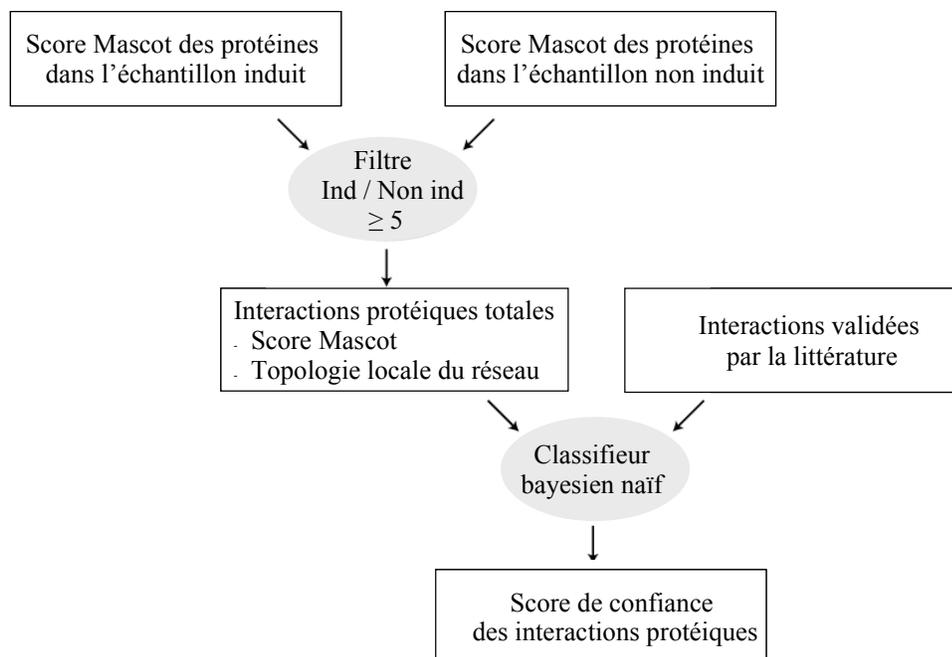
L'ensemble des données générées est collecté et soumis aux banques de données grâce au moteur de recherche MASCOT (version 2.1) (.Matrix Science). Ce moteur de recherche est un outil d'analyse de données très puissant dont l'algorithme est basé sur une approche probabilistique (Perkins *et al.*, 1999). Cela signifie que les masses des fragments mesurées sont comparées aux masses des fragments théoriques générés *in silico* à partir des peptides présents dans les banques de données et la probabilité d'identité est mesurée par un score Mascot. MASCOT permet d'interroger les principales banques de données protéiques comme SwissProt<sup>®</sup>, Prosite<sup>®</sup> et NCBI (*US National Center for Biotechnology Information*). Il contient aussi les séquences traduites des collections de séquences d'ADN maintenues dans la GeneBank, ainsi que les résultats de clivage par les protéases de séquences peptidiques connues.

### 2.1.5. Interprétation bioinformatique et représentation graphique des interactions protéine-protéine

Les outils de spectrométrie de masse nous permettent de générer une liste de candidats-partenaires d'interaction, pouvant contenir une certaine fraction de faux positifs. Dans notre cas, ces interactions non-spécifiques sont estimées à 13% (Jeronimo *et al.*, 2007) et sont dues surtout à des interactions transitoires ou à des contaminants.

Afin de minimiser le taux des faux positifs et de faux négatifs nous avons utilisé une approche bioinformatique basée sur un algorithme développé dans notre laboratoire (Proteus). Ainsi, les résultats obtenus sont triés et validés manuellement pour donner en final une liste des protéines présentes dans l'échantillon initial. Notre système de validation comporte plusieurs filtres. Le premier filtre, basé sur le score Mascot, nous permet d'éliminer les protéines qui se fixent de manière non-spécifique aux colonnes d'affinité durant les étapes de purification. Les interactions retenues sont celles pour lesquels le score Mascot dans l'éluât induit est au moins cinq fois supérieur comparativement à l'éluât équivalent non induit (**Figure 3-5**).

Le deuxième filtre est basé sur la notion de score de confiance pour une interaction donnée ou « Interaction Reliability (IR) Score ». Ce score permet d'estimer la pertinence des interactions en fonction d'une évaluation statistique considérant les facteurs suivants : score Mascot, score Mascot pour un peptide, nombre de partenaires d'interaction communs, bidirectionnalité de l'interaction et nombre de protéines-appâts par partenaire-proie. Le score IR est calculé à l'aide d'un classifieur bayésien naïf à partir d'une série de 135 interactions positives, validées par la littérature, et une série de 53 interactions négatives, considérés comme étant des faux positifs (Jeronimo *et al.*, 2007). En se basant sur le calcul de la spécificité et de la sensibilité de la méthode (taux d'interactions prédites qui sont correctes), le seuil de fiabilité du score IR est estimé à 0,6729. En appliquant ce seuil de confiance, il est estimé que environ 13% des interactions sont des faux positifs et un ratio équivalent sont des faux négatifs (83% sensibilité et 83% spécificité) (Jeronimo *et al.*, 2007). Les interactions pour lesquelles le score IR dépasse cette valeur sont retenues et, à partir de celles-ci, des réseaux d'interactions protéine-protéine sont créés.



**Figure 3-5. Schéma de validation des interactions protéiques**

Le premier filtre permet d'éliminer les interactions non spécifiques en comparant les scores Mascot des échantillons (expression induites / non induite  $\geq 5$ ). Le deuxième filtre permet d'estimer la pertinence des interactions en fonction d'un score de confiance calculé selon les critères indiqués (voir le texte pour plus de détails). Modifié à partir de (Jeronimo *et al.*, 2007).

Les réseaux d'interactions sont représentés par un graphique à l'aide du logiciel VisANT, où les nœuds représentent les protéines et les arcs les interactions entre elles. On utilise un arc non-orienté pour décrire une interaction bidirectionnelle et une flèche pour une interaction non symétrique. Dans nos représentations graphiques, les nœuds sont rassemblés selon leurs annotations proposées par le « *Gene Ontology Consortium* » (Ashburner *et al.*, 2000) et/ou le degré de connectivité entre les nœuds.

## RÉSULTATS

La purification par la méthode TAP des complexes protéiques associés aux protéines étiquetées a été réalisée comme décrite à la section 3.1. Lors de chaque analyse, les résultats obtenus ont été compilés sous forme d'un tableau indiquant le score IR pour chaque protéine identifiée. Une analyse bioinformatique de la structure de chaque protéine étudiée ainsi que leurs séquences sont présentées sous forme de figures. Les représentations graphique des motifs et domaines ont été recueillis à partir du navigateur Ensembl (<http://uswest.ensembl.org/index.html>). L'ensemble des résultats est récapitulé à la section 4.3 où toutes les interactions ont été rassemblées sous forme d'un réseau.

### 2.2. Étude du complexe de l'ARN polymérase I

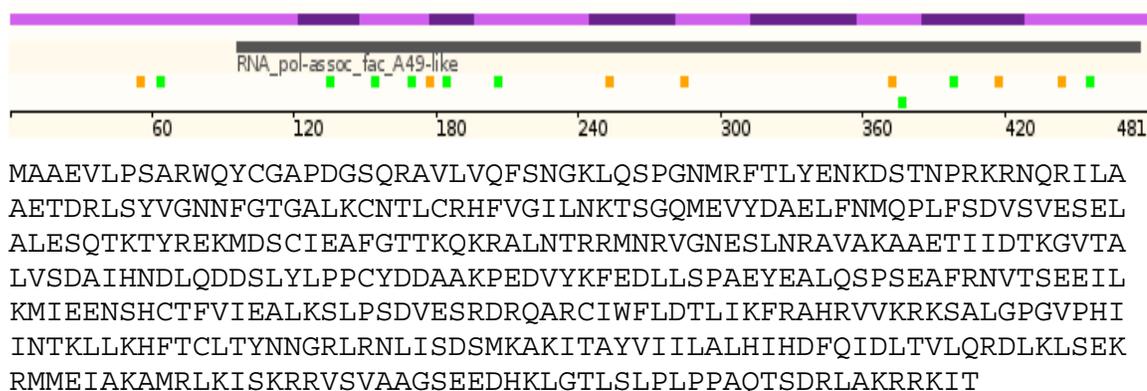
L'ARN Pol I humaine est constituée de treize sous-unités, parmi lesquelles cinq sont spécifiques (A194, A127, A12.2/ZNRD1, hRPA43/TWISTNB hPAF53/POLR1E et hPAF49/CD3EAP), deux communes aux ARN Pol I et III (AC40 et AC19) et cinq communes aux ARN Pol I, II et III (Rpb5, Rpb6, Rpb8, Rpb10 et Rpb12) (Panov *et al.*, 2006; Cramer *et al.*, 2008). La composition et la structure de l'enzyme chez l'humain a été en partie déduite grâce à des études basées sur l'homologie de séquences chez la levure. Par contre, les résultats complets de purification de la machinerie de transcription de l'ARN Pol I assemblée dans les cellules humaines ne sont pas encore disponibles dans la littérature scientifique. Afin de purifier ce complexe chez l'humain nous avons sélectionné et étiqueté les sous-unités spécifiques POLR1E et TWISTNB, à partir desquelles nous avons créé les lignées cellulaires correspondantes.

#### 2.2.1. Purification TAP de POLR1E/hPAF53

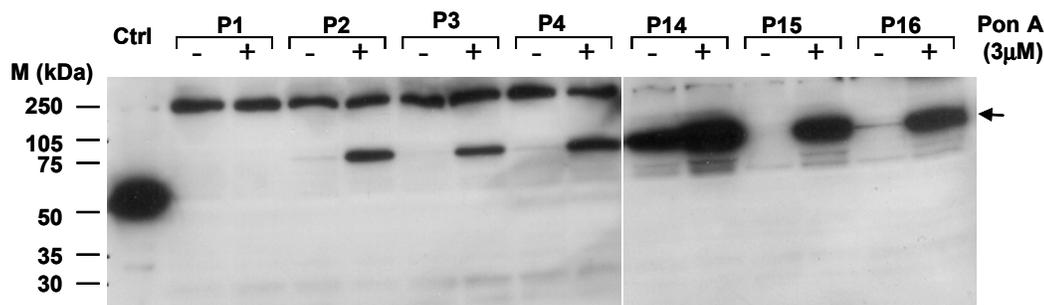
POLR1E (Polymerase (RNA) I polypeptide E) est une sous-unité spécifique de l'ARN Pol I. Elle joue un rôle critique dans l'initiation de la transcription des précurseurs des ARNr 5.8S, 18S et 28S en tant qu'intermédiaire d'interaction entre l'ARN Pol I et le facteur UBF (Hanada *et al.*, 1996; Yamamoto *et al.*, 2004). Situé près

de l'entrée du pore de l'enzyme, POLR1E forme un hétérodimère avec CD3EAP/hPAF49, qui ensemble servent de facteur d'élongation intégré ayant un effet de stimulation similaire au facteur TFIIF (Kuhn *et al.*, 2007).

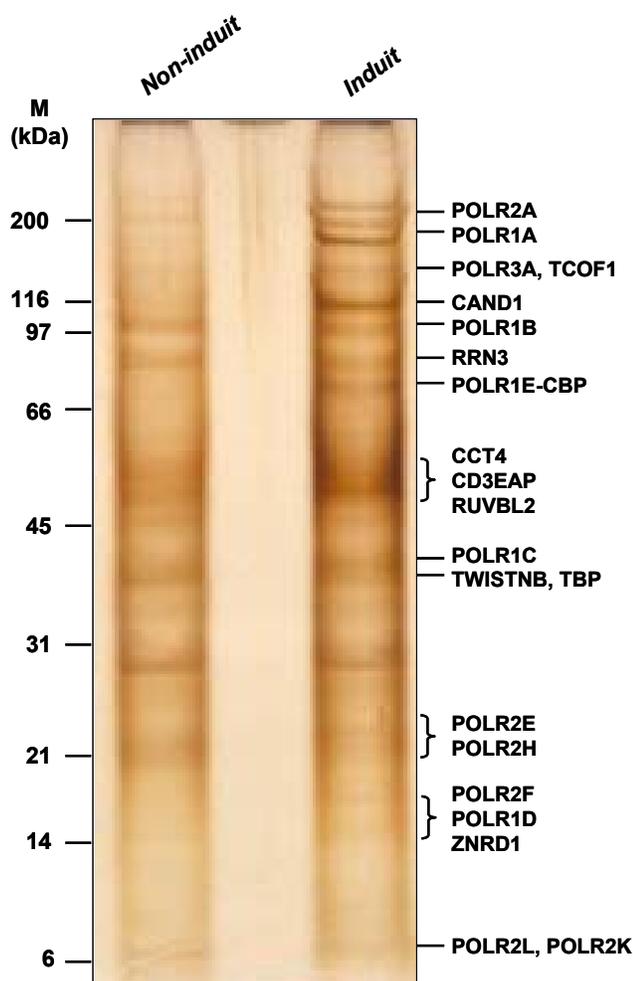
Le gène POLR1E encode une protéine de 481 aa ayant 30% d'identité avec RPA49, l'homologue de POLR1E chez la levure *S. cerevisiae*. RPA49 est une sous-unité non-essentielle, dont la délétion confère un phénotype conditionnel thermosensible (Liljelund *et al.*, 1992). Le polypeptide d'environ 54 kDa est composé principalement par le domaine A49 ARN polymérase I (aa 96 à 478) (**Figure 4-1**).



**Figure 4-1. Séquence d'acides aminés et représentation graphique des motifs et domaines de POLR1E.** RNA\_pol\_assoc\_fac\_A49-like – région d'homologie avec le facteur A49 chez *S. cerevisiae*; Carreaux verts et oranges – formes synonymes et non synonymes de variations d'une paire de bases.



**Figure 4-2. Analyse d'expression inductible de POLR1E.** Pour chaque lignée l'expression a été induite ou non avec 3 µM de Ponasterone A et le lysat cellulaire analysé par immunobuvardage de type western. La flèche indique la bande correspondant à la protéine étiquetée TAP.



**Figure 4-3. Purification des protéines associées à POLR1E.** Éluât de purification TAP des lignées stables ECR-293- POLR1E-TAP induites ou non avec 3  $\mu$ M de Ponasterone A migré sur gel SDS-PAGE 4-12 % et analysé par spectrométrie de masse (**Tableau 1** en annexes).

Le vecteur d'expression pMZI-POLR1E contient l'ADNc de POLR1E (Numéro d'accèsion : BC014331) fusionné à son extrémité 3' à la séquence de l'étiquette TAP. L'expression du gène a été dévoilée par immunobuvardage de type western à partir des lysats cellulaires. Cinq clones positifs ont été retenus et identifiés, soit P2, P3, P4, P15 et P16, affichant une bande intense à la hauteur approximative de 75 kDa, compatible avec la masse de POLR1E fusionné à l'étiquette TAP (taille approximative de 20 kDa) (**Figure 4-2**).

La purification TAP représentée à la **Figure 4-3** a été réalisée à partir du clone P15. L'analyse protéomique des résultats de purification met en évidence le complexe enzymatique de l'ARN Pol I contenant au moins les 13 sous-unités connues. Il s'agit de la forme compétente pour l'initiation de la transcription, étant donné que le complexe comprend le facteur de recrutement RRN3/TIF-IA et une partie du complexer SL1, soit les sous-unités TAF1C/TAFI110 et TBP. La protéine CAND1 connue pour interagir avec TBP et POLR2E se retrouvent aussi dans le complexe purifié à partir de POLR1E-TAP. TCOF1 est une phosphoprotéine nucléolaire connue pour avoir un rôle dans la transcription de l'ARNr. Sa présence dans le complexe est particulièrement intéressante et son rôle sera analysé plus en détails dans le chapitre Discussion.

D'autres protéines non reliées à la transcription de l'ARNr ont aussi été co-purifiées, notamment, POLR2A, POLR3A, CCT4 et RUVBL2 (**Figure 4-3**). Leur présence pourrait être expliquée par la formation de complexes intermédiaires via POLR2E et les autres sous-unités communes aux ARN polymérases I, II et III (POLR2F, POLR2H, POLR2K et POLR2L) et celles communes aux ARN Pol I et III (POLR1C et POLR1D). De plus, on constate la présence des chaperonnes moléculaires Hsp70, des tubulines (alpha et beta) et des karyophérines (**Tableau 1** en annexes).

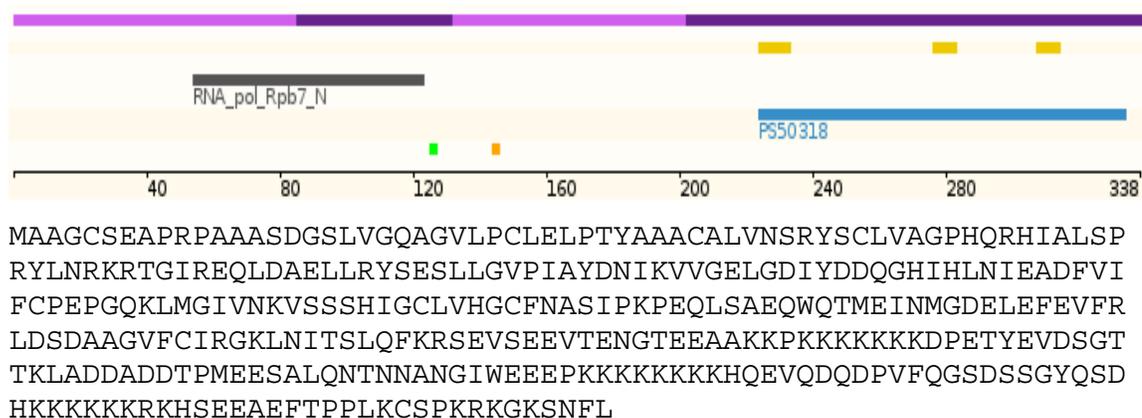
### 2.2.2. Purification TAP de TWISTNB/hRPA43

TWISTNB (TWIST *neighbor*), est une sous-unité spécifique de l'ARN Pol I qui est homologue à Rpb7 et Rpc25 des ARN Pol II et III eucaryotes, respectivement. L'homologue de TWISTNB chez la levure, RPA43, est essentiel à la viabilité cellulaire et à la synthèse de l'ARNr *in vivo* (Thuriaux *et al.*, 1995). Située à l'interface entre le cœur catalytique (via ABC23) (Bischler *et al.*, 2002) et les facteurs généraux de transcription (via RRN3) (Beckouet *et al.*, 2008), cette sous-unité occupe une position clé au sein de l'enzyme. Chez la levure, RPA14/RPA43 forment un hétérodimère structurellement et fonctionnellement équivalent à RPB4/RPB7 (Peyroche *et al.* 2002), qui est considérée comme une plateforme spécifique pour l'assemblage du complexe d'initiation (Kuhn *et al.*, 2007). Chez l'humain, l'homologue de RPA14 n'a pas encore été identifié, cependant, il est possible qu'une protéine de fonction similaire à RPA14

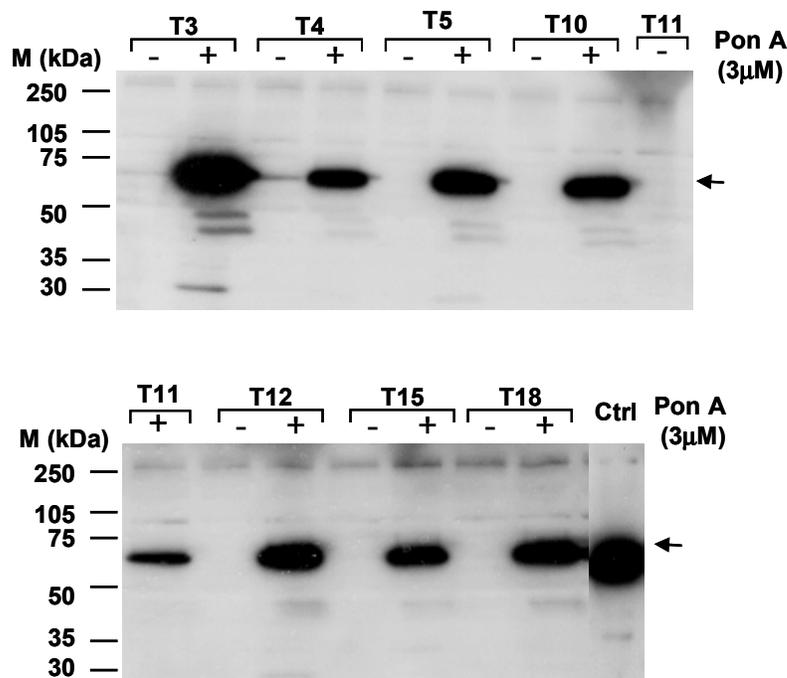
soit présente dans le génome humain mais reste non identifiée en raison de la divergence de séquence.

Le gène TWISTNB encode une protéine de 338 aa, ayant une masse moléculaire prédite de 37,4 kDa. Il fut identifié au sein d'une région génomique (7p21) qui est supprimée chez des patients atteints du syndrome de Saethre-Chotzen et de difficultés d'apprentissage spécifiques (Kosan *et al.*, 2002). La protéine partage 72 % d'identité avec son orthologue chez la souris. Elle contient une région d'homologie avec le domaine N-terminal de Rpb7 (aa 51 à 123) et 2 signaux de localisation nucléaire potentiels à son extrémité C-terminale, riche en résidus lysine (PS50318) (**Figure 4-4**).

L'étiquette TAP a été placée en position C-terminale de TWISTNB (Numéro d'accès : BC130300) et l'expression inductible de la protéine recombinante a été testée. Plusieurs clones positifs ont été retenus : T5, T10, T11, T12, T15 et T18. Les bandes de faible intensité observées en-dessous de la bande principale sont des artefacts (**Figure 4-5**). Le clone T11 a fait l'objet de purification TAP et les résultats sont représentés à la **Figure 4-6**.

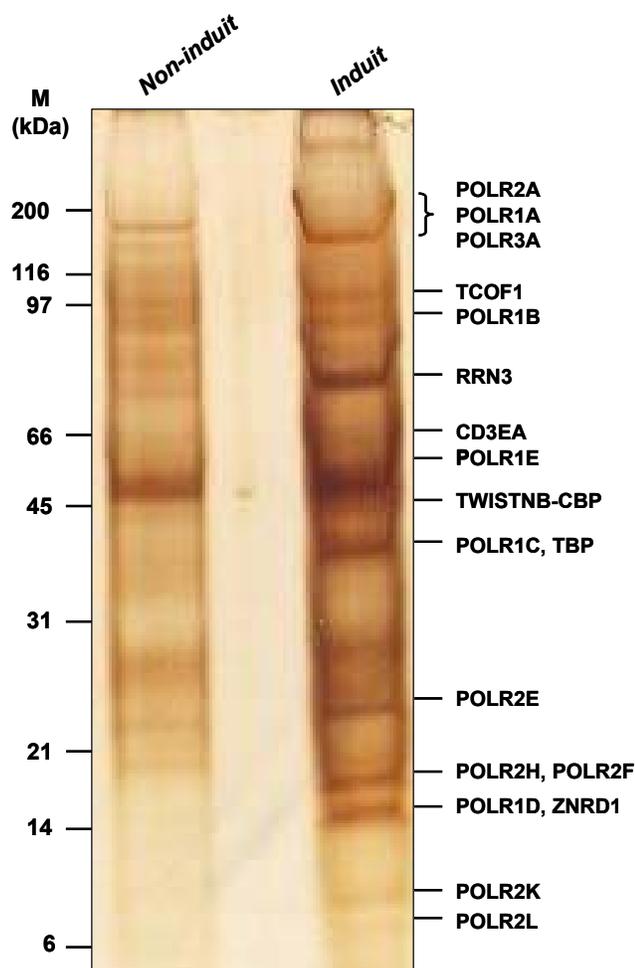


**Figure 4-4. Séquence d'acides aminés et représentation graphique des motifs et domaines de TWISTNB.** En jaune – séquences de faible complexité; carreaux verts et oranges – formes synonymes et non synonymes de variations d'une paire de bases.



**Figure 4-5. Analyse d'expression inducible de TWISTNB.** Pour chaque lignée l'expression a été induite ou non avec 3  $\mu$ M de Ponasterone A et le lysat cellulaire analysé par immunobuvardage de type western. La flèche indique la bande correspondant à la protéine étiquetée TAP.

La purification TAP de TWISTNB nous a permis de valider l'existence d'un réseau d'interaction comprenant les 13 sous-unités du complexe de l'ARN Pol I, les sous-unités TAF1B/TAFI63 et TBP du facteur de transcription basal SL1, les facteurs RRN3 et TCOF1. On note la présence d'autres interacteurs communs avec POLR1E, comme les sous-unités POLR2A et POLR3A et plusieurs karyophérines (**Figure 4-3** et **Tableau 2** en annexes).



**Figure 4-6. Purification des protéines associées à TWISTNB.** Éluât de purification TAP des lignées stables ECR-293- TWISTNB-TAP, induites ou non avec 3  $\mu$ M de Ponasterone A, migré sur gel SDS-PAGE 4-12 % et analysé par spectrométrie de masse (**Tableau 2** en annexes).

### 2.3. Caractérisation du réseau d'interactions de la machinerie de transcription de l'ARN Pol II

Dans le but de poursuivre l'expansion du réseau d'interactions protéiques de la machinerie de transcription de l'ARN Pol II humaine, nous avons procédé à la purification des complexes protéiques solubles impliquant POLR2E, PFDN4, MBD2, XPA, CAND1 et PDCD5 comme appât. Ces protéines ont été identifiées antérieurement comme étant des composantes du réseau de l'ARN Pol II et/ou des partenaires d'interactions possibles de RPAP3, RPAP4/XAB1 et CCT5 (sous-unité du complexe

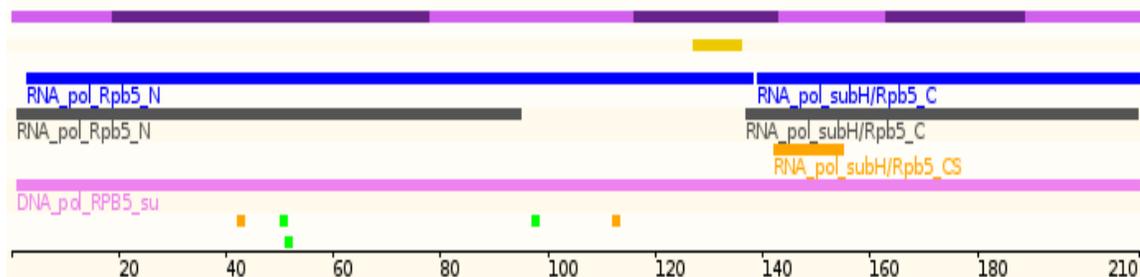
TRiC) (Jeronimo *et al.*, 2007; Cloutier *et al.*, 2009; Nitta *et al.*, 2000; Lembo *et al.*, 2003).

### 2.3.1. Purification TAP de POLR2E/RPB5

POLR2E (polymerase (RNA) II polypeptide E) est une des cinq sous-unités communes des ARN polymérase I, II et III. Dans la structure de l'ARN Pol II, POLR2E fait partie de la mâchoire inférieure et se compose d'un domaine exposé (aa 1 à 139) et un domaine incorporé (aa 140 à 210), qui sont bien conservés entre levure et l'homme (Cramer *et al.*, 2001) (**Figure 4-7**). Chez la levure, Rpb5 et Rpb9 ensemble interagissent avec Rpb1 et cette interaction pourrait avoir un rôle dans le maintien de la structure de Rpb1 au niveau de la bulle de transcription (Zaros *et al.*, 2007). Le domaine exposé sert aux interactions avec les régulateurs transcriptionnels et comprend deux sous-domaines, la partie N-terminale (aa 1 à 47) et la partie centrale de RPB5 (aa 57 à 139) reliées à travers une petite boucle. Le premier est critique pour la liaison de TFIIB et le dernier comporte les sites d'interaction avec la sous-unité RAP30 de TFIIF, la protéine X du virus de l'hépatite B (HBx) et RMP/URI (Wei *et al.*, 2001; Luvsanjav *et al.*, 2004; Le *et al.*, 2005).

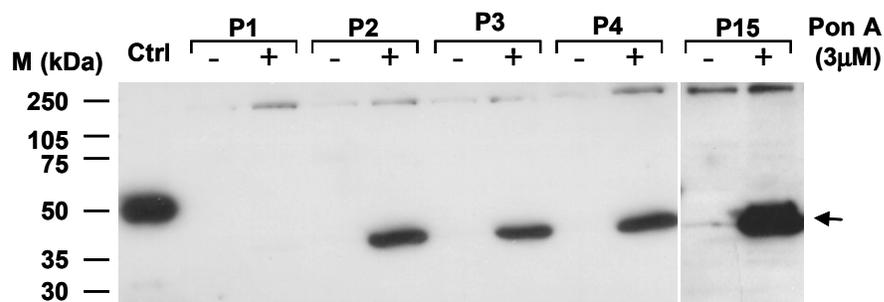
POLR2E a été copurifié avec le complexe RPAP3/R2TP/*prefoldin-like* (Gstaiger *et al.* 2003; Sardu *et al.* 2008; Cloutier *et al.* 2009), mais il n'était pas certain s'il faisait partie du complexe ou s'il était plutôt un substrat, vue l'activité chaperonne de ce complexe. Ici, nous avons procédé à la purification TAP de POLR2E afin de vérifier la bidirectionnalité des interactions dans le complexe.

Le sous-clonage de POLR2E (Numéro d'accèsion : BC034144) a été réalisé initialement avec l'étiquette TAP à l'extrémité C-terminale de la protéine. De cette manière, le complexe RPAP3 a été purifié mais pas les ARN polymérase (résultats non-présentés). Ce résultat suggérait que la présence de l'étiquette TAP en C-terminal de POLR2E pouvait avoir un impact négatif sur l'assemblage des sous-unités polymérase. Nous avons, donc, refait l'expérience en plaçant l'étiquette en N-terminal, au niveau du domaine exposé. La **Figure 4-8** montre que les clones P2, P3, P4 et P15 expriment la protéine étiquetée de façon inductible (bandes intenses vers 45 kDa) et peuvent, donc, être utilisés pour la purification TAP.

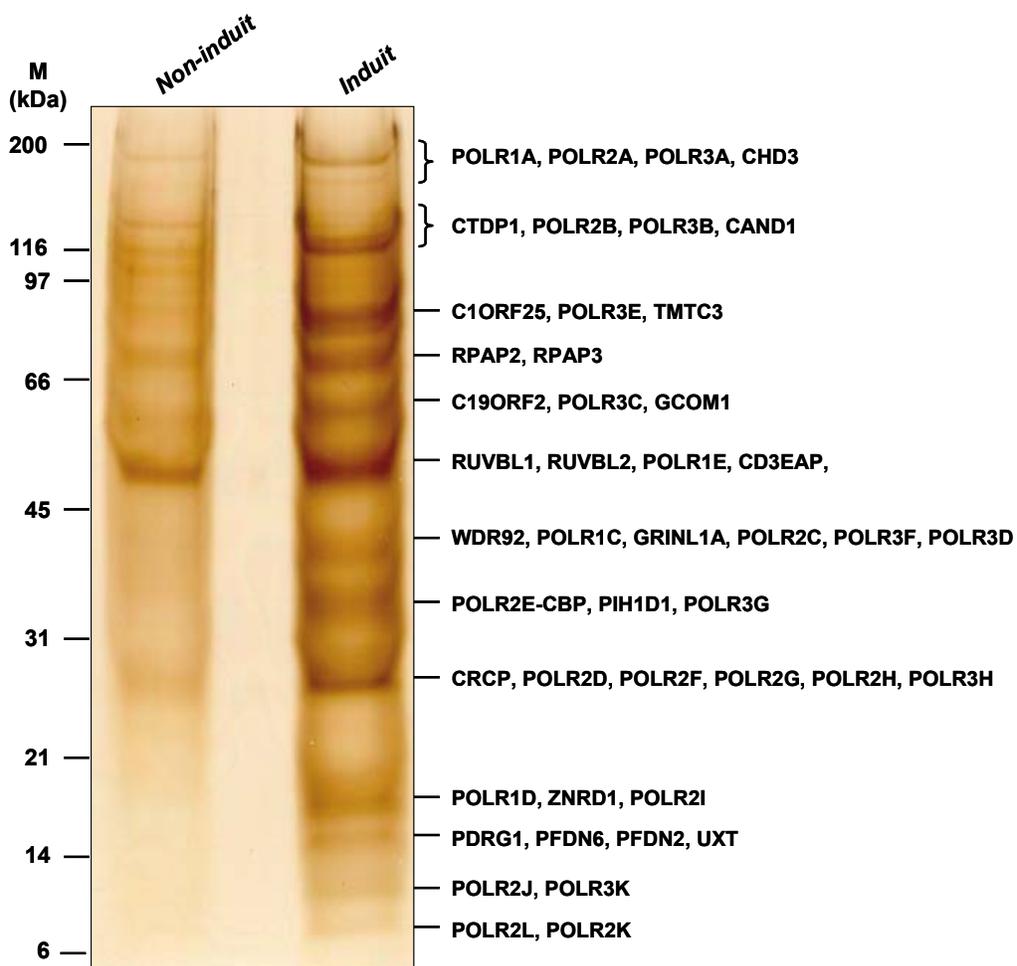


MDDEEETYRLWKIRKKTIMQLCHDRGYLVTQDELDTLEEFKAQSGDKPSEGRPRRTDLTVL  
 VAHNDDPTDQMFVFFPEEPKVGIKTIKVYCQRMQEENITRALIVVQQGMTPSAKQSLVDMA  
 PKYILEQFLQQELLINITEHELVPHEVVMTKEEVTELLARYKLRENQLPRIQAGDPVARYF  
 GIKRGQVVKIIRPSETAGRYITYRLVQ

**Figure 4-7. Séquence d'acides aminés et représentation graphique des motifs et domaines de POLR2E.** En jaune – séquences de faible complexité; carreaux verts et oranges – formes synonymes et non synonymes de variations d'une paire de bases.



**Figure 4-8. Analyse d'expression inductible de POLR2E.** Pour chaque lignée l'expression a été induite ou non avec 3  $\mu$ M de Ponasterone A et le lysat cellulaire analysé par immunobuvardage de type western. La flèche indique la bande correspondant à la protéine étiquetée TAP.



**Figure 4-9. Purification des protéines associées à POLR2E.** Éluât de purification TAP des lignées stables ECR-293-POLR2E-TAP, induites ou non avec 3  $\mu$ M de Ponasterone A, migré sur gel SDS-PAGE 4-12 % et analysé par spectrométrie de masse (**Tableau 3** en annexes).

Les partenaires d'interaction les plus abondants de POLR2E dans la fraction soluble sont des sous-unités du complexe RPAP3 et des ARN polymérase I, II et III, accompagnées de régulateurs transcriptionnels. Ceci démontre que POLR2E est réellement une composante du complexe RPAP3/R2TP/*prefoldin-like*. Parmi les régulateurs transcriptionnels on retrouve le complexe Médiateur avec son régulateur GRINL1A, certains facteurs généraux de transcription, CAND1, la sous-unité SUPT4H1 du facteur d'élongation DSIF, la phosphatase du CTD Fcp1/CTDP1, l'hélicase CHD3 et la protéine RPAP2 (**Figure 4-9** et **Tableau 3** en annexes).

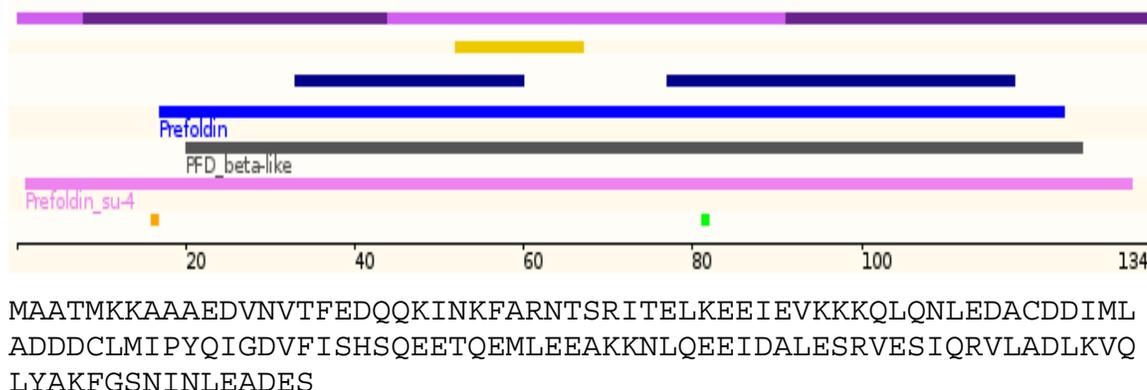
Parmi les autres protéines purifiées à partir de POLR2E-TAP nous retrouvons des facteurs impliqués dans la maturation des ARNm (hnRNPD, hnRNPE1, hnRNPE2, hnRNPE3, SNRPB), une ARNt méthyl-transférase hypothétique (similarité de séquence) C1orf25/TRM1L, des karyophérines (KPNA1, KPNA2, KPNB1, XPO1 et XPO5), des chaperonnes moléculaires HSP90 (HSP90AA1, HSP90AB1) et HSP70 (HSPA5, HSPA8) et des tubulines alpha (TUBA3E, TUBA4A, TUBA4B) et beta (TUBB2A, TUBB3, TUBB4, TUBB4Q).

### 2.3.2. Purification TAP de PFDN4

Le complexe RPAP3/R2TP/*prefoldin-like* que nous avons purifié contient plusieurs protéines de type préfoldines (C19orf2/URI, PDRG1 et UXT) en plus des préfoldines classiques (PFDN2 et PFDN6). Au tout début du projet de purification de ce complexe des résultats similaires avaient été rapportés dans deux études indépendantes (Gstaiger *et al.*, 2003; Sardu *et al.*, 2008). Une de ces études décrivait un complexe d'environ 1 MDa composé de Tip49/RUVBL1, Tip48/RUVBL2, PFD2/PFDN2, C19orf2/Uri (*unconventional prefoldin RBP5 interactor*), RPB5/POLR2E et plusieurs protéines aux fonctions inconnues similaires aux préfoldines. Parmi ces protéines il y avait deux préfoldines, une de type beta, PFDN4r (*PFD4-related*), et l'autre de type alpha, STAP1 (*SKP2-associated alpha PFD1*) (Gstaiger *et al.*, 2003). Les deux protéines y étaient mentionnées sans donner de référence à un gène spécifique. Nous étions intéressés à savoir si URI faisait partie de plusieurs complexes impliquant d'autres sous-unités préfoldines classiques. Pour vérifier cette hypothèse nous avons choisi PFDN4, une préfoldine classique de type beta (**Figure 4-10**) (voir section 2.3) et l'avons soumis à une purification TAP.

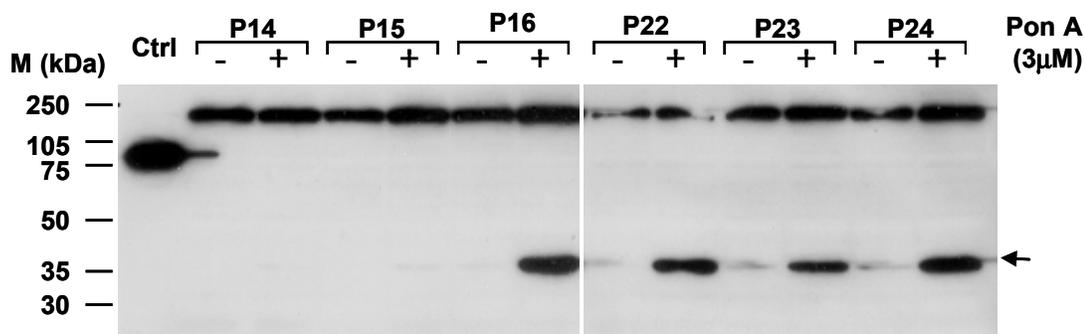
L'expression inductible de la protéine étiquetée (Numéro d'accension : BC010953) est constatée pour plusieurs clones, dont P16, P22, P23 et P24 par les bandes intenses situées vers 35 kDa. Les bandes de faible intensité, correspondant aux clones non-induits, ont été considérées comme des artéfacts.

L'analyse protéomique des résultats de purification TAP de PFDN4 met en évidence le complexe préfoldine sous sa forme canonique, mais aucune autre composante du complexe RPAP3/R2TP/*prefoldin-like* (**Figure 4-12** et **Tableau 4** en annexes).

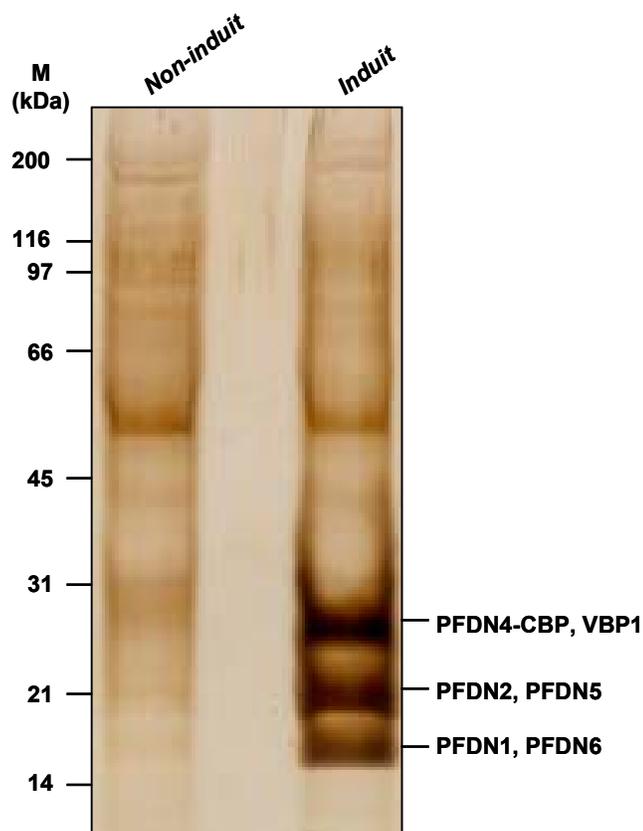


**Figure 4-10. Séquence d'acides aminés et représentation graphique des motifs et domaines de PFDN4.** En jaune – séquences de faible complexité; carreaux verts et oranges – formes synonymes et non synonymes de variations d'une paire de bases.

Comme appui, la purification TAP d'une autre sous-unité préfoldine, VBP1/PFDN3 (résultats non présentés), a également abouti vers le complexe préfoldine classique uniquement. Plus tard, en effectuant une recherche d'homologie pour PDRG1 nous avons constaté une ressemblance entre cette protéine avec PFDN4 (environ 25 % d'identité et 53 % de similitude) et nous avons conclu que PDRG1 est probablement la protéine PFDN4r mentionnée dans Gstaiger *et al.*, (2003). De la même manière la protéine STAP1 fut par la suite identifiée à UXT ce qui permit d'identifier le complexe décrit dans cet article comme étant identique au complexe RPAP3/R2TP/*prefoldin-like*. Cependant, la purification de PFDN4 a permis l'identification de KIN, une protéine nucléaire impliquée dans la réponse au stress génotoxique (Masson *et al.*, 2003), qui fut également identifiée dans les purifications précédentes de PFDN2 et PFDN6. La purification ultérieure de KIN n'a pas permis de confirmer la bidirectionnalité de cette interaction (résultats obtenus par Philippe Cloutier).



**Figure 4-11. Analyse d'expression inductible de PFDN4.** Pour chaque lignée l'expression a été induite ou non avec 3  $\mu\text{M}$  de Ponasterone A et le lysat cellulaire analysé par immunobuvardage de type western. La flèche indique la bande correspondant à la protéine étiquetée TAP.



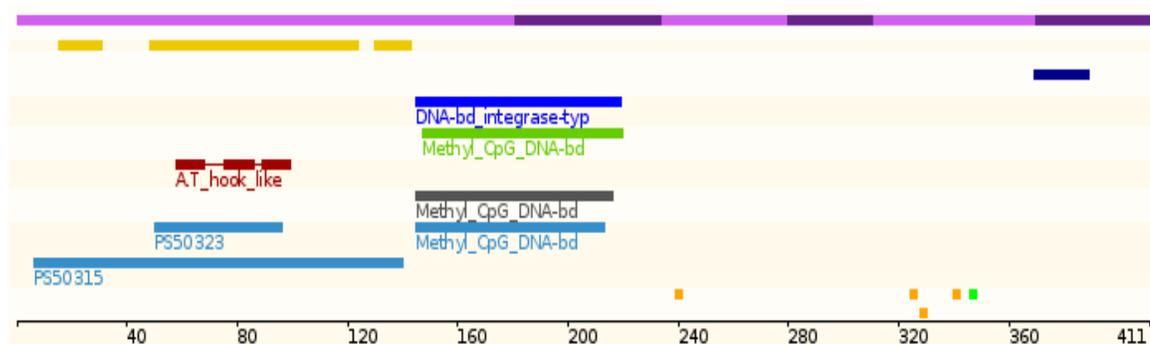
**Figure 4-12. Purification des protéines associées à PFDN4.** Éluât de purification TAP des lignées stables ECR-293-PFDN4-TAP, induites ou non avec 3  $\mu\text{M}$  de Ponasterone A, migré sur gel SDS-PAGE 4-12 % et analysé par spectrométrie de masse (**Tableau 4** en annexes).

### 2.3.3. Purification TAP de MBD2 et XPA

MBD2 est un facteur impliqué dans le remodelage de la chromatine. Il appartient à une famille de protéines capables de reconnaître et de se lier aux îlots CpG méthylés par un motif conservé appelé MBD (*methyl-CpG binding domain*) (Hendrich et Bird 1998). Cette protéine possède en plus un domaine TRD (transcriptional repression domain) en C-terminal qui se recouvre avec le domaine MBD et un motif GR (Glycine/Arginine-rich repeat) en N-terminal (**Figure 4-13A**). Cette protéine existe dans deux isoformes, la protéine complète, MBD2a, et la forme tronquée en N-terminal ne pouvant pas se lier aux CpG méthylés, MBD2b. En fonction de ses partenaires d'interaction, MBD2 peut agir à la fois comme un répresseur ou comme activateur de la transcription. En tant que répresseur, MBD2 fait partie des complexes MeCP1 (*Methyl-CpG binding Protein 1*) (Zhang *et al.* 1999) et Mi2/NuRD (*Nucleosome Remodeling and histone Deacetylation*) (Sansom *et al.*, 2007), interagit avec MBD3 et DNMT1 (Tatematsu *et al.*, 2000) ainsi qu'avec MIZF, un répresseur transcriptionnel spécifique (Sekimata et Homma, 2004). MBD2 est impliqué dans l'activation transcriptionnelle grâce à l'interaction avec les composants du complexe CREB (cyclic AMP-responsive element binding), comme l'ARN hélicase A (RHA), TACC3 et Tax (Fujita *et al.* 2003; Angrisano *et al.*, 2006; Ego *et al.*, 2005).

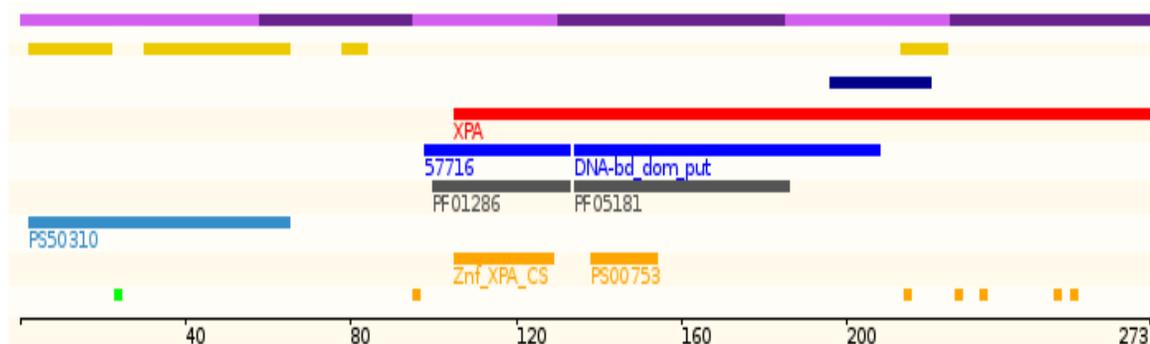
XPA (*Xeroderma pigmentosum group A*) est une protéine de 32 kDa impliquée dans la réparation de l'ADN par excision de nucléotide (NER, *nucleotide excision repair*) (Sancar et Reardon, 2006). Elle contient un motif à doigts de zinc qui lui confère une affinité pour les distorsions de la double hélice d'ADN (aa 98 à 219) (**Figure 4-13B**). Son domaine N-terminal comporte un domaine d'interaction avec RPA2 (aa 4 à 29) et un signal de localisation nucléaire (aa 30 à 42). La région contenant le groupe de l'acide glutamique ou E-cluster (aa 78 à 84) est importante pour l'interaction avec ERCC1. XPA lie aussi TFIIH via la région C-terminale (aa 226 à 273). La région centrale (aa 98 à 187) est nécessaire pour la liaison à RPA1 (Nitta *et al.*, 2000). Dans le processus du NER, XPA est recruté à l'ADN endommagé sous forme de complexe avec RPA et son rôle est de stabiliser l'ouverture de la double hélice autour de la lésion.

A.



MRAHPGGGRCCPEQEEGESAAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRRREGARGGGGRGR  
GRWKQAGRGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSLGGDGGGCGGGSGGGGAP  
RREPVPFPGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPS  
GKKFRSKPQLARYLGNLTVDLSSFDRTGKMMPSKLOKNKQRLRNDPLNQNGKPKDLNNTLP  
IRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMEL  
PKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTD  
EDIRKQEERVQQVRKKLEALMADILSRAADTEEMDIEMDSGDEA

B.



MAAADGALPEAAALEQPAELPASVRASIERKRQALMLRQARLAARPYSATAAAATGGMAN  
VKAAPKI IDTGGGFIL EEEEEEEQKIGKVVHQPGPVMEFDYVIC EECGKEFMDSYLMNHFD  
LPTCDNCRDADDKHLITKTEAKQEYLLKDCLEKREPPLKFI VKKNP HHSQWGMKLYLK  
LQIVKRSLEVWGSQEAL EEAKEVRQENREKMKQKKF DKKVKELRRAVRSSVWKRETIVHQH  
EYGPEENLEDDMYRKTCTMCGHELTYEKM

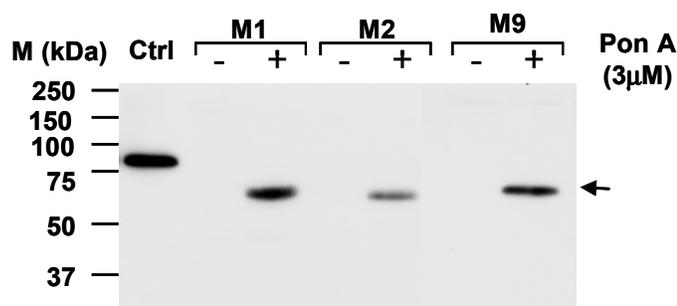
**Figure 4-13. Séquences d'acides aminés et représentations graphiques des motifs et domaines de MBD2 (A) et XPA (B).** En jaune – séquences de faible complexité; carreaux verts et oranges – formes synonymes et non synonymes de variations d'une paire de bases.

RPAP4/GPN1 (*RNAPII-associated protein 4/GPN-loop GTPase 1*) (aussi appelé MBDin/XAB1 pour *MBD2 interactor/XPA binding protein 1*) a été décrit précédemment comme étant une GTPase qui interagit avec MBD2 et le facteur de réparation XPA pour former un complexe (Lembo *et al.*, 2003; Nitta *et al.*, 2000). Malgré ces études le rôle de

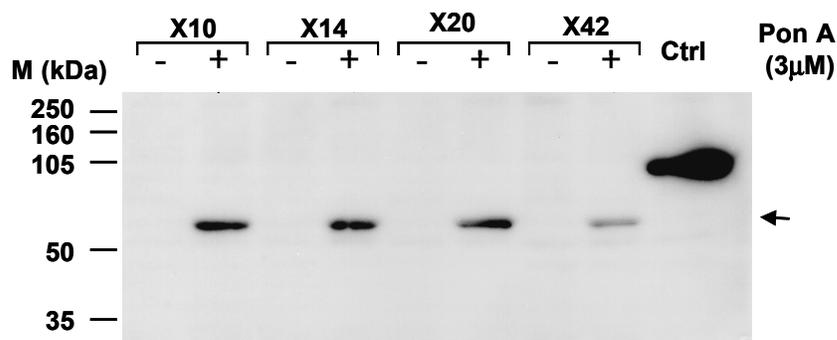
RPAP4/GPN1 n'était toujours pas élucidé. Dans notre réseau d'interaction (Jeronimo *et al.*, 2007) RPAP4 se situe à l'interface entre l'ARN Pol II et un groupe de facteurs impliqués dans le repliement et l'assemblage de complexes protéiques, incluant les complexes RPAP3, TRiC/CCT, préfoldines et certaines protéines de choque thermique, mais pas de MBD2 ou XPA. Pour confirmer ou infirmer l'interaction de ces protéines avec GPN1, nous avons procédé à l'étude du réseau d'interaction des complexes associés à MBD2 et XPA.

La **Figure 4-14** montre que la version complète de la protéine MBD2 (Numéro d'accèsion : BC032638) ainsi que la protéine XPA (Numéro d'accèsion : BC014965) recombinantes sont exprimées de façon inductible (bandes à 63 et 52 kDa respectivement).

A.



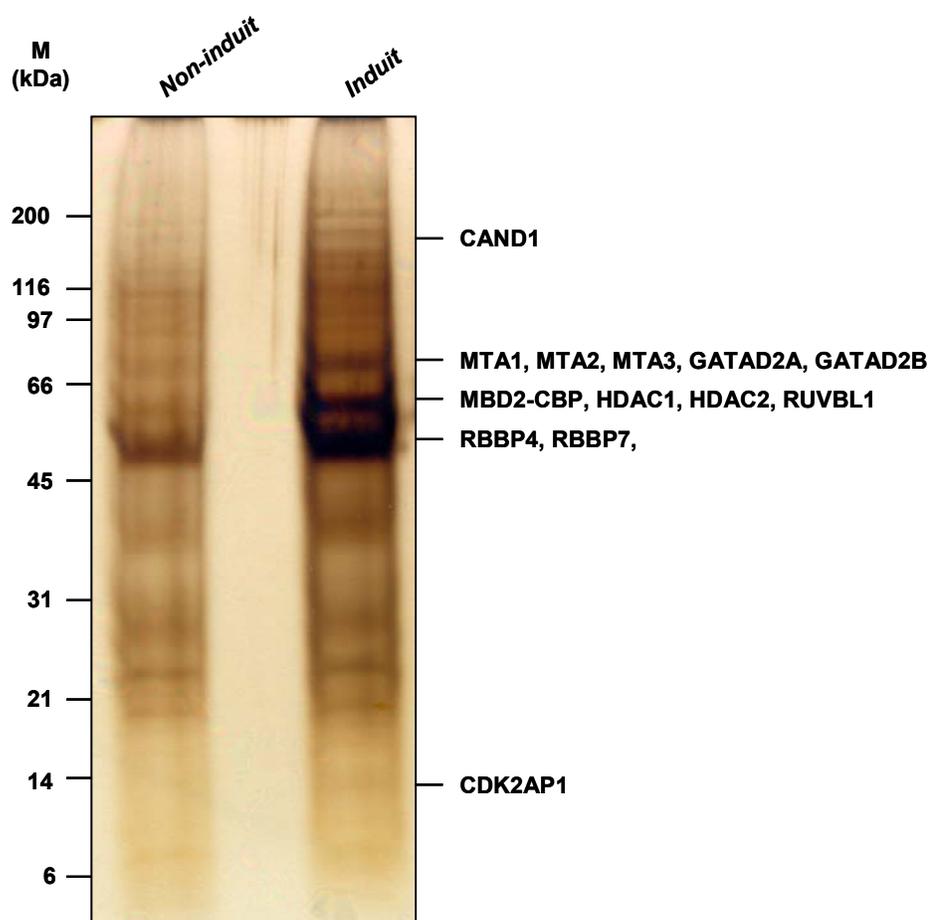
B.



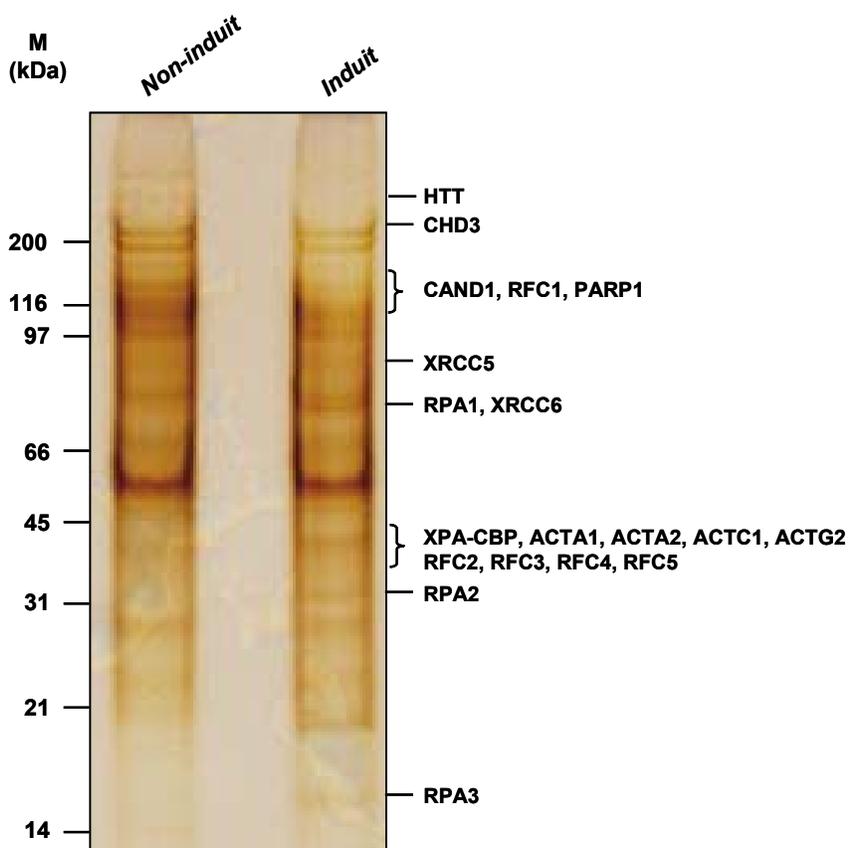
**Figure 4-14. Analyse d'expression inductible de MBD2 (A) et XPA (B).** Pour chaque lignée l'expression a été induite ou non avec 3  $\mu$ M de Ponasterone A et le lysat cellulaire analysé par immunobuvardage de type western. La flèche indique la bande correspondant à la protéine étiquetée TAP.

Les résultats de l'analyse protéomique n'ont pas permis de soutenir l'hypothèse sur le rôle de RPAP4/GPN1 dans la régulation chromatinienne ou la réparation de l'ADN, car la protéine n'a pas co-purifié ni avec MBD2, ni avec XPA, du moins pas dans la phase soluble (**Figure 4-15, Tableaux 5 et 6** en annexes). Ceci est en accord avec nos résultats sur son rôle dans la biogenèse de l'ARN Pol II, sujet qui sera détaillé dans le chapitre Discussion.

A.



B.



**Figure 4-15. Purification des protéines associées à MBD2 (A) et XPA (B).** Éluât de purification TAP des lignées stables ECR-293-MBD2-TAP (A) et ECR-293-XPA-TAP (B), induites ou non avec 3  $\mu$ M de Ponasterone A, migré sur gel SDS-PAGE 4-12 % et analysé par spectrométrie de masse (**Tableaux 5 et 6** en annexes).

Le réseau d'interactions de MBD2 met en évidence le complexe Mi-2/NuRD formé par les histones désacétylases HDAC1 et HDAC2, les protéines RbAp46/RBBP7, RbAp48/RBBP4, GATAD2A/p66 $\alpha$ , GATAD2B/p66 $\beta$ , MTA1-3, CDK2AP1/ DOC-1 et dépourvu des sous-unités Mi-2 $\alpha$ /CHD3 et Mi-2 $\beta$ /CHD4 (CHD : *chromodomain helicase DNA-binding protein*), responsables de l'activité de remodelage de la chromatine dépendante de l'ATP. Les interactions avec ces sous-unités ont probablement été perdues durant la purification, ou contrecarrées par l'étiquette TAP. Les protéines RUVBL1, CAND1, des chaperonnes moléculaires HSP90 et 70 et des tubulines alpha et beta font aussi partie du réseau d'interactions de MBD2 (**Figure 4-15A** et **Tableau 6** en annexes).

En ce qui concerne XPA, on le retrouve surtout en association avec des protéines impliqués dans le processus de reconnaissance et de réparation des lésions de l'ADN, comme le complexe RPA, le complexe RFC, l'hétérodimère Ku (XRCC5-XRCC6), PARP1 et PRKDC/ XRCC7. Parmi les autres protéines présentes dans cette purification se trouve CAND1, CHD3, HTT, HSPA2 ainsi que l'actine alpha (ACTA1, ACTA2 et ACTC1) et gamma (ACTG2) (**Figure 4-15B** et **Tableau 7** en annexes). Ces résultats permettent de soutenir l'hypothèse sur le rôle de la huntingtine (HTT) dans le maintien de l'intégrité génétique et les conséquences de son dysfonctionnement dans la maladie de Huntington (Goehler *et al.*, 2004; Enokido *et al.*, 2010) et seront développés plus en détail dans le chapitre Discussion.

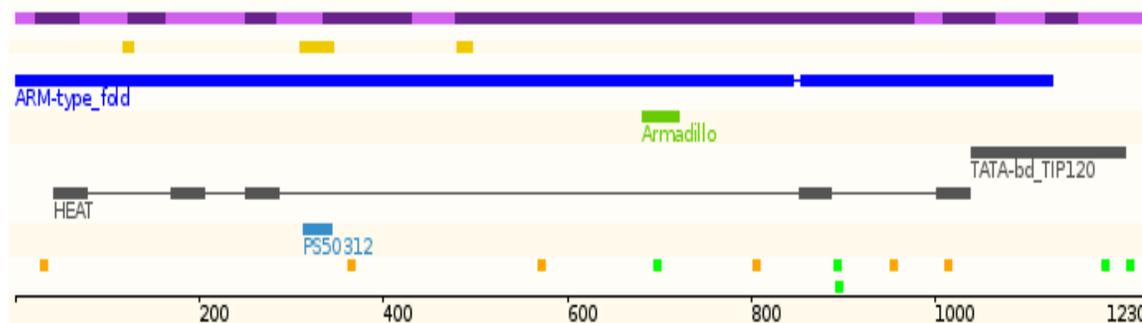
#### 2.3.4. Purification TAP de CAND1/TIP120A

CAND1/TIP120A (cullin-associated and neddylation-dissociated protein-1/TATA-binding protein interacting protein-120A) est une protéine de 120 kDa composée de 27 répétitions HEAT (huntingtin-elongation-A subunit-TOR) en tandem, dont l'empilement continu produit une structure super hélicoïdale extrêmement sinieuse (**Figure 4-19**) (Goldenberg *et al.*, 2004). CAND1/TIP120A a été initialement isolé par des techniques de purification d'affinité *in vitro* comme interagissant avec TBP. Elle avait été identifiée en tant que facteur de transcription capable de stimuler la transcription par les ARN polymérase I, II et III (Yogosawa *et al.*, 1996). Des études ultérieures ont révélé le rôle de CAND1 dans la régulation de l'activité des Culline-RING ubiquitine Ligases (CRLs), le dernier protagoniste de la réaction d'ubiquitination. CAND1 agit comme inhibiteur des CRLs en s'associant de manière spécifique avec les sous-unités Cullins déneddylées pour les séquestrer dans un état non assemblé et inactif (revu dans Bosu et Kipreos, 2008).

Il a été montré que le système ubiquitine-protéasome est impliqué à différents niveaux dans la régulation de l'expression des gènes autant via son activité protéolytique que non-protéolytique (revu dans Muratani et Tansey, 2003). Parmi les protéines que nous avons étiquetées au laboratoire, CAND1 a été co-purifiée avec LARP7, MEPCE et TCEA2, des facteurs qui modulent de façon directe ou indirecte le taux d'élongation de

la transcription par l'ARN Pol II dans les cellules humaines. Nous avons donc proposé un rôle possible de CAND1 comme régulateur de la transcription via le système ubiquitine-protéasome. Les autres protéines qui ont également co-purifié CAND1 sont la sous-unité CCT5 du complexe TRiC et la sous-unité UXT du complexe RPAP3.

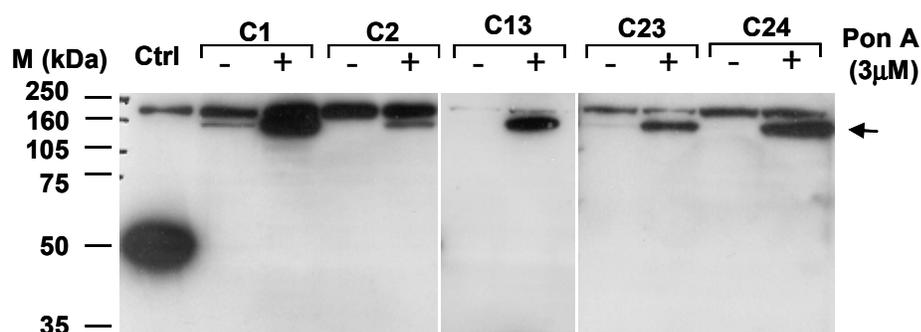
Le sous-clonage de CAND1 (Numéro d'accension : BC050341) a été réalisé avec l'étiquette TAP en C-terminal. Parmi les clones positifs présentés à la **Figure 4-17** (bande intense vers 150 kDa), le C-13 a été choisi pour la purification. L'analyse protéomique de l'éluât de purification CAND1-TAP n'a pas permis de confirmer les interactions avec LARP7, MEPCE ou TCEA2. Seulement les sous-unités CCT4 et CCT7 du complexe TRiC se retrouve dans la liste d'interacteurs potentiels avec de faibles scores IR (**Figure 4-18** et **Tableau 7** en annexes). Par contre, la présence de XPA et CHD3 n'était pas prévue et elle parait, pour le moins, intrigante. Des études récentes ont caractérisé le complexe Cul4A-DDB comme régulateur dans la voie NER en ciblant DDB2 et XPC (Liu *et al.*, 2009), mais aucun lien n'a encore été fait, à notre connaissance, entre CAND1 et XPA. Ce sujet sera discuté en plus de détails au chapitre Discussion.



```
MASASYHISNLLKMTSSDKDFRFRMATNDLMTELQKDSIKLDDDSERKVVKMILKLLLEDKN
GEVQNLAVKCLGPLVSKVKEYQVETIVDTLCTNMLSDEQLRDISSIGLKTIVIGELPPASS
GSALAAVCKKITGRRLTSAIAKQEDVSVQLEALDIMADMLSRQGGLLVNFHPSILTCLLPQ
LTSPRLAVRKRTIIALGHLVMSCGNIVFVDLIEHLLSELSKNDMSMSTRTYIQCIAAISRQ
AGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEVYPHVSTIINICLKYL
TYDPNYNYDDEDEDENAMDADGGDDDDQGSDDDEYSDDDDMSWKVRRRAAKCLDAVVSTRHE
MLPEFYKTVSPALISRFKEREENVKADVFHAYLSLLKQTRPVQSWLCPDAMEQGETPLTM
LQSQVPNIVKALHKQMKESVKTRQCCFNMLTELNVNLPGALTQHIPVLVPGIIFSLNDKS
SSSNLKIDALSCLYVILCNHSPQVFHHPVQALVPPVVACVGDPFYKITSEALLVTQQLVKV
IRPLDQPPSSFDATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNLGDNLGSDDLPN
TLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPIILASFLRKNQRALKGLTL
```

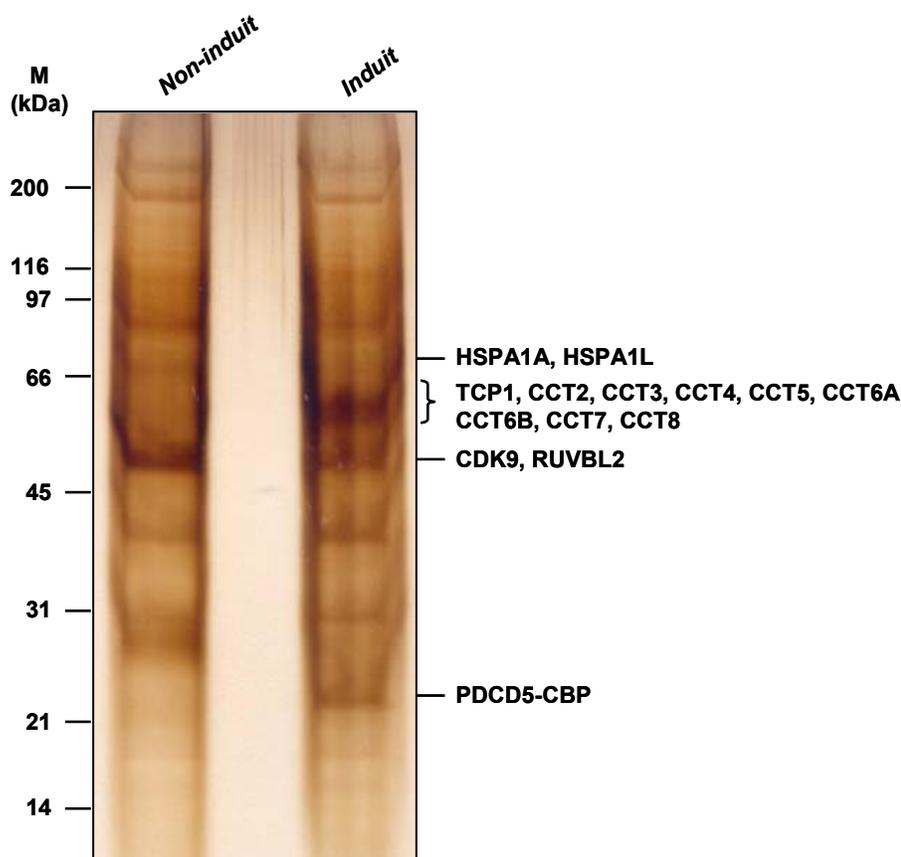
SALDILIKNYSDSLTAAMIDAVLDELPLISESDMHVSQMAISFLTTLAKVYPSSLSKISG  
 SILNELIGLVRSPLLQGGALSAMLDFQALVVTGTNNLGYMDLLRMLTGPVYSQSTALTHK  
 QSYYSIAKCVAAALTRACPKEGPAVVGQFIQDVKNRSTDSIRLLALLSLGEVGHHDLSGQ  
 LELKSVILEAFSSPSEEVKSAASYALGSI SVGNLPEYLPFVQLQEITSQPKRQYLLHSLKE  
 I I SSASVVGLKPYVENIWALLLKHCECAEEGTRNVVAECLGKLTLLIDPETLLPRLKGYLIS  
 GSSYARSSVVTAVKFTISDHPQPIDPLLKNCIGDFLKTLED PDLNVRVALVTFNSAAHNK  
 PSLIRDLLDTVLPFLYNETKVRKELIREVEMGPFKHTVDDGLDIRKAAFECMYTLLDSCLD  
 RLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQRLDRLVEPLRATCTTKVKA  
 NSVKQEFEKQDELKRSAMRAVAALLTIPEAEKSPLMSEFQSQISSNPELAAIFESI QKDSS  
 STNLESMDTS

**Figure 4-16. Séquence et représentation graphique des motifs et domaines de CAND1.** En jaune – séquences de faible complexité; carreaux verts et oranges – formes synonymes et non synonymes de variations d’une paire de bases.



**Figure 4-17. Analyse d’expression inductible de CAND1.** Pour chaque lignée l’expression a été induite ou non avec 3  $\mu$ M de Ponasterone A et le lysat cellulaire analysé par immunobuvardage de type western. La flèche indique la bande correspondant à la protéine étiquetée TAP.

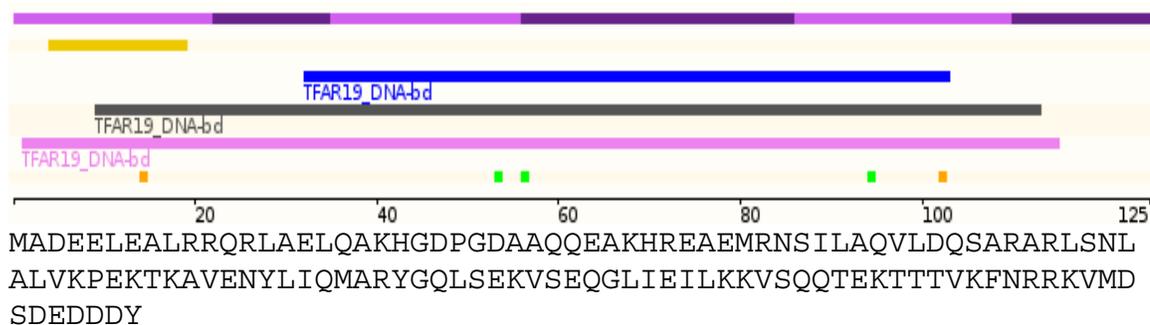
Les culline-1, -3, -4B et -5 sont présentes dans le réseau d’interaction de CAND1, mais pas culline-2, -4A et -7. Ceci est en accord avec l’observation selon laquelle, dans certains types cellulaires, CAND1 lie préférentiellement un sous-ensemble spécifique de cullines (Bossu et Kipreos, 2008). Le réseau d’interactions de CAND1 contient aussi les protéines Rbx1, RNF7, DDB1 ainsi que les chaperonnes moléculaires HSPA1L et HSPA2 (**Figure 4-18** et **Tableau 7** en annexes).



**Figure 4-18. Purification des protéines associées à CAND1.** Éluât de purification TAP des lignées stables ECR-293-CAND1-TAP, induites ou non avec 3  $\mu$ M de Ponasterone A, migré sur gel SDS-PAGE 4-12 % et analysé par spectrométrie de masse (**Tableau 7** en annexes).

### 2.3.5. Purification TAP de PDCD5/TFAR19

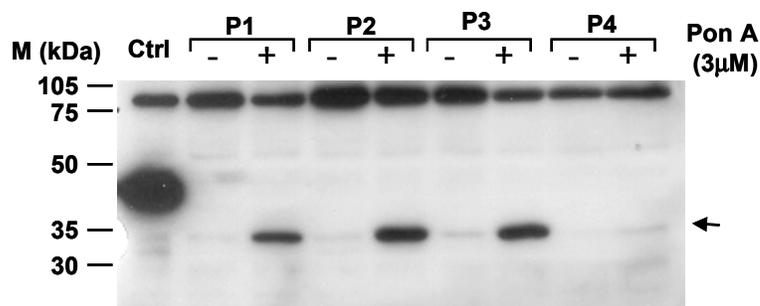
PDCD5/TFAR19 (human programmed cell death 5/ TF-1 cell apoptosis-related gene 19) est un gène humain cloné à partir des cellules de TF-1 en apoptose. La protéine PDCD5 humaine (125 aa) comporte un domaine potentiel de liaison à l'ADN (aa 22 à 112) (**Figure 4-19**) et peut être divisée en trois régions structurales : une région centrale rigide et deux régions terminales distinctes. La région centrale (aa 41 à 101) représente un sous-domaine rigide formé de trois hélices  $\alpha$ . Les résidus 102 à 125 en C-terminal représentent une région flexible non structurée avec un segment très mobile (aa 116 à 125). Le fragment N-terminal correspondant aux résidus 1 à 26 peut former une hélice  $\alpha$  stable indépendamment du noyau de la protéine (Liu *et al.*, 2005).



**Figure 4-19. Séquence et représentation graphique des motifs et domaines de PDCD5.** En jaune – séquences de faible complexité; carreaux verts et oranges – formes synonymes et non synonymes de variations d’une paire de bases.

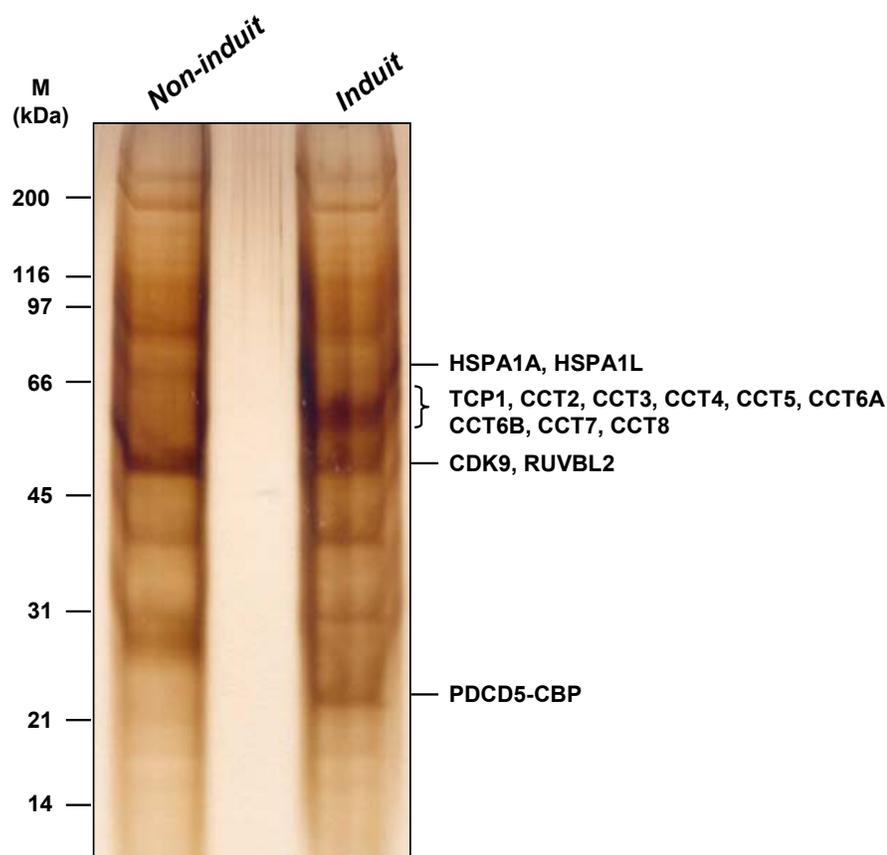
La protéine PDCD5 est un régulateur important de l’apoptose et de la paraptose. Elle est surexprimée pendant ces deux processus distincts de mort cellulaire programmée (Liu *et al.*, 1999; Wang *et al.*, 2004) et sous-exprimée dans certaines tumeurs humaines (Chen *et al.*, 2010). Avant l’induction de l’apoptose, PDCD5 est distribué autant dans le noyau que dans le cytoplasme. Une fois l’apoptose induite, la quantité de PDCD5 augmente et s’accumule dans le noyau (Chen *et al.*, 2001). Il a été montré que le domaine N-terminal de PDCD5 comprenant l’hélice  $\alpha$  serait nécessaire pour exercer l’effet pro-apoptotique (Liu *et al.*, 2005), tandis que le domaine C-terminal flexible était essentiel pour le transport intercellulaire (via les radeaux lipidiques et la liaison aux chaînes héparane sulfate des protéoglycanes de la matrice extracellulaire) (Wang *et al.*, 2006). Une étude récente a aussi montré que PDCD5 lie Tip60/KAT5, une histone acétyltransférase, et agit en tant que co-activateur de l’apoptose via la voie de signalisation Tip60-p53 (Xu *et al.*, 2009).

Dans le réseau de l’ARN Pol II, PDCD5 a été identifié comme partenaire d’interaction de la sous-unité epsilon (CCT5) du complexe TRiC. Cette interaction a soulevé certains questionnements concernant le rôle que cette chaperonine pourrait avoir dans la régulation de PDCD5 dans le contexte de son repliement et son transport à travers les membranes. Mais, avant de formuler des hypothèses sur ce sujet, nous avons voulu d’abord vérifier la bidirectionnalité des interactions PDCD5-TRiC.



**Figure 4-20. Analyse d'expression inductible de PDCD5.** Pour chaque lignée l'expression a été induite ou non avec 3  $\mu$ M de Ponasterone A et le lysat cellulaire analysé par immunobuvardage de type western. La flèche indique la bande correspondant à la protéine étiquetée TAP.

PDCD5 (Numéro d'accension : BC015519) a été étiquetée en C-terminal et des lignées cellulaires ECR-293 ont été créées. L'expression inductible a été constatée pour plusieurs clones identifiés par la bande intense située au niveau de 35 kDa à la **Figure 4-20**. Le clone P1 a été retenu pour la purification TAP. L'analyse protéomique des résultats de purification TAP de PDCD5 montre que les huit sous-unités de chaque anneau du complexe TRiC/CCT co-purifient avec ce dernier (**Figure 4-21** et **Tableau 8** en annexes). De plus, deux chaperonnes moléculaires HSP70 (HSPA1A et HSPA1L), RUVBL2 et CDK9 se retrouvent dans la composition du complexe de purification PDCD5-TAP.



**Figure 4-21. Purification des protéines associées à PDCD5.** Éluât de purification TAP des lignées stables ECR-293-PDCD5-TAP, induites ou non avec 3  $\mu$ M de Ponasterone A, migré sur gel SDS-PAGE 4-12 % et analysé par spectrométrie de masse (**Tableau 8** en annexes).

#### 2.4. Représentation graphique des interactions protéine-protéine

L'ensemble des interactions protéine-protéine provenant de toutes les purifications TAP effectuées durant ce projet de maîtrise ont été rassemblées sous forme d'un un réseau, monté à l'aide du logiciel VisANT (<http://visant.bu.edu>). La totalité des interactions qui ont été validées statistiquement (score de confiance supérieur à 0.6729) est illustrée à la **Figure 4-22**. Le graphique comprend les complexes ARN polymérase I, II et III, RPAP3/R2TP/*prefoldin-like*, PFD, TRiC/CCT, Mi-2/NuRD, des facteurs de transcription et de réparation de l'ADN, etc. Il s'agit de complexes purifiés à partir de la fraction cellulaire soluble ECR-293.

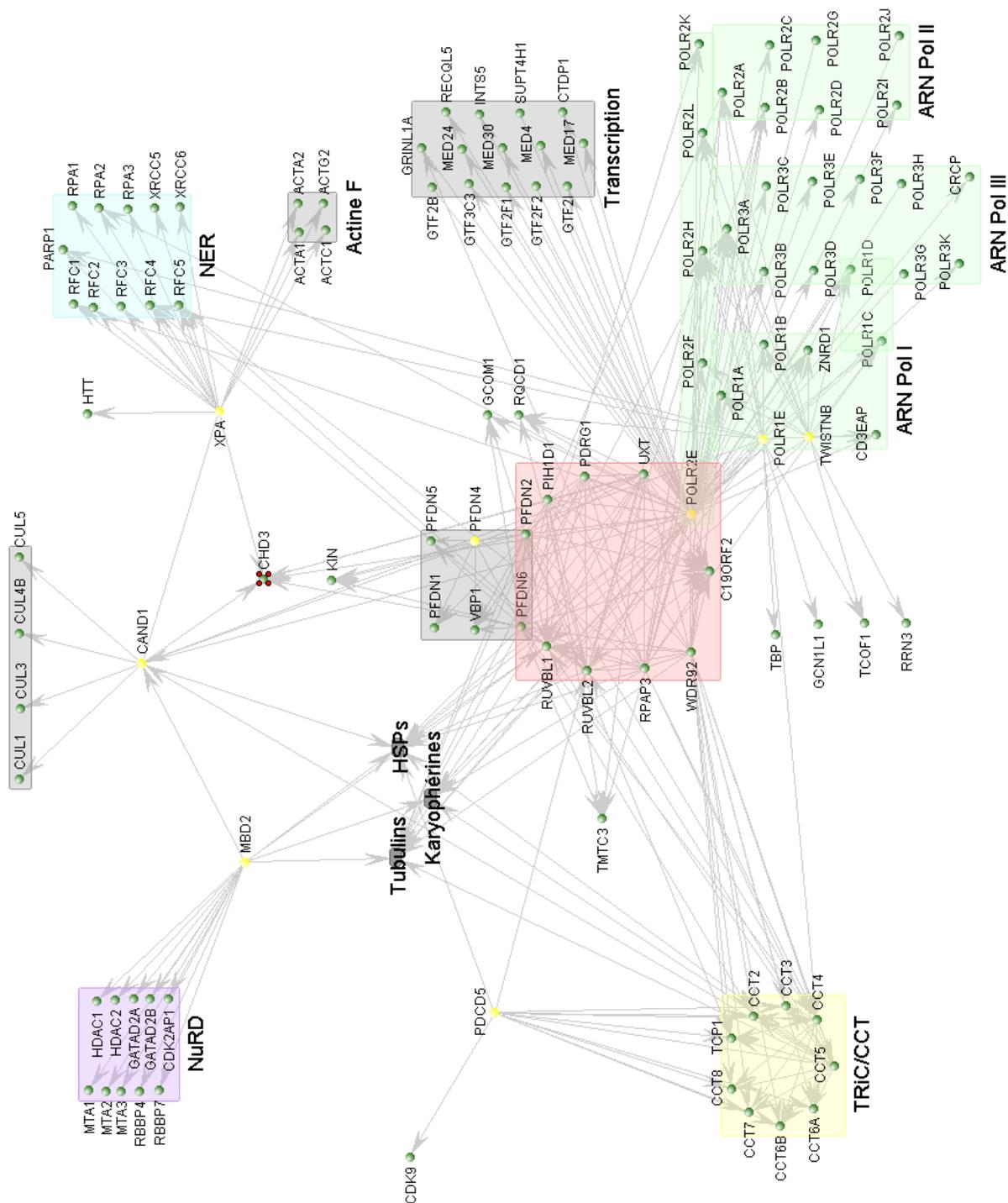


Figure 4-22. Réseau d'interactions connectant les ARN Pol I, II et III, les complexes RPAP3/R2TP/*prefoldin-like*, TRiC/CCT, Mi-2/NuRD et des facteurs de transcription et de réparation de l'ADN.

### 3. DISCUSSION

#### 3.1. Identification à haut débit des interactions protéine-protéine

La combinaison de méthodes de purification de protéines et de la spectrométrie de masse est une des méthodes les plus utilisées pour détecter les interactions protéiques. La méthode de purification TAP (*tandem affinity purification*) permet la purification de complexes protéiques sous forme native de manière hautement reproductible et à un degré de pureté compatible avec l'analyse en spectrométrie de masse. Elle est applicable de façon systématique, ce qui permet de l'utiliser dans des projets de construction de cartes d'interactions protéine-protéine à l'échelle du protéome.

Dans notre laboratoire, cette méthode a été adaptée pour la purification de complexes protéiques à partir de cultures de cellules humaines. On utilise un système d'expression inductible à l'ecdysone qui empêche la surexpression de la protéine étiquetée, ce qui limite la formation d'interactions protéiques non-spécifiques. De plus, l'approche bioinformatique, basée sur un algorithme développé dans notre laboratoire, nous permet de sélectionner des interactions protéiques avec un haut niveau de confiance (Jeronimo *et al.*, 2007). Les cartes d'interactions créées à partir de l'ensemble de données permettent non seulement d'obtenir une vue d'ensemble du fonctionnement du protéome, mais représente aussi une source précieuse de nouvelles interactions protéine-protéine, pouvant mener à la découverte de nouveaux modules fonctionnels.

Malgré tous les avantages mentionnés, l'utilisation de la méthode de purification TAP a des limites. Elle permet, avant tout, d'étudier les complexes protéiques stables formés *in vivo* dans la fraction cellulaire soluble, mais pas les complexes présents dans la fraction insoluble (ADN, membranes), ni leur comportement dynamique. Il faut aussi garder à l'esprit qu'il s'agit d'une méthode de co-purification qui ne permet pas d'établir si les interactions sont directes. Finalement, comme pour toute autre méthode, sa performance est affectée par son taux de spécificité. Afin de minimiser les taux de faux négatifs et de faux positifs, des améliorations constantes sont apportées surtout au niveau des outils de détection qui y sont associés, comme la spectrométrie de masse et la bioinformatique (Cloutier *et al.*, 2009; Lavalée-Adam *et al.*, 2010).

### 3.2. Caractérisation du complexe de l'ARN Pol I

Les connaissances actuelles sur la composition de l'ARN Pol I proviennent des études génétiques et biochimiques menées chez la levure. Contrairement aux sous-unités des ARN Pol II et III, les sous-unités de l'ARN Pol I sont généralement moins bien conservées (Werner *et al.*, 2009). À partir des 14 sous-unités constitutives de l'ARN Pol I de *S. cerevisiae*, l'orthologue mammifère pour la sous-unité A14 n'a pas encore été identifié. Dans le but de cartographier le réseau d'interaction de l'ARN Pol I humaine nous avons étiqueté les sous-unités spécifiques POLR1E et TWISTNB et soumis à la procédure de purification TAP couplée à la spectrométrie de masse.

Ensemble, les deux protéines étiquetées permettent la purification d'un module comprenant, entre autre, le complexe de l'ARN Pol I avec les 13 sous-unités connues, le facteur de recrutement RRN3/TIF-IA, une fraction du facteur SL1 (TAF1B/TAFI63, TAF1C/TAFI110 et TBP) et la protéine TCOF1. Il a été montré que l'ARN Pol I - RRN3, la seule forme d'enzyme compétente pour l'initiation de la transcription, est recrutée sur le promoteur de l'ADNr grâce aux interactions entre RRN3 et les sous-unités TAF1B/TAFI63 et TAF1C/TAFI110 du complexe SL1/TIF-IB (Miller *et al.*, 2001). Les faibles scores IR des facteurs TAF1B/TAFI63 et TAF1C/TAFI110 (0.06 et 0.05) (**Tableau 1** en annexes) peuvent être perçus soit comme un indice de leur faible représentativité, soit comme des interactions transitoires. On peut donc supposer que le complexe de l'ARN Pol I – RRN3 peut s'associer à un sous-complexe SL1 (TAF1B/TAFI63 – TAF1C/TAFI110 – TBP) pour former un dernier intermédiaire métastable (tel que proposé par Dundr *et al.*, 2002) avant le recrutement de l'enzyme à la matrice d'ADN via les autres sous-unités du complexe SL1 et UBF.

Treacle, le produit du gène TCOF1, est une protéine nucléolaire connue pour avoir un rôle dans la transcription de l'ADNr, mais dont le mécanisme précis d'action demeure incertain. Des mutations non-sens dans le gène TCOF1 entraînent une apoptose des cellules de la crête neurale, causant la dysostose mandibulofaciale (Treacher Collins syndrome), en réponse à l'incapacité de ces cellules à compléter la synthèse de l'ADN ribosomal (Valdez *et al.*, 2004; Dixon *et al.*, 2006). Treacle co-localise avec la

machinerie de transcription de l'ARN Pol I au nucléole et aux NOR (*Nucleolar Organizer Regions*) lors des différentes phases du cycle cellulaire (Valdez *et al.*, 2004; Lin et Yeh, 2009). Lin et Yeh ont montré que Treacle est essentielle au recrutement de l'ARN Pol I au nucléole. Elle aurait aussi un rôle dans la méthylation du pré-ARNr (Gonzales *et al.*, 2005). La protéine possède un signal de localisation nucléaire à son extrémité N-terminale, un domaine d'interaction avec l'ARN Pol I dans la région centrale, un signal de localisation nucléolaire, un domaine d'interaction avec UBF et le promoteur ribosomal en C-terminal (Lin et Yeh, 2009). C'est une phosphoprotéine contenant 1411 aa dont les résidus forment un grand nombre de sites de phosphorylation pour la caséine kinase 2 et la protéine kinase C.

Chez la levure RPA14 est une sous-unité spécifique non-essentielle, dont la déplétion résulte en une enzyme moins performante, mais toujours fonctionnelle. En revanche, le double mutant RPA14 avec une autre sous-unité non-essentielle (RPA12.2, RPA 34.5 ou RPA 49) est létal (Gadal *et al.*, 1997). Il a été montré que RPA14 forme un hétérodimère stable avec RPA43 qui interagit à son tour avec la sous-unité commune ABC23/Rpb6. Dans ce contexte, le rôle de RPA14 est de stabiliser l'interaction entre ces sous-unités au sein du complexe (Bishler *et al.*, 2002; Peyroche *et al.*, 2002). Dans la structure tridimensionnelle de l'enzyme, l'hétérodimère RPA14/RPA43 forme une protubérance positionnée au même endroit que le CTD de Rpb1 occupe dans l'ARN Pol II et pourrait remplir des fonctions similaires (Peyroche *et al.*, 2002). Étant donné que des homologues de RPA14 ont été identifiés uniquement chez des Ascomycètes, il est possible que chez l'humain TCOF1 ne serve pas uniquement au recrutement de l'ARN Pol, mais qu'il puisse remplir des fonctions similaires à RPA14.

Pour confirmer cette hypothèse, il serait intéressant de vérifier d'abord, si Treacle interagit directement avec TWISTNB, en effectuant un immunobuvardage de type far western, et deuxièmement, vérifier s'il pouvait avoir un rôle dans la transcription de l'ADNr à partir d'un promoteur spécifique.

### 3.3. Étude protéomique de la machinerie de l'ARN polymérase II

#### 3.3.1. POLR2E/RPB5 est une sous-unité constitutive du complexe RPAP3/R2TP/*prefoldin-like*

Comme mentionné dans les sections précédentes, POLR2E/RPB5 est une sous-unité commune aux trois ARN polymérases nucléaires et a été copurifiée dans des études indépendantes avec le complexe RPAP3/R2TP/*prefoldin-like* (Gstaiger *et al.* 2003, Sardu *et al.*, 2008; Cloutier *et al.* 2009). Nos résultats de purification confirment que POLR2E est une composante constitutive du complexe RPAP3/R2TP/*prefoldin-like* et pas uniquement la sous-unité des ARN polymérases par laquelle ce dernier interagit.

Le complexe RPAP3/R2TP/*prefoldin-like* est composé de protéines jouant un rôle dans l'assemblage des complexes protéiques, comme certains cofacteurs de chaperonne et des préfoldines (voir section 2.3). Cette constatation suggère que le complexe pourrait agir comme chaperonne ciblant directement l'assemblage des sous-unités polymérases, plus spécifiquement des sous-unités catalytiques, via POLR2E (Cloutier et Coulombe, 2010). Cette hypothèse a été confirmée dernièrement par les résultats d'une étude publiés durant la rédaction de ce manuscrit (Boulon *et al.*, 2010). Dans cette étude, les auteurs montrent que la fraction cytoplasmique des sous-unités catalytiques des ARN Pol I et II (RPA194 et Rpb1) lie spécifiquement RPAP3 dans des conditions où l'assemblage de ces enzymes est compromis et que la coopération entre RPAP3 et HSP90 est indispensable pour l'activité de stabilisation. De plus, leurs données indiquent que l'ARN Pol II est assemblée dans le cytoplasme et révèlent un mécanisme de contrôle de qualité qui associe HSP90 à l'importation nucléaire d'enzymes entièrement assemblées. Des études concomitantes réalisées dans notre laboratoire montrent que RPB1 et RPB2, les deux plus grandes sous-unités de l'ARN Pol II, sont importées dans le noyau grâce à un mécanisme qui nécessite la fonction de navette de RPAP4/GPN1 et l'assemblage des microtubules, tandis que les autres sous-unités restantes emploieraient des mécanismes alternatifs (Forget *et al.*, 2010). Des études supplémentaires seront nécessaires afin d'établir un consensus quant au mécanisme de biogénèse des ARN polymérases nucléaires.

### 3.3.2. Lien entre la méthylation des gènes et la réparation NER

RPAP4/GPN1 appartient à une famille de protéines « *ATP bind 1* » possédant, comme son nom l'indique, un domaine de liaison à l'ATP, lequel pourrait apporter l'énergie nécessaire aux changements conformationnels impliqués dans la formation des complexes protéiques. De plus, la purification TAP de RPAP4 montre qu'en plus de l'ARN Pol II, il est associé avec un nombre de facteurs impliqués dans le repliement ou l'assemblage des protéines, comme des préfoldines et certaines chaperonnes moléculaires. Des études précédentes ont démontré que GPN1 interagit avec XPA et qu'il participe à la régulation de la transcription via son interaction avec MBD2 (Nitta *et al.*, 2000 ; Lembo *et al.*, 2003). L'identification de ces interactions ainsi que l'observation que ces trois protéines appartiennent au même complexe avaient soulevé la possibilité d'un lien entre le système de méthylation de l'ADN et celui de réparation NER. Nos résultats de purification pour les trois protéines mentionnées n'ont pas permis de soutenir l'hypothèse sur le rôle de RPAP4/GPN1 dans ce lien, du moins pas au niveau de la fraction soluble.

Par ailleurs, l'intervention finement régulée du système ubiquitine-protéasome dans la reconnaissance des dommages à l'ADN a été bien documentée à ce jour. Le facteur UV-DDB (*UV-damaged-DNA-binding protein*) représente un hétérodimère formé de DDB1 et DDB2 qui se lie fortement et spécifiquement aux lésions de l'ADN causées par les irradiations UV (Keeney *et al.*, 1993). Il agit de concert avec le complexe XPC/HR23B dans la détection des lésions durant la réparation globale du génome. En même temps, UV-DDB fait partie d'un complexe ubiquitine E3 ligase contenant la culline 4A, dont les substrats sont DDB2, XPC et la protéine p21 (Liu *et al.*, 2009). Ce complexe intervient également dans l'ubiquitination des histones H2A autour de la lésion, ce qui pourrait déstabiliser la structure du nucléosome et faciliter l'accès des facteurs NER au site endommagé. Par exemple, il a été montré que l'ubiquitination par le complexe UV-DDB autour des sites endommagés dans les nucléosomes favorise le recrutement de XPA (Takedachi *et al.*, 2010). Cependant, l'ubiquitination des histones par ce complexe n'est pas suffisante pour déstabiliser les nucléosomes. Dans la réponse aux cassures double-brin de l'ADN, la cascade UBC13–RNF8–RNF168 serait impliquée

dans l'ubiquitination des histones H2A et H2AX et la régulation de cette réponse serait coordonnée par la sous-unité CHD4 du complexe NuRD (Smeenk *et al.*, 2010; Larsen *et al.*, 2010).

Dans le réseau d'interactions de la **Figure 4-22**, le seul lien constaté entre les modules purifiés MBD2 et XPA passe par le biais de CAND1, un régulateur négatif de l'activité des E3 culline-RING ubiquitine ligases. La purification subséquente de CAND1 montre une interaction bidirectionnelle CAND1–XPA (scores IR de 1.0 et 0.9), ayant CHD3 comme interacteur commun (scores IR de 0.9 et 0.8). Par contre, le complexe NuRD purifié à partir de MBD2 étiqueté ne contient pas les sous-unités CHD3 et CHD4. Cette composition atypique du complexe reste à vérifier, étant donné que la purification du complexe entier à partir de la phase soluble a été confirmée par le passé (Xue *et al.*, 1998; Zhang *et al.*, 1999; Spruijt *et al.*, 2010). On peut aussi mentionner la présence de UBE2N/UBC13 dans la composition du module MBD2-TAP, malgré un faible score (0.07).

Ces résultats préliminaires supportent la possibilité d'un lien entre le complexe de remodelage de la chromatine NuRD et la machinerie de réparation NER via CHD3 et/ou CHD4 et le système ubiquitine-protéasome. Pour en apprendre d'avantage sur le rôle de NuRD, en tant que régulateur de la réponse aux cassures double-brin de l'ADN, il faudra d'abord confirmer les interactions de CHD3 et CHD4 avec MBD2, CAND1 et XPA dans la fraction soluble. Ensuite, on peut envisager l'analyse de ces complexes en réponse à l'irradiation UV et finalement procéder à des analyses au niveau de la chromatine.

### 3.3.3. Rôle de la Huntingtine dans le maintien de l'intégrité génétique

Dans le module de purification XPA-TAP, en plus des facteurs de réparation de l'ADN comme RPA, RFC, Ku (XRCC5-XRCC6), PRKDC/XRCC7 et PARP1, nous avons identifié la protéine Huntingtine (Htt).

La Htt est une protéine ubiquitaire indispensable au développement embryonnaire, à l'hématopoïèse, à la neurogénèse et à la survie neuronale à l'âge adulte. La protéine normale (3144 aa) est complètement soluble, localisée principalement au cytoplasme, associée au cytosquelette et à diverses structures cellulaires (vésicules à clathrine, endosomes, vésicules synaptiques) (Cattaneo *et al.*, 2005). L'expansion anormale du nombre de répétition CAG (36 à 180 répétitions) dans l'exon 1 du gène *IT15* codant pour la Huntingtine est à l'origine de la maladie de Huntington (HD), une maladie neurodégénérative incurable caractérisée par des troubles cognitifs et des mouvements involontaires. Cette mutation se traduit en une longue chaîne polyglutamine (polyQ) dans la région amino-terminale. Les changements de conformation induits par l'extension polyQ rendent la protéine sensible à un clivage protéolytique conduisant à la formation de fragments amino-terminaux qui s'accumulent sous forme d'agrégats. Chez les individus affectés les dommages sont restreints à certaines régions cérébrales, le striatum en particulier. Diverses fonctions cellulaires, y compris la transcription, l'épissage, l'activité protéolytique, l'autophagie, le transport axonal, la transmission synaptique, l'endocytose, la stabilité de la membrane mitochondriale, le métabolisme énergétique et la réponse au stress oxydatif sont affectées dans les neurones exprimant la protéine mutée Htt-polyQ (revu dans (Cattaneo *et al.*, 2005; Zuccato *et al.*, 2010)). Ainsi, il a été suggéré que la pathologie de HD soit liée à une perte de fonction de la Htt normale et à une augmentation de la toxicité liée aux extensions polyQ, qui pourraient diminuer la fonction de certaines protéines.

La présence de la Htt dans le module de purification de XPA (score IR de 0.9) nous permet de soutenir l'hypothèse sur la possibilité d'un certain rôle de cette protéine dans le mécanisme de réparation de l'ADN. Des expériences de double hybride chez la levure ont montré que la Htt normale (20Q) peut interagir avec les protéines

Ku70/XRCC6 et CHD3 via le fragment N-terminal (aa 1 à 506) (Goehler *et al.*, 2004). Une étude plus récente rapporte que l'expression de la Htt mutée (111Q) dans les neurones provoque des cassures double-brins de l'ADN génomique. En effet, il s'agirait d'une accumulation de cassures due à un défaut de réparations à cause de l'interaction entre la Htt muté et Ku70 (Enokido *et al.*, 2010). Ku70 et Ku80 sont des sous-unités régulatrices de la kinase PRKDC (protein kinase, DNA-activated, catalytic polypeptide) avec laquelle ils forment un complexe ternaire impliqué dans la voie NHEJ (*non-homologous end-joining*) de réparation des cassures double-brin de l'ADN (Sancar et Reardon, 2004). Ku70 et Ku80 sont présents dans notre module avec des scores de 1.0 et 0.9 respectivement, tandis que PRKDC a un score de seulement 0.07 (**Tableau 6** en annexes). Ceci suggère que dans la fraction soluble des cellules ECR-293 au phénotype neuronal les deux protéines seraient liées à la Htt normale<sup>1</sup>. Cette interaction pourrait servir à protéger Ku70 et Ku80 contre la dégradation par le protéasome entre les événements de réparation d'ADN. Au contraire, dans les cellules affectées par HD l'affinité entre Ku70 et la Htt mutée serait plus forte, à cause des changements de conformations induits par l'extension polyQ. Ainsi, une partie du complexe Ku70-Htt est dégradée pendant que l'autre forme des agrégats intracellulaires (Enokido *et al.*, 2010). Ceci dit, il serait intéressant de confirmer cette interaction par des purifications TAP de Ku70 et Ku80.

---

<sup>1</sup> La Htt que nous avons copurifié correspond en théorie à la protéine normale, étant donné que la forme non mutée est totalement soluble et qu'elle a été identifiée dans la portion supérieure du gel (Figure 4-15B).

### 3.3.4. PDCD5 est un substrat physiologique de la chaperonine TRiC/CCT

Tel que mentionné, PDCD5 est un régulateur important de l'apoptose et de la paraptose et sa fonction est étroitement reliée à sa structure. Avec la stratégie de purification TAP nous avons déterminé que PDCD5 interagit avec le complexe chaperonine TRiC/CCT dans la fraction soluble des cellules humaines ECR-293. Ce résultat suggère que TRiC/CCT serait responsable du repliement adéquat de PDCD5 et que cette étape serait essentielle dans le processus de mort cellulaire programmée induite par cette protéine. Le fait que 40 % de l'interactome de TRiC/CCT comprend des gènes essentiels (Yam *et al.*, 2008) reflète une voie de repliement qui doit éviter les interactions aberrantes entre les domaines hydrophobes. De plus, dans un communiqué récent il a été montré que PhLP1 (phosducin-like protein 1), une co-chaperonne de TRiC/CCT, pouvait inhiber l'interaction entre PDCD5 et la chaperonine (Gray *et al.*, 2010). Ceci nous conduit à nous intéresser ultérieurement à la cinétique de l'association. Est-ce que PDCD5 fait uniquement transiter la chaperonine durant sa biogénèse ou les CCT servent de réservoir pour le stabiliser contre l'agrégation ou la dégradation ? Est-ce que la libération de PDCD5 nécessiterait un autre niveau de régulation ? Des études ultérieures seront nécessaires afin de répondre à ces questions.

#### 4. CONCLUSION ET PERSPECTIVES

L'analyse de complexes protéiques et de réseaux d'interactions protéine-protéine ainsi que de leur dynamique de formation en fonction du temps et de l'état cellulaire est d'une importance capitale en recherche biologique. Notre laboratoire s'est intéressé au réseau d'interaction de la machinerie de transcription des ARN polymérases nucléaires dans le but de mieux comprendre leurs mécanismes de régulation.

L'objectif de mon projet de maîtrise était de purifier le complexe de l'ARN Pol I ainsi que de poursuivre l'expansion du réseau d'interactions protéine-protéine de la machinerie de transcription de l'ARN Pol II humaine. Pour ce faire, la méthode de purification TAP couplée à la spectrométrie de masse, adaptée pour permettre la purification de complexes protéiques solubles à partir de cellules humaines, a été utilisée.

D'abord, la purification de sous-unités spécifiques de l'ARN Pol I (POLR1E et TWISTNB) nous a permis de mettre en évidence un réseau d'interactions protéine-protéine impliquées dans la transcription de l'ARNr, dont le complexe de l'ARN Pol I et la protéine Treacle. Ainsi, il serait intéressant de vérifier le rôle précis de Treacle dans la régulation de l'activité de l'ARN Pol I et plus particulièrement dans la transcription de l'ADNr à partir d'un promoteur spécifique.

Par la suite, ce travail a permis de confirmer la bidirectionnalité des interactions entre POLR2E et le complexe RPAP3/R2TP/*prefoldin-like* ainsi que celle entre PDCD5 et le complexe TRiC/CCT. Même si le rôle du complexe RPAP3/R2TP/*prefoldin-like* dans l'assemblage de l'ARN Pol II a déjà été démontré, il serait intéressant de connaître quels sont ses autres substrats et cofacteurs, ainsi que la contribution de chaque sous-unité en particulier dans ce processus. Quant à PDCD5, le fait que son activité soit étroitement reliée à sa structure souligne l'importance de son interaction avec le complexe chaperonine. Étant donné son implication dans divers types de tumeurs, la découverte de tout nouveau mécanisme de régulation de PDCD5 semble extrêmement pertinente.

Parallèlement, nous avons montré que RPAP4/GPN1 ne faisait partie ni du réseau d'interactions de XPA, ni de celui de MBD2, ce qui est en accord avec nos résultats sur son rôle dans la biogénèse de l'ARN Pol II. Cependant, ensemble, les résultats de purification de MBD2, XPA et CAND1 nous permettent de soutenir l'importance du système ubiquitine-protéasome dans la régulation des processus de réparation de l'ADN et d'entrevoir un certain lien entre la méthylation des gènes et la réparation NER. Par contre, pour démontrer ce lien il faudra envisager une autre approche de purification permettant l'analyse des complexes assemblés sur la chromatine, plutôt que dans la fraction soluble. De plus, l'analyse du réseau d'interactions de XPA nous a amené à proposer un rôle pour la Huntingtine normale dans le maintien de l'intégrité génétique. Cette hypothèse pourrait être facilement vérifiée par des purifications TAP et sa confirmation éventuelle apporterait des progrès significatifs dans la compréhension des mécanismes de la maladie de Huntington.

Outre le fait que ce type d'analyse nous a permis d'identifier et de caractériser de nouveaux régulateurs de la machinerie de transcription, il reste particulièrement intéressant de suivre ses applications subséquentes en biologie clinique comme la recherche de nouvelles cibles thérapeutiques ou encore de marqueurs diagnostiques.

## RÉFÉRENCES

- Anelli, T. and Sitia, R. (2008). Protein quality control in the early secretory pathway. *EMBO J.* 27, 315-327.
- Angrisano, T., Lembo, F., Pero, R., Natale, F., Fusco, A., Avvedimento, V.E., Bruni, C.B. et Chiariotti, L. (2006). TACC3 mediates the association of MBD2 with histone acetyltransferases and relieves transcriptional repression of methylated promoters. *Nucleic Acids Res* 34, 364-72.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. et Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 25, 25-29.
- Awrey, D.E., Weilbaecher, R.G., Hemming, S.A., Orlicky, S.M., Kane, C.M. et Edwards, A.M. (1997). Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J Biol Chem.* 272, 14747-14754.
- Baumann, M., Pontiller, J. et Ernst, W. (2010). Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview. *Mol. Biotechnol.* 45, 241-247.
- Beckouet, F., Labarre-Mariotte, S., Albert, B., Imazawa, Y., Werner, M., Gadal, O., Nogi, Y. et Thuriaux, P. (2008). Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. *Mol Cell Biol.* 28, 1596-1605
- Birch, J.L., Tan, B.C., Panov, K.I., Panova, T.B., Andersen, J.S., Owen-Hughes, T.A., Russell, J., Lee, S.C. et Zomerdijk, J.C. (2009) FACT facilitates chromatin transcription by RNA polymerases I and III. *EMBO J.* 28, 854-865.

- Bischler, N., Brino, L., Carles, C., Riva, M., Tschochner, H., Mallouh, V. et Schultz, P. (2002). Localization of the yeast RNA polymerase I-specific subunits. *EMBO J.* 21, 4136-4144.
- Bosu, D.R. et Kipreos, E.T. (2008). Cullin-RING ubiquitin ligases: global regulation and activation cycles. *Cell Div.* 3:7.
- Boulon, S., Pradet-Balade, B., Verheggen, C., Molle, D., Boireau, S., Georgieva, M., Azzag, K., Robert, M.C., Ahmad, Y., Neel, H., Lamond, A.I. et Bertrand, E. (2010). HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. *Mol Cell.* 39, 912-24.
- Bradsher J, Auriol J, Proietti de Santis L, Iben S, Vonesch JL, Grummt I, Egly JM. (2002). CSB is a component of RNA pol I transcription. *Mol. Cell* 10, 819-829.
- Buratowski, S., Hahn, S., Guarente, L. et Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 56, 549-561.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* 38, 626-635.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* 38, 626-635.
- Cattaneo, E., Zuccato, C. et Tartari, M. (2005). Normal huntingtin function: an alternative approach to Huntington's disease. *Nat Rev Neurosci.* 6, 919-30.
- Chédin, S., Riva, M., Schultz, P., Sentenac, A. et Carles, C. (1998). The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes Dev.* 12, 3857-3871.

- Chen, C., Zhou, H., Xu, L., Xu, D., Wang, Y., Zhang, Y., Liu, X., Liu, Z., Ma, D., Ma, Q. et Chen, Y. (2010). Recombinant human PDCD5 sensitizes chondrosarcomas to cisplatin chemotherapy in vitro and in vivo. *Apoptosis*. 15, 805-13.
- Chen, Y., Sun, R., Han, W., Zhang, Y., Song, Q., Di, C. et Ma, D. (2001). Nuclear translocation of PDCD5 (TFAR19): an early signal for apoptosis? *FEBS Lett*. 509, 191–196.
- Cheng, B. et Price, D.H. (2007). Properties of RNA Polymerase II Elongation Complexes Before and After the P-TEFb-mediated Transition into Productive Elongation. *J. Biol. Chem*. 282, 21901-21912.
- Choudhary, C., Kumar, C., Gnad, F., Nielsen M.L., Rehman, M., Walther, T.C., Olsen, J.V. et Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834-840.
- Cloutier, P. et Coulombe, B. (2010) New insights into the biogenesis of nuclear RNA polymerases? *Biochem Cell Biol*. 88, 211-21.
- Cloutier, P., Al-Khoury, R., Lavallée-Adam, M., Faubert, D., Jiang, H., Poitras, C., Bouchard, A., Forget, D., Blanchette, M. et Coulombe, B. (2009). High-resolution mapping of the protein interaction network for the human transcription machinery and affinity purification of RNA polymerase II-associated complexes. *Methods* 48, 381-386.
- Coulombe, B. et Burton, Z.F. (1999). DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. *Microbiol. Mol. Biol. Rev*. 63, 457-478.
- Coulombe, B., Jeronimo, C., Langelier, M.F., Cojocar, M., et Bergeron, D. (2004). Interaction networks of the molecular machines that decode, replicate and maintain the integrity of the human genome. *Mol. Cell Proteomics* 3, 851-856.
- Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., Jawhari, A., Jennebach, S., Kamenski, T.,

- Kettenberger, H., Kuhn, C.D., Lehmann, E., Leike, K., Sydow, J.F. et Vannini, A. (2008). Structure of eukaryotic RNA polymerases. *Annu Rev Biophys.* 37, 337-52.
- Cramer, P., Bushnell, D.A. et Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292, 1863-76.
- Delgermaa, L., Hayashi, N., Dorjsuren, D., Nomura, T., Thuy, T.T., et Murakami, S. (2004). Subcellular localization of RPB5-mediating protein and its putative functional partner. *Mol. Cell. Biol.* 24, 8556–8566.
- Denissov, S., van Driel, M., Voit, R., Hekkelman, M., Hulsen, T., Hernandez, N., Grummt, I., Wehrens, R. et Stunnenberg, H. (2007). Identification of novel functional TBP-binding sites and general factor repertoires. *EMBO J.* 26, 944–954
- Dixon, J., Jones, N.C., Sandell, L.L., Jayasinghe, S.M., Crane, J., Rey, J.P., Dixon, M.J. et Trainor, P.A. (2006). Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13403–13408.
- Dorjsuren, D., Lin, Y., Wei, W., Yamashita, T., Nomura, T., Hayashi, N., et Murakami, S. (1998). RMP, a novel RNA polymerase II subunit 5-interacting protein, counteracts transactivation by hepatitis B virus X protein. *Mol. Cell. Biol.* 18, 7546–7555.
- Drygin, D., Rice, W.G. et Grummt, I. (2010). The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu. Rev. Pharmacol. Toxicol.* 50, 131-156.
- Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L.I., Phair, R.D. et Misteli, T. (2002). A kinetic framework for a mammalian RNA polymerase in vivo. *Science* 298, 1623-1626
- Ego, T., Tanaka, Y. et Shimotohno, K. (2005). Interaction of HTLV-1 Tax and methyl-CpG-binding domain 2 positively regulates the gene expression from the hypermethylated LTR. *Oncogene.* 24, 1914-23.

- Ellis, R.J., Van der Vies, S.M. et Hemmingsen, S.M. (1989). The molecular chaperone concept. *Biochem. Soc. Symp.* 55, 145-153.
- Enokido, Y., Tamura, T., Ito, H., Arumughan, A., Komuro, A., Shiwaku, H., Sone, M., Foulle, R., Sawada, H., Ishiguro, H., Ono, T., Murata, M., Kanazawa, I., Tomilin, N., Tagawa, K., Wanker, E.E. et Okazawa, H. (2010). Mutant huntingtin impairs Ku70-mediated DNA repair. *J Cell Biol.* 189, 425-43.
- Forget, D., Lacombe, A.A., Cloutier, P., Al-Khoury, R., Bouchard, A., Lavallée-Adam, M., Faubert, D., Jeronimo, C., Blanchette, M. et Coulombe, B. The protein interaction network of the human transcription machinery reveals a role for the conserved GTPase RPAP4/GPN1 and microtubule assembly in nuclear import and biogenesis of RNA polymerase II. *Mol Cell Proteomics.* 9, 2827-39.
- Forget, D., Robert, F., Grondin, G., Burton, Z.F., Greenblatt, J. et Coulombe, B. (1997). RAP74 induces promoter contacts by RNA polymerase II upstream and downstream of a DNA bend centered on the TATA box. *Proc. Natl. Acad. Sci. U.S.A* 94, 7150-7155.
- Fuda, N.J., Ardehali, M.B., Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 461, 186-192.
- Fujita, H., Fujii, R., Aratani, S., Amano, T., Fukamizu, A. et Nakajima, T. (2003). Antithetic effects of MBD2a on gene regulation. *Mol Cell Biol.* 23, 2645-57.
- Gadal, O., Mariotte-Labarre, S., Chedin, S., Quemeneur, E., Carles, C., Sentenac, A. et Thuriaux, P. (1997). A34.5, a nonessential component of yeast RNA polymerase I, cooperates with subunit A14 and DNA topoisomerase I to produce a functional rRNA synthesis machine. *Mol Cell Biol.* 17, 1787-95.
- Gingras, A.C., Aebersold, R. et Raught B. (2005). Advances in protein complex analysis using mass spectrometry. *J. Physiol.* 563, 11-21.
- Gingras, A.C., Stager, S., Raught, B. et Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* 8, 645- 54.

- Goehler, H., Lalowski, M., Stelzl, U., Waelter, S., Stroedicke, M., Worm, U., Droege, A., Lindenberg, K.S., Knoblich, M., Haenig, C., Herbst, M., Suopanki, J., Scherzinger, E., Abraham, C., Bauer, B., Hasenbank, R., Fritzsche, A., Ludewig, A.H., Büssow, K., Coleman, S.H., Gutekunst, C.A., Landwehrmeyer, B.G., Lehrach, H. et Wanker, E.E. (2004). A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol Cell*. 15, 853– 65.
- Goldenberg, S.J., Cascio, T.C., Shumway, S.D., Garbutt, K.C., Liu, J., Xiong, Y., et Zheng, N. (2004). Structure of the Cand1-Cull1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. *Cell* 119, 517-528.
- Gonzales, B., Henning, D., So, R.B., Dixon, J., Dixon, M.J. et Valdez, B.C. (2005). The Treacher Collins syndrome (TCOF1) gene product is involved in pre-rRNA methylation. *Hum Mol Genet*. 14, 2035-43.
- Gorski, J.J., Pathak, S., Panov, K., Kasciukovic, T., Panova, T., Russell, J. et Zomerdijk, J.C. (2007). A novel TBP-associated factor of SL1 functions in RNA polymerase I transcription. *EMBO J*. 26, 1560-1568.
- Gray, A., Howlett, A., Prince, J., Ahn, N., Willardson, B. (2010). Novel phospho-ucin-like protein 1 interactions. *FASEB J*. 24, 834.2.
- Greenblatt, J. (1997). RNA polymerase II holoenzyme and transcriptional regulation. *Curr. Opin. Cell Biol*. 9, 310-319.
- Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C. et Gelbart, W. M. (2002). *Introduction à l'analyse génétique (3e éd.)* Éditions De Boeck Université, Bruxelles, 860 p.
- Grummt, I.(2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev*. 17, 1691-1702.

- Gstaiger, M., Luke, B., Hess, D., Oakeley, E.J., Wirbelauer, C., Blondel, M., Vigneron, M., Peter, M. et Krek, W. (2003). Control of nutrient-sensitive transcription programs by the unconventional prefoldin URI. *Science* 302, 1208-12.
- Hanada, K., Song, C.Z., Yamamoto, K., Yano, K., Maeda, Y., Yamaguchi, K. et Muramatsu M. (1996). RNA polymerase I associated factor 53 binds to the nucleolar transcription factor UBF and functions in specific rDNA transcription. *EMBO J.* 15, 2217–2226.
- Hendrich, B. et Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol.* 18, 6538-47.
- Hernandez-Verdun, D. et Louvet, E. (2004). Le nucléole: structure, fonctions et maladies associées. *Médecine Sciences* 20, 37-44.
- Huber, L.A. (2003). Is proteomics heading in the wrong direction? *Nat. Rev. Mol. Cell Biol.* 4, 74-80
- Hull, M.W., McKune, K. et Woychik, N.A. (1995). RNA polymerase II subunit RPB9 is required for accurate start site selection. *Genes Dev.* 9, 481-490.
- Iben S, Tschochner H, Bier M, Hoogstraten D, Hozák P, Egly JM, Grummt I. (2002). TFIIF plays an essential role in RNA polymerase I transcription. *Cell* 109, 297-306.
- Jansa, P. et Grummt, I. (1999). Mechanism of transcription termination: PTRF interacts with the largest subunit of RNA polymerase I and dissociates paused transcription complexes from yeast and mouse. *Mol. Gen. Genet.* 262, 508-514.
- Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J., Chabot, B., Poirier, G.G., Hughes, T.R., Blanchette, M., Price, D.H. et Coulombe, B. (2007). Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol Cell* 27, 262-274.
- Jeronimo, C., Langelier, M.F., Zeghouf, M., Cojocaru, M., Bergeron, D., Baali, D., Forget, D., Mnaimneh, S., Davierwala, A.P., Pootoolal, J., Chandy, M., Canadien,

- V., Beattie, B.K., Richards, D.P., Workman, J.L., Hughes, T.R., Greenblatt, J. et Coulombe, B. (2004). RPAP1, a novel human RNA polymerase II-associated protein affinity purified with recombinant wild-type and mutated polymerase subunits. *Mol. Cell Biol.* 24, 7043- 7058.
- Juven-Gershon, T., Hsu, J.Y., Theisen, J.W. et Kadonaga, J.T. (2008). The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol.* 20, 253-259.
- Keeney, S., Chang, G.J. et Linn, S. (1993). Characterization of a human DNA damage binding protein implicated in xeroderma pigmentosum E. *J Biol Chem.* 268, 21293-300.
- Kettenberger, H., Armache, K.J. et Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* 114, 347–357.
- Khorasanizadeh, S. (2004). The nucleosome: From genomic organization to genomic regulation. *Cell* 116, 259- 272.
- Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H. et Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77, 599-608.
- Kosan, C. et Kunz, J. (2002). Identification and characterisation of the gene TWIST NEIGHBOR (TWISTNB) located in the microdeletion syndrome 7p21 region. *Cytogenet Genome Res.* 97, 167-70.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.
- Krueger, B.J., Jeronimo, C., Roy, B.B., Bouchard, A., Barrandon, C., Byers, S.A., Searcey, C.E., Cooper, J.J., Bensaude, O., Cohen, E.A., Coulombe, B. et Price, D.H. (2008). LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. *Nucleic Acids Res.* 36, 2219-2229

- Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R. et Cramer, P. (2007). Functional architecture of RNA polymerase I. *Cell* 131, 1260-72.
- Kwapisz, M., Beckouët, F. et Thuriaux, P. (2008). Early evolution of eukaryotic DNA-dependent RNA polymerases. *Trends Genet.* 24, 211-215.
- Langelier, M.F., Trinh, V., et Coulombe, B. (2002). Gros plan sur l'ARN polymérase II. *Médecine Sciences* 18, 211-216.
- Larsen, D.H., Poinsignon, C., Gudjonsson, T., Dinant, C., Payne, M.R., Hari, F.J., Danielsen, J.M., Menard, P., Sand, J.C., Stucki, M., Lukas, C., Bartek, J., Andersen, J.S. et Lukas, J. (2010). The chromatin-remodeling factor CHD4 coordinates signaling and repair after DNA damage. *J Cell Biol.* 190, 731-40.
- Lavallée-Adam, M., Cloutier, P., Coulombe, B. et Blanchette, M. (2010). Modeling Contaminants in AP-MS/MS Experiments. *J Proteome Res.*
- Le, T.T., Zhang, S., Hayashi, N., Yasukawa, M., Delgermaa, L. et Murakami, S. (2005). Mutational analysis of human RNA polymerase II subunit 5 (RPB5): the residues critical for interactions with TFIIF subunit RAP30 and hepatitis B virus X protein. *J Biochem.* 138, 215-24.
- Lee, D., Gershenzon, N., Gupta, M., Ioshikhes, I., Reinberg, D. et Lewis, B. (2005). Functional characterization of core promoter elements: The downstream core element is recognized by TAF1. *Molecular and Cellular Biology* 25, 9674–9686.
- Lembo, F., Pero, R., Angrisano, T., Vitiello, C., Iuliano, R., Bruni, C. B. et Chiariotti, L.(2003). MBDin, a novel MBD2-interacting protein, relieves MBD2 repression potential and reactivates transcription from methylated promoters. *Mol. Cell. Biol.* 23, 1656–1665.
- Lescuyer, P., Chevallet, M. et Rabilloud, T. (2004). L'analyse protéomique : concepts, réalités et perspectives en thérapeutique. *Médecine Sciences* 20, 587-592.

- Liljelund, P., Mariotte, S., Buhler, J.M. et Sentenac, A. (1992). Characterization and mutagenesis of the gene encoding the A49 subunit of RNA polymerase A in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 89, 9302-5.
- Lin, C.I. et Yeh, N.H. (2009). Treacle recruits RNA polymerase I complex to the nucleolus that is independent of UBF. *Biochem. Biophys. Res. Commun.* 386, 396-401.
- Liu, D., Yao, H., Chen, Y., Feng, Y., Chen, Y. et Wang, J. (2005). The N-terminal 26-residue fragment of human programmed cell death 5 protein can form a stable alpha-helix having unique electrostatic potential character. *Biochem. J.* 392, 47-54.
- Liu, H., Wang, Y., Zhang, Y., Song, Q., Di, C., Chen, G., Tang, J. et Ma, D. (1999). TFAR19, a novel apoptosis-related gene cloned from human leukemia cell line TF-1, could enhance apoptosis of some tumor cells induced by growth factor withdrawal. *Biochem Biophys Res Commun.* 254, 203-10.
- Liu, L., Lee, S., Zhang, J., Peters, S.B., Hannah, J., Zhang, Y., Yin, Y., Koff, A., Ma, L. et Zhou, P. (2009). CUL4A abrogation augments DNA damage response and protection against skin carcinogenesis. *Mol Cell.* 34, 451-60.
- Lodish, H.F. (2005). *Biologie moléculaire de la cellule* (3e éd.). Éditions De Boeck Université, Bruxelles, 973 p.
- Luo, X., Huang, Y., et Sheikh, M.S. (2003). Cloning and characterization of a novel gene PDRG that is differentially regulated by p53 and ultraviolet radiation. *Oncogene*, 22, 7247–7257.
- Luvsanjav, D., Hayashi, N., Dorjsuren, D., Nomura, T., Le, T.T.- T., and Murakami, S. (2004). Subcellular localization of RPB5-mediating protein and its putative functional partner. *Mol.Cell Biol.* 24, 8556–8566.
- Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J. et Reinberg, D. (2004). "Functional interactions of RNA-capping enzyme with factors that positively and

- negatively regulate promoter escape by RNA polymerase II." *Proc. Natl. Acad. Sci. U.S.A.* 101, 7572-7577.
- Mantovani, R. (1999). The molecular biology of the CCAAT-binding factor NF-Y. *Gene* 239, 15-27.
- Masson, C., Menea, F., Pinon-Lataillade, G., Frobert, Y., Chevillard, S., Radicella, J.P., Sarasin, A. et Angulo, J.F. (2003). Global genome repair is required to activate KIN17, a UVC-responsive gene involved in DNA replication. *Proc Natl Acad Sci U S A.* 100, 616-21.
- Maston, G.A., Evans, S.K. et Green, M.R. (2006). Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* 7, 29-59.
- Melnikov, E.E. et Rotanova, T.V. (2010). Molecular chaperones. *Russian J. of Bioorg. Chem.* 36, 1–10.
- Melville, M.W., McClellan, A.J., Meyer, A.S., Darveau, A. et Frydman, J. (2003). The Hsp70 and TRiC/CCT chaperone systems cooperate in vivo to assemble the von Hippel-Lindau tumor suppressor complex. *Mol. Cell Biol.* 23, 3141-51.
- Miller, G., Panov, K.I., Friedrich, J.K., Trinkle-Mulcahy, L., Lamond, A.I. et Zomerdijk, J.C. (2001). hRRN3 is essential in the SL1-mediated recruitment of RNA Polymerase I to rRNA gene promoters. *EMBO J.* 20, 1373–1382.
- Miller, O. L. Jr. et Beatty, B. R. (1969). Visualization of nucleolar genes. *Science* 164, 955-957.
- Moss, T., Langlois, F., Gagnon-Kugler, T. et Stefanovsky, V. (2007). A housekeeper with power of attorney: the rRNA genes in ribosomal biogenesis. *Cell. Mol. Life Sci.* 64, 29-49.
- Muratani, M. et Tansey, W.P. (2003). How the ubiquitin-proteasome system controls transcription. *Nat Rev Mol Cell Biol.* 4, 192-201.

- Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D. et Bird, A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet.* 23, 58-61.
- Nitta, M., Saijo, M., Kodo, N., Matsuda, T., Nakatsu, Y., Tamai, H. et Tanaka, K.(2000). A novel cytoplasmic GTPase XAB1 interacts with DNA repair protein XPA. *Nucleic Acids Res.* 28, 4212–4218.
- No, D., Yao, T.P. et Evans, R.M. (1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.* 93, 3346-3351.
- Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., Brunak, S. et Mann, M. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* 3, ra3.
- Panov, K.I., Panova, T.B., Gadal, O., Nishiyama, K., Saito, T., Russell, J. et Zomerdijk, J.C. (2006). RNA polymerase I-specific subunit CAST/hPAF49 has a role in the activation of transcription by upstream binding factor. *Mol. Cell Biol.* 26, 5436-48.
- Perkins, D.N., Pappin, D.J., Creasy, D.M. et Cottrell J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-3567.
- Petsko, G.A. et Ringe, D. (2009). *Structure et fonction des protéines (1<sup>e</sup> éd.)*. Éditions De Boeck Université, Bruxelles, 190 p.
- Peyroche, G., Levillain, E., Siaut, M., Callebaut, I., Schultz, P., Sentenac, A., Riva, M. et Carles, C. (2002). The A14-A43 heterodimer subunit in yeast RNA pol I and their relationship to Rpb4-Rpb7 pol II subunits. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14670-14675.
- Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C. et Riva, M. (2000). The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J* 19, 5473-5482.

- Porquet, D. (2005). Analyse protéomique : stratégies méthodologiques et applications en biologie clinique. *Spectra Biol.* 145, 55-59.
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M. et Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 24, 218–229.
- Raska, I., Koberna, K., Malínský, J., Fidlerová, H. et Masata, M. (2004). The nucleolus and transcription of ribosomal genes. *Biol. Cell* 96, 579-594.
- Richard, P. et Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. *Genes Dev.* 23, 1247-1269.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., et Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030-1032.
- Robert, F., Douziech, M., Forget, D., Egly, J.M., Greenblatt, J., Burton, Z.F. et Coulombe, B. (1998). Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. *Mol. Cell* 2, 341-351.
- Rose, K.M., Szopa, J., Han, F.S., Cheng, Y.C., Richter, A. et Scheer, U. (1988). Association of DNA topoisomerase I and RNA polymerase I: a possible role for topoisomerase I in ribosomal gene transcription. *Chromosoma* 96, 411–416.
- Russell, J. et Zomerdijs, J.C. (2005). RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem. Sci.* 30, 87-96.
- Sancar, A. et Reardon, J.T. Nucleotide excision repair in *E. coli* and man. *Adv Protein Chem.* 69, 43-71.
- Sansom, O.J., Maddison, K. et Clarke, AR. (2007). Mechanisms of Disease: methyl-binding domain proteins as potential therapeutic targets in cancer. *Nat. Clin. Prac. Onc.* 4, 305-315.

- Sardiu, M.E., Cai, Y., Jin, J., Swanson, S.K., Conaway, R.C., Conaway, J.W., Florens, L. et Washburn, M.P. (2008). Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc Natl Acad Sci U S A.* 105, 1454-9.
- Schneider, D.A., French, S.L., Osheim, Y.N., Bailey, A.O., Vu, L., Dodd, J., Yates, J.R., Beyer, A.L. et Nomura, M. (2006). RNA polymerase II elongation factors Spt4p and Spt5p play roles in transcription elongation by RNA polymerase I and rRNA processing. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12707-12.
- Sekimata, M. et Homma, Y. (2004). Sequence-specific transcriptional repression by an MBD2-interacting zinc finger protein MIZF. *Nucleic Acids Res.* 32, 590-7.
- Shilatifard, A., Conaway, R.C. et Conaway JW. (2003). The RNA polymerase II elongation complex. *Annu. Rev. Biochem.* 72, 693-715.
- Siegers, K., Bölter, B., Schwarz, J.P., Böttcher, U.M., Guha, S. et Hartl, F.U. (2003). TRiC/CCT cooperates with different upstream chaperones in the folding of distinct protein classes. *EMBO J.* 22, 5230-5240.
- Siegert, R., Leroux, M.R., Scheufler, C., Hartl, F.U. et Moarefi, I. (2000). Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell* 103, 621-32.
- Sikorski, T.W. et Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Curr. Opin. Cell Biol.* 21, 344-351.
- Sirri, V., Urcuqui-Inchima, S., Roussel, P. et Hernandez-Verdun, D. (2008). Nucleolus: the fascinating nuclear body. *Histochem Cell Biol.* 129, 13-31.
- Smeenk, G., Wiegant, W.W., Vrolijk, H., Solari, A.P., Pastink, A. et van Attikum, H. (2010). The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage. *J Cell Biol.* 190, 741-9.

- Spruijt, C.G., Bartels, S.J., Brinkman, A.B., Tjeertes, J.V., Poser, I., Stunnenberg, H.G. et Vermeulen, M. CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex. *Mol Biosyst.* 6, 1700-6.
- Stefanovsky, V.Y., Langlois, F., Gagnon-Kugler, T., Rothblum, L.I. et Moss, T. (2006). Growth factor signaling regulates elongation of RNA polymerase I transcription in mammals via UBF phosphorylation and chromatin remodeling. *Mol. Cell* 21, 629–639.
- Strahl, B.D. et Allis, C.D. (2000). The language of covalent histone modifications. *Nature* 403, 41-45.
- Takedachi, A., Saijo, M. et Tanaka, K. (2010). DDB2 complex-mediated ubiquitylation around DNA damage is oppositely regulated by XPC and Ku and contributes to the recruitment of XPA. *Mol Cell Biol.* 30, 2708-23.
- Tatematsu, K.I., Yamazaki, T. et Ishikawa, F. (2000). MBD2-MBD3 complex binds to hemi-methylated DNA and forms a complex containing DNMT1 at the replication foci in late S phase. *Genes Cells.* 5, 677-88.
- Thomas, M.C. et Chiang, C.M. (2006). The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol.* 41, 105-178.
- Thompson, C.M., Young, R.A. (1995). General requirement for RNA polymerase II holoenzymes in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 92, 4587–4590.
- Thuriaux, P., Mariotte, S., Buhler, J.M., Sentenac, A., Vu, L., Lee, B.S. et Nomura, M. (1995). Gene RPA43 in *Saccharomyces cerevisiae* encodes an essential subunit of RNA polymerase I. *J Biol Chem.* 270, 24252-7.
- Vainberg, I.E., Lewis, S.A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H.L. et Cowan, N.J. (1998). Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* 93, 863-73.
- Valdez, BC., Henning, D., So, R.B., Dixon, J. et Dixon, M.J. (2004). The Treacher Collins syndrome (TCOF1) gene product is involved in ribosomal DNA gene

- transcription by interacting with upstream binding factor. *Proc Natl Acad Sci U S A*. 101, 10709-14.
- Wang, Y., Li, D., Fan, H., Tian, L., Zhong, Y., Zhang, Y., Yuan, L., Jin, C., Yin, C. et Ma, D. (2006). Cellular uptake of exogenous human PDCD5 protein. *J Biol Chem*. 281, 24803-17.
- Wang, Y., Li, X., Wang, L., Ding, P., Zhang, Y., Han, W. et Ma, D. (2004). An alternative form of paraptosis-like cell death, triggered by TAJ/TROY and enhanced by PDCD5 overexpression. *J. Cell Sci*. 117, 1525–1532.
- Wei, W., Dorjsuren, D., Lin, Y., Qin, W., Nomura, T., Hayashi, N., and Murakami, S. (2001). Direct interaction between the subunit RAP30 of transcription factor IIF (TFIIF) and RNA polymerase subunit 5, which contributes to the association between TFIIF and RNA polymerase II. *J. Biol. Chem*. 276, 12266–12273.
- Werner, M., Hermann-Le Denmat, S., Treich, I., Sentenac, A. et Thuriaux, P. (1992) Effect of mutations in a zinc-binding domain of yeast RNA polymerase C (III) on enzyme function and subunit association. *Mol. Cell Biol*. 12, 1087-1095.
- Werner, M., Thuriaux, P. et Soutourina, J. (2009). Structure–function analysis of RNA polymerases I and III. *Curr. Opin. Struct. Biol*. 19, 740-745.
- Xu, L., Chen, Y., Song, Q., Xu, D., Wang, Y. et Ma, D. (2009). PDCD5 interacts with Tip60 and functions as a cooperator in acetyltransferase activity and DNA damage-induced apoptosis. *Neoplasia* 11, 345–354.
- Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J. et Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell*. 2, 851-61.
- Yam, A.Y., Xia, Y., Lin, H.T., Burlingame, A., Gerstein, M. et Frydman, J. Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nat Struct Mol Biol*. 15, 1255-62.

- Yamamoto, K., Yamamoto, M., Hanada, K., Nogi, Y., Matsuyama, T. et Muramatsu, M. (2004). Multiple protein-protein interactions by RNA polymerase I-associated factor PAF49 and role of PAF49 in rRNA transcription. *Mol Cell Biol.* 24, 6338-49.
- Yogosawa, S., Makino, Y., Yoshida, T., Kishimoto, T., Muramatsu, M. et Tamura, T. (1996). Molecular cloning of a novel 120-kDa TBP-interacting protein. *Biochem Biophys Res Commun.* 229, 612-7.
- Zaros, C., Briand, J.F., Boulard, Y., Labarre-Mariotte, S., Garcia-Lopez, M.C., Thuriaux, P. et Navarro, F. (2007). Functional organization of the Rpb5 subunit shared by the three yeast RNA polymerases. *Nucleic Acids Res.* 35, 634-47.
- Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A. et Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* 13, 1924-35.
- Zhang, Y., Sikes, M.L., Beyer, A.L. et Schneider, D.A. (2009). The Paf1 complex is required for efficient transcription elongation by RNA polymerase I. *Proc. Natl. Acad. Sci. U. S. A.* 106, 2153-8.
- Zhao, H., Wang, Q., Zhang, H., Liu, Q., Du, X., Richter, M., et Greene, M.I. (2005). UXT is a novel centrosomal protein essential for cell viability. *Mol. Biol. Cell* 16, 5857–5865.
- Zuccato, C., Valenza, M. et Cattaneo, E. (2010). Molecular mechanisms and potential therapeutical targets in Huntington's disease. *Physiol Rev.* 90, 905-81.

## ANNEXES

Tableau 1. Liste des protéines partenaires d'interaction co-purifiées avec POLR1E.

Protéine	Score IR	Protéine	Score IR	Protéine	Score IR
ARID3A	1	ARID3B	0.6	KHDRBS1	0.09
CD3EAP	1	SNRPB	0.59	REXO4	0.09
GCN1L1	1	SNRPN	0.59	TUBB2B	0.09
PCBP2	1	RRN3	0.58	KIF11	0.07
POLR1A	1	SMARCC2	0.58	NSUN5	0.07
POLR1B	1	ARF6	0.55	TRIM23	0.07
POLR1C	1	LRRC40	0.53	KLC2	0.06
POLR2H	1	CNOT1	0.49	AMOTL1	0.05
TUBA1A	1	LOC100129958	0.35	DHX40	0.05
TUBB2A	1	RFC1	0.34	DNM1L	0.05
TUBB3	1	XPO5	0.34	GRHPR	0.05
TUBB4	1	IPO5	0.32	JMJD1C	0.05
CAND1	0.99	RPA2	0.32	JMJD7	0.05
KPNA2	0.99	TBPL2	0.27	JMJD7-PLA2G4B	0.05
POLR1D	0.99	PES1	0.25	SRP14	0.05
PHGDH	0.98	BAT2L	0.24	TAF1C	0.05
HSPA6	0.95	SPTBN1	0.23	TBC1D4	0.05
TWISTNB	0.95	PCNT	0.22	WDR87	0.05
POLR2L	0.92	POLR2K	0.22	ZBTB46	0.05
RQCD1	0.9	RPS7	0.2	AGPAT2	0.04
RUVBL2	0.89	RPA1	0.19	ANKRD26	0.04
CCT4	0.87	SMARCC1	0.19	ANXA5	0.04
KPNB1	0.84	SUPT6H	0.19	ARPC3	0.04
POLR2A	0.84	VAV2	0.19	DUSP15	0.04
PARP1	0.77	LOC730210	0.18	EMR1	0.04
POLR3A	0.76	LOC730230	0.18	PAK2	0.04
ZNRD1	0.75	NOLC1	0.18	RXRA	0.04
HSPA5	0.74	RNF213	0.17	VPS28	0.04
SAMD1	0.72	DOCK7	0.15	EEF1A1	0
TCOF1	0.72	hCG_1988300	0.1	TUBA1B	0
RFC4	0.7	HEATR6	0.1	TUBA1C	0
TBP	0.69	HNRNPK	0.1	TUBB	0
TUBB1	0.68	RIF1	0.1	TUBB2C	0
POLR2F	0.63	CSE1L	0.09	XPO1	0

**Tableau 2. Liste des protéines partenaires d'interaction co-purifiées avec TWISTNB.**

Protéine	Score IR	Protéine	Score IR	Protéine	Score IR
CD3EAP	1	KPNA6	0.7	CYP4F11	0.04
KPNA2	1	TBP	0.58	CYP4F12	0.04
POLR1A	1	HNRNPH1	0.45	EEF1G	0.04
POLR1B	1	HNRNPH2	0.45	IQCA1	0.04
POLR1C	1	PCNT	0.22	PIBF1	0.04
POLR1D	1	VAV2	0.22	PMM1	0.04
POLR1E	1	LOC730210	0.19	PRR11	0.04
POLR2E	1	LOC730230	0.19	RP4-691N24.1	0.04
POLR2H	1	TBPL2	0.19	SERPINI2	0.04
RRN3	1	TRIP12	0.19	TRD@	0.04
TCOF1	1	RNF213	0.17	UNC13B	0.04
POLR2L	0.99	ALB	0.08	BAT2D1	0
ZNRD1	0.94	BAT2	0.06	EEF1A1	0
POLR2A	0.86	TAF1B	0.06	ELAVL1	0
TUBB4	0.84	TOP2B	0.06	HSPA1A	0
POLR2F	0.79	HIST1H1T	0.05	HSPA1B	0
POLR3A	0.79	SF3B2	0.05	POLR2K	0
KPNA1	0.7	113425206	0.04		
KPNA5	0.7	119569250	0.04		

**Tableau 3. Liste des protéines partenaires d'interaction co-purifiées avec POLR2E.**

Protéine	Score IR	Protéine	Score IR	Protéine	Score IR
ARID3A	1	EIF4A1	0.96	BRD2	0.19
C19orf2	1	MED24	0.96	CCDC72	0.19
C1orf25	1	HNRNPD	0.95	EXOG	0.19
CAND1	1	MED30	0.95	MED11	0.19
CD3EAP	1	BAT1	0.94	MRPL52	0.19
CKAP5	1	DDX39	0.94	PPP2R5A	0.19
CTDP1	1	DDX5	0.94	SMYD3	0.19
DDX42	1	RECQL5	0.94	STK11IP	0.19
GRINL1A	1	KPNA1	0.93	MAF1	0.18
GTF2B	1	KPNA6	0.93	RPS7	0.18
GTF2F1	1	TUBB4Q	0.93	WBSCR16	0.18
GTF2F2	1	INTS5	0.91	PRKCSH	0.17
GTF2I	1	ZNRD1	0.91	RACGAP1	0.17
HSPA8	1	ARHGEF1	0.9	GTF2IP1	0.14
PCBP1	1	PTBP1	0.9	FIBP	0.12
PDRG1	1	RQCD1	0.9	PDCD6	0.11
PFDN2	1	CRCP	0.89	NXF1	0.08
PFDN6	1	SUPT4H1	0.88	PRPS1L1	0.08
PIH1D1	1	CHD3	0.87	RAVER1	0.08

POLR1A	1	RFC1	0.87	ARMC5	0.07
POLR1C	1	SNRPB	0.87	CDKL3	0.07
POLR1D	1	SNRPN	0.87	DNAH1	0.07
POLR1E	1	Gcom1	0.86	PRPS1	0.07
POLR2A	1	GTF3C3	0.86	REPS1	0.07
POLR2B	1	EIF4A2	0.85	RIC8A	0.07
POLR2C	1	TMTC3	0.84	SFRS15	0.07
POLR2D	1	XPO1	0.84	ZFP91	0.07
POLR2G	1	LOC283767	0.82	ZFP91-CNTF	0.07
POLR2H	1	PCBP3	0.82	DIAPH2	0.06
POLR2I	1	LOC730826	0.81	LCK	0.06
POLR2J	1	KPNA5	0.78	MCM4	0.06
POLR2L	1	LOC100129958	0.74	MDN1	0.06
POLR3A	1	XPO5	0.73	SMTN	0.06
POLR3C	1	GFAP	0.72	119584193	0.05
POLR3D	1	RPA2	0.71	CHPF	0.05
POLR3E	1	PICK1	0.69	FAM120B	0.05
POLR3F	1	MED17	0.68	FLJ36131	0.05
POLR3H	1	ARMCX3	0.66	FMNL3	0.05
PPP2R1A	1	BBX	0.63	KIAA1543	0.05
RP11-631M21.2	1	MED15	0.62	LSM3	0.05
RPAP2	1	SAT1	0.62	OBSL1	0.05
RPAP3	1	MED31	0.59	TOP1	0.05
RPRD1A	1	RPL38	0.59	XPNPEP3	0.05
RUVBL1	1	GBP2	0.58	C6orf204	0.04
RUVBL2	1	SH2D4B	0.57	CCDC144A	0.04
SPTAN1	1	POLE	0.56	CCDC144B	0.04
TRIP12	1	LOC653781	0.55	CKAP2	0.04
TUBA3E	1	F7	0.47	COL1A2	0.04
TUBB2A	1	SNRPD3	0.44	KIAA1539	0.04
TUBB3	1	MED10	0.42	LOC440243	0.04
TUBB4	1	MED27	0.41	LOC645202	0.04
UXT	1	TRAP1	0.38	LOC727832	0.04
WDR92	1	C14orf166	0.37	TUFT1	0.04
HSP90AA1	0.99	GRINL1B	0.34	HNRNPU	0
HSP90AB1	0.99	PRCC	0.34	HSPA1A	0
MED4	0.99	POLR3GL	0.31	HSPA1B	0
PCBP2	0.99	CCDC88A	0.27	HSPA6	0
POLR3B	0.99	PTCD3	0.26	IPO5	0
POLR3G	0.99	RBM16	0.26	MTHFD2	0
POLR3K	0.99	MED1	0.24	POLR2F	0
SMARCA4	0.99	MAP4K4	0.22	POLR2K	0
HSPA5	0.98	TFDP1	0.21	TUBA1A	0
PHGDH	0.98	GCDH	0.2	TUBA1B	0
TUBA4A	0.98	MUC19	0.2	TUBB	0
TUBA4B	0.98	TPR	0.2	TUBB2C	0
SPTBN1	0.97	119607622	0.19		

**Tableau 4. Liste des protéines partenaires d'interaction co-purifiées avec PFDN4.**

Protéine	Score IR	Protéine	Score IR	Protéine	Score IR
PFDN6	1	NEFH	0.06	MYHa	0.04
TUBB3	0.96	PACSIN2	0.06	PTCH1	0.04
KIN	0.84	C19orf60	0.05	PTPRM	0.04
SH2D4B	0.62	CTAG1A	0.05	RLN1	0.04
HNRNPA1	0.59	CTAG1B	0.05	RLN2	0.04
Gcom1	0.58	DNAH17	0.05	SF4	0.04
MKL2	0.58	DNM1	0.05	TBX18	0.04
PCBP2	0.42	FLJ36131	0.05	TR2IT1	0.04
DYX1C1	0.3	GBP3	0.05	TUT1	0.04
STK11IP	0.21	TOP1	0.05	EEF1A1	0
119607622	0.2	C13orf35	0.04	PFDN1	0
DNAH9	0.2	CHRM1	0.04	PFDN5	0
GBP2	0.19	CRYZL1	0.04	TUBA1B	0
MRPL52	0.19	CYB	0.04	TUBA1C	0
NSBP1	0.18	GCDH	0.04	VBP1	0
SLIT1	0.18	MKI67	0.04		
LOC100126583	0.06	MYH10	0.04		

**Tableau 5. Liste des protéines partenaires d'interaction co-purifiées avec MBD2.**

Protéine	Score IR	Protéine	Score IR	Protéine	Score IR
GATAD2A	1	KPNB1	0.73	ENAH	0.06
GATAD2B	1	SRI	0.69	MCM4	0.06
HDAC1	1	KPNA6	0.68	PHGDH	0.06
HDAC2	1	KIF5B	0.66	119568149	0.05
MTA2	1	HNRNPH1	0.65	BEND3	0.05
MTA3	1	BAT2D1	0.39	CAMK2D	0.05
RBBP4	1	TIMM13	0.32	CCDC109B	0.05
RBBP7	1	CALM2	0.31	EIF4A1	0.05
TUBA1A	1	KPNA1	0.26	EIF4A2	0.05
TUBA1B	1	KPNA5	0.26	INF2	0.05
TUBA3C	1	TUBA1C	0.24	21754970	0.04
TUBB	1	DDX5	0.18	CAMKK1	0.04
TUBB2A	1	SMARCE1	0.18	CDH7	0.04
TUBB2C	1	STARD3	0.18	CEP78	0.04
LOC642954	0.97	MCM7	0.14	GPR123	0.04
MTA1	0.97	LOC400248	0.12	GPR22	0.04
HSPA1A	0.95	ARID3A	0.11	JARID1A	0.04
HSPA8	0.92	WDR77	0.11	KDM5A	0.04
CDK2AP1	0.9	C2	0.1	P4HB	0.04
KPNA3	0.9	CFB	0.1	SSBP1	0.04
RUVBL1	0.9	CALM3	0.08	EEF1A1	0

HSPA1L	0.89	RBM14	0.08	EEF1AL3	0
CAND1	0.81	CAPNS2	0.07	ELAVL1	0
HNRNPU	0.79	MATR3	0.07	HNRNPA2B1	0
HSPA1B	0.79	UBE2N	0.07	KHSRP	0
TUBA8	0.79	UBE2NL	0.07		

**Tableau 6. Liste des protéines partenaires d'interaction co-purifiées avec XPA.**

Protéine	Score IR	Protéine	Score IR	Protéine	Score IR
CAND1	1	UTRN	0.17	ACTR1A	0.06
PARP1	1	NCAPD3	0.16	CDK2AP1	0.06
RFC1	1	SFRS7	0.16	CDK2AP2	0.06
RFC2	1	BCR	0.15	COMMD3	0.06
RFC3	1	PRSS3	0.14	DDX49	0.06
RFC4	1	MED23	0.13	DIDO1	0.06
RFC5	1	RPS6KB2	0.13	DLK1	0.06
RPA1	1	DRAP1	0.12	FGFR3	0.06
RPA2	1	RBM25	0.12	IGF2BP1	0.06
RPA3	1	ALDH18A1	0.11	LOC100129492	0.06
XRCC6	1	PRDX3	0.11	LOC730839	0.06
ACTA1	0.99	REPS1	0.11	LYN	0.06
ACTA2	0.99	RPL24	0.11	NXF1	0.06
ACTC1	0.99	TRAP1	0.1	PCTK2	0.06
ACTG2	0.99	TRIO	0.1	PLG	0.06
HSPA2	0.99	TRRAP	0.1	RPS24	0.06
HTT	0.92	UNC45A	0.1	RPS6	0.06
XRCC5	0.88	DUSP9	0.09	TBP	0.06
CHD3	0.84	LYPLA2	0.09	ARPC3	0.05
DDX3Y	0.8	UBA5	0.09	BTF3	0.05
OSBPL3	0.56	ABCF2	0.08	DUSP23	0.05
TUBB1	0.54	ARF1	0.08	EIF3K	0.05
DDX46	0.46	C8orf42	0.08	HAUS5	0.05
U2AF1	0.41	COBRA1	0.08	KALRN	0.05
LOC653269	0.38	DDX25	0.08	KIAA0841	0.05
NFATC2IP	0.33	GFAP	0.08	RBBP5	0.05
LOC100129958	0.31	HSP90AA5P	0.08	SNRPE	0.05
NDEL1	0.31	MAD2L1	0.08	SPTBN1	0.05
IDH3B	0.28	PRPS1	0.08	TCEB1	0.05
HSPA5	0.26	RPL23AP42	0.08	223365	0.04
SRCAP	0.26	RPL23AP5	0.08	CDC34	0.04
DDX6	0.25	RPS18	0.08	CSPG4	0.04
KPNA5	0.23	SUPT6H	0.08	DDX43	0.04
SNRPA	0.23	USP24	0.08	FAM184B	0.04
HRNR	0.22	BLK	0.07	LOC391656	0.04
LIMS1	0.2	CHAF1B	0.07	LOC652875	0.04
VASP	0.2	EIF2AK2	0.07	LOC730429	0.04

SHOC2	0.19	LOC731751	0.07	PFKFB2	0.04
CALM2	0.17	PRKDC	0.07	SLC5A5	0.04
CALM3	0.17	RAN	0.07	UBR5	0.04
CUL4A	0.17	SNRPB2	0.07	LOC727761	0
IK	0.17	ABCB10	0.06		

**Tableau 7. Liste des protéines partenaires d'interaction co-purifiées avec CAND1.**

Protéine	Score IR	Protéine	Score IR	Protéine	Score IR
CUL1	1	CDC42EP4	0.08	CAMKK1	0
CUL4B	1	CDK2	0.08	CAPNS1	0
CUL5	1	EXOSC1	0.08	CCT4	0
HSPA2	1	SCYL2	0.08	CKAP5	0
KIF5C	1	YWHAB	0.08	CNOT1	0
CUL3	0.99	ANXA7	0.07	DDB1	0
HSPA1L	0.99	CUX1	0.07	DDX1	0
PPP3CA	0.99	FGFR2	0.07	DDX3X	0
CHD3	0.92	FLG2	0.07	DDX42	0
XPA	0.91	RNF7	0.07	DDX5	0
EIF3D	0.88	C9orf40	0.06	DIAPH1	0
KPNA6	0.88	CAMK1	0.06	DOCK6	0
TIMM13	0.75	CCDC85C	0.06	DOCK7	0
BAT2	0.7	CCRK	0.06	DYNC1H1	0
BAT2L	0.66	CDC2L5	0.06	EEF1A1	0
DDX3Y	0.65	FHL1	0.06	EIF3L	0
SMARCC2	0.62	GTF2H2	0.06	EIF4A1	0
LOC653269	0.54	GTF2H2B	0.06	ENAH	0
CCT7	0.48	LOC730394	0.06	GCN1L1	0
FAM82B	0.43	PCMT1	0.06	GTF2I	0
ANXA11	0.38	SNX24	0.06	GTF3C1	0
LOC727761	0.36	ABL2	0.05	HECTD1	0
XRN2	0.34	ACACA	0.05	HSPA1B	0
TCOF1	0.32	BPTF	0.05	HSPA8	0
LRCH2	0.29	CARD14	0.05	INF2	0
FUBP1	0.24	CCDC75	0.05	IPO5	0
C14orf133	0.21	CTAG1A	0.05	KHSRP	0
HRNR	0.2	HIRIP3	0.05	KIF5B	0
PPP2R1A	0.2	LOC119358	0.05	KLC2	0
DKC1	0.19	LRBA	0.05	KPNA2	0
PFKP	0.19	MUTED	0.05	KPNB1	0
CCDC72	0.18	NOP10	0.05	LRRC40	0
EXOG	0.18	RBX1	0.05	LRRC41	0
EXOSC4	0.18	SNRPD2	0.05	NOC4L	0
PPP2R5A	0.18	TSC1	0.05	PCBP1	0
SMYD3	0.18	TSPAN32	0.05	PCBP2	0

YWHAG	0.17	37574658	0.04	RPA1	0
CSNK2A1P	0.16	119570769	0.04	SBF1	0
JMJD1B	0.16	C15orf38	0.04	SF3B2	0
YWHAZ	0.16	C18orf10	0.04	SUPT6H	0
CAMK2B	0.15	CHRAC1	0.04	TIA1	0
MOBK13	0.15	KIAA0195	0.04	TRIP12	0
ARF5	0.14	LAMB4	0.04	TRIP13	0
PRSS1	0.14	NUP210	0.04	TUBA1A	0
CSNK2B	0.13	STRN	0.04	TUBA1B	0
PLEC1	0.13	STYXL1	0.04	TUBA1C	0
PRPF31	0.13	TBC1D9B	0.04	TUBB	0
RPL5	0.13	WDR43	0.04	TUBB2A	0
MTMR1	0.12	ARID3A	0	TUBB2C	0
TBC1D2B	0.12	ARNT	0	TUBB3	0
PPP3R1	0.11	BAT2D1	0	TUBB4	0
DNAJC12	0.1	C1orf163	0	VPS25	0
EPS8	0.1	C22orf28	0	XPO1	0

**Tableau 8. Liste des protéines partenaires d'interaction co-purifiées avec PDCD5.**

Protéine	Score IR	Protéine	Score IR	Protéine	Score IR
CCT2	1	ELAVL2	0.09	FLJ21369	0.04
CCT3	1	ELAVL4	0.09	HIRIP3	0.04
CCT4	1	EXOSC6	0.09	JMJD1C	0.04
CCT5	1	TUBA4A	0.09	N4BP3	0.04
CCT6A	1	TUBA4B	0.09	NOM1	0.04
CCT7	1	EHBP1	0.08	OFD1	0.04
CCT8	1	UACA	0.08	SNRPE	0.04
HSPA1L	1	DST	0.07	SPANXA1	0.04
TCP1	1	LOC440589	0.07	STYXL1	0.04
DDX3Y	0.99	LOC645018	0.07	ARID3A	0
HSPA1A	0.98	MLL3	0.07	BAT2D1	0
CCT6B	0.96	RPS2P40	0.07	C1orf163	0
TUBA8	0.93	VPS25	0.07	CAMKK1	0
CAB39	0.89	C19orf43	0.06	CKAP5	0
CDK9	0.88	CDK2	0.06	DDX3X	0
RUVBL2	0.86	CDK5	0.06	DDX42	0
LOC391656	0.68	CNP	0.06	EEF1A1	0
HAS1	0.64	DSN1	0.06	ELAVL1	0
XPO5	0.36	EXOSC4	0.06	ENAH	0
PPP1CC	0.32	HELLS	0.06	HECTD1	0
CAMK2D	0.31	RPS4Y1	0.06	HNRNPA1	0
EIF4A1	0.31	BCR	0.05	HNRNPA2B1	0
EIF4A2	0.31	C1orf49	0.05	HNRNPH1	0
FAM82B	0.28	C9orf82	0.05	HSPA1B	0
MAK	0.28	CAPZB	0.05	HSPA6	0

DNAJC10	0.24	CHCHD1	0.05	HSPA8	0
PSMA5	0.24	CRKRS	0.05	KHSRP	0
GAPVD1	0.22	CYB	0.05	KIF5B	0
PCTK1	0.22	FGFR1	0.05	PCBP1	0
DLK1	0.21	FKSG2	0.05	PCBP2	0
RBM22	0.2	IL18	0.05	PDCD2L	0
HNRNPH2	0.18	NBEAL2	0.05	PPP3CA	0
DDX1	0.15	ODZ2	0.05	PRPS1	0
NOC4L	0.15	PIK3CA	0.05	RPS2	0
PPIH	0.15	POU5F1	0.05	SUPT6H	0
YWHAG	0.15	PQBP1	0.05	TCEAL4	0
YWHAH	0.15	RBM16	0.05	TRIP12	0
YWHAQ	0.15	RPA3	0.05	TUBA1A	0
ZGPAT	0.14	RPL5	0.05	TUBA1B	0
CAB39L	0.13	ABCB10	0.04	TUBA1C	0
SYNE1	0.12	C7orf70	0.04	TUBB	0
LRCH2	0.11	C8orf40	0.04	TUBB2C	0
SLAIN2	0.11	CCDC14	0.04	TUBB3	0
MBD3	0.1	CCKAR	0.04	ZBED1	0
BAT2	0.09	DHX16	0.04		
DHX15	0.09	DPP6	0.04		