

Direction des bibliothèques

AVIS

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Étude structure / fonction des sous-unités catalytiques de l'ARN polymérase II

par :

Céline Domecq

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
en biochimie
avril, 2008

©, Céline Domecq, 2008



Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :
Étude structure / fonction des sous-unités catalytiques de l'ARN polymérase II

présenté par :
Céline Domecq

président-rapporteur

a été évalué par un jury composé des personnes suivantes :

Pascal Chartrand

président-rapporteur

Benoit Coulombe

directeur de recherche

Jacques Archambault

membre du jury

RÉSUMÉ

L'ARN polymérase (ARNP) II est une enzyme multimérique qui transcrit l'ensemble des gènes codant les protéines chez les eucaryotes. Ce mémoire rapporte les résultats d'une étude structure / fonction de l'ARNP II réalisée dans le but de mieux comprendre le fonctionnement moléculaire de l'enzyme.

Dans ce mémoire, je décrirai d'abord mon travail visant à poursuivre la caractérisation des mutants de l'ARNP II humaine déjà commencée au sein du laboratoire du Dr. Coulombé. Pour ce faire, une méthode de purification de l'ARNP II humaine a été mise au point en utilisant des sous-unités doublement étiquetées (TAP) dans des lignées cellulaires ECR 293. Le complexe étiqueté de l'ARNP II est fonctionnellement actif *in vitro* et *in vivo*. En utilisant les données cristallographiques comme guide, des mutants de l'ARNP II ont été créés dans les boucles « switch » 1 et « switch » 2 ainsi que dans la fermeture à glissière. Ces ARNP II mutants ont été purifiés et caractérisés *in vitro* par des essais de retards sur gel et de transcription.

Par la suite, je décrirai mon travail visant à caractériser l'ARNP II de la levure, *Saccharomyces cerevisiae*. Nous avons choisi d'utiliser cet organisme pour la suite du projet car le choix des mutations les plus informatives s'avérait difficile à faire avec notre modèle humain. *S. cerevisiae*, est un organisme de choix car il nous a permis de créer et de caractériser facilement et rapidement un grand nombre de mutants. Nous avons choisi de caractériser d'autres éléments structuraux clés du site catalytique, la boucle d'embranchement 2 et les acides aminés du site actif même. De nouveaux mutants de l'ARNP II ont été créés, purifiés et caractérisés. L'étude de ces mutations a été faite *in vitro* par des essais de transcription et *in vivo* par des tests de croissance sur milieu sélectif.

Les études structure / fonction de l'ARNP II menées chez l'Humain et chez *S. cerevisiae* nous ont permis de mieux comprendre le mécanisme moléculaire de l'enzyme et les différentes étapes de la transcription génique.

Mots clés : Transcription, ARNP II, mutagenèse, purification TAP, retard sur gel, cellules EcR 293, *S. cerevisiae*.

ABSTRACT

RNAPII is a central multimeric enzyme that transcribes all genes coding for proteins in eukaryotes. The goal of my master's project was to perform a structure / function study of RNAPII in order to obtain a better understanding of the enzyme's molecular mechanism but also a better knowledge of the different steps involved in gene transcription.

In the first aim of my project, I pursued the characterization of human RNAPII mutants, already begun in Dr. Coulombe's laboratory. The purification method of human RNAPII was developed using double tagged subunits in EcR293 cells. Tagged RNAPII complex containing 12 sub-units is functionally active *in vitro* and *in vivo*. Using structural data obtained by crystallography, as a guide, RNAPII mutants were created in three loops of the active site: switch 1, switch 2 and zipper. These mutant RNAPII were purified and characterized *in vitro* by gel shift and transcription assays.

The second aim of my master's project consisted in characterizing RNAPII from the yeast *Saccharomyces cerevisiae*. We have chosen to use this organism because deciding on which exact mutations (e.g. substitution or deletion, number of residues in alanine scans, etc.) are likely to be the most informative represents a difficulty in our method. *S. cerevisiae* is a well suited organism for this project and has allowed us to create and characterize easily and quickly numerous mutants. We have chosen to characterize other structural elements of the catalytic site, mainly the fork loop 2 and amino acids of the active site. New RNAPII mutants were created, purified and characterized. Protocols for protein purification and biochemical assays were adapted to yeast. Study of yeast RNAPII was done *in vitro* by transcription assays and *in vivo* by assessing growth on selective media.

These structure / function studies of RNAPII from human cells and yeast have led to a better understanding of the molecular mechanism of this enzyme.

Key words : Transcription, RNAPII, mutagenesis, TAP purification, gel shift, EcR 293 cells, *S. cerevisiae*.

TABLE DES MATIÈRES	
RÉSUMÉ	iii
ABSTRACT	iv
TABLE DES MATIÈRES	v
LISTE DES TABLEAUX	ix
LISTE DES FIGURES	x
LISTE DES ABBRÉVIATIONS	xi
REMERCIEMENTS	xv
1 INTRODUCTION	1
1-1 Structure d'un gène de classe II	1
1-1.1 Promoteur basal	1
1-1.1.1 La boîte TATA	2
1-1.1.2 L'initiateur	3
1-1.1.3 Le DPE	4
1-1.1.4 Le DCE	5
1-1.1.5 Le BRE	5
1-1.1.6 Le MTE	6
1-1.2 Séquences régulatrices	6
1-1.2.1 Séquences activatrices	6
1-1.2.1.1 Les UAS	6
1-1.2.1.2 Les enhanceurs	7
1-1.2.1.3 Les LCR	7
1-1.2.2 Séquences répressives	7
1-1.2.2.1 Les URS	7
1-1.2.2.2 Les silencers	7
1-1.2.3 Insulator	8
1-2 Chromatine : obstacle à la transcription	8
1-2.1 Modifications des histones	9
1-2.2 Facteurs de remodelage	10
1-3 Facteurs généraux de la transcription	10
1-3.1 TFIID	12
1-3.1.1 TBP	12
1-3.1.2 TAF	12
1-3.1.3 TRF	13

2-1.1.2	Transfection stable des cellules avec l'ADNc d'intérêt	41
2-1.1.3	Purification par double affinité	41
2-1.1.3.1	Extraction des protéines cellulaires	42
2-1.1.3.2	Méthode TAP	43
2-1.1.3.2.1	La colonne d'IgG	43
2-1.1.3.2.2	La colonne de calmoduline	43
2-1.1.4	Concentration et identification des complexes purifiés	44
2-1.2	Levures	44
2-1.2.1	Construction du plasmide contenant le gène d'intérêt fusionné à l'étiquette TAP	44
2-1.2.1.1	PCR	45
2-1.2.1.2	Digestion enzymatique	46
2-1.2.1.3	Ligation	46
2-1.2.1.4	Transformation des bactéries compétentes	47
2-1.2.1.5	Mutagenèse	47
2-1.2.2	Transformation des levures	48
2-1.2.3	Échange de plasmide : « plasmid shuffling »	48
2-1.2.4	Sélection sur 6-azauracile	49
2-1.2.5	Purification par double affinité	49
2-1.2.5.1	Extraction des protéines cellulaires	49
2-1.2.5.2	Méthode TAP et identification des complexes	49
2-2	Création des souches <i>RPB2</i> -TAP et KO <i>DST1</i>	50
2-3	Essai de transcription <i>in vitro</i>	50
2-4	Essai d'élongation <i>in vitro</i>	51
2-5	Essai de retard sur gel (EMSA)	52
2-5.1	Synthèse des sondes radioactives	52
2-5.2	Assemblage	52
2-5.2.1	Au promoteur	52
2-5.2.2	ADN non spécifique	53
2-5.3	Gel de retardement	53

3	RÉSULTATS	54
3-1	Purification par double affinité de l'ARNPII	54
3-2	Étude structure / fonction du site catalytique de l'ARNPII humaine	55
3-2.1	Purification des ARNPII humaines	57
3-2.2	Caractérisation des ARNPII humaines mutées	58
3-3	Étude structure / fonction du site catalytique de l'ARNPII chez <i>S. cerevisiae</i>	61
3-3.1	Purification des ARNPII de levures	64
3-3.2	Caractérisation des ARNPII mutées de levures	65
4	DISCUSSION	68
4-1	Purification simple et efficace de l'ARNPII	68
4-2	Choix de la levure comme système modèle d'étude	69
4-3	Étude structure / fonction du site catalytique de l'ARNPII	70
4-3.1	Chez l'Humain	70
4-3.2	Chez <i>S. cerevisiae</i>	73
5	CONCLUSION ET PERSPECTIVES	75
6	LISTE DES RÉFÉRENCES	77

LISTE DES TABLEAUX

Tableau 1 : Facteurs généraux de la transcription associés au promoteur basal.

Tableau 2 : Les facteurs généraux de la transcription et leur rôle.

Tableau 3 : Les sous-unités de l'ARNPII.

Tableau 4 : Les domaines du site catalytique de l'ARNPII.

Tableau 5 : Les facteurs d'élongation de l'ARNPII et leurs rôles.

LISTE DES FIGURES

- Figure 1 : Promoteur basal d'un gène de classe II.
- Figure 2 : Structure de l'ARNPII à 12 sous-unités.
- Figure 3 : Éléments structuraux du site actif de l'ARNPII.
- Figure 4 : Schéma de la structure de l'étiquette TAP et de la méthode de purification TAP.
- Figure 5 : Schéma résumant la construction des souches de levures.
- Figure 6 : Immunobuvardage de préparations d'ARNPII humaine sauvage purifiées avec des concentrations différentes de sels.
- Figure 7 : Représentation schématique des mutations de la protéine Rpb1 humaine dans la séquence et la structure.
- Figure 8 : Gel SDS-PAGE 4-12% des ARNPII humaines mutantes purifiées.
- Figure 9 : Gel SDS-PAGE 4-12% des ARNPII humaines mutantes purifiées.
- Figure 10 : Essai de transcription *in vitro*.
- Figure 11 : Essai de retard sur gel de liaison non spécifique des ARNPII humaines à l'ADN.
- Figure 12 : Essai de retard sur gel des ARNPII humaines au promoteur.
- Figure 13 : Représentation schématique de la méthode ayant permis de générer les souches de levures et de purifier les ARNPII sauvage et mutantes chez la levure.
- Figure 14 : Représentation schématique des mutations de la protéine Rpb2 de *S. cerevisiae* dans la séquence et la structure.
- Figure 15 : Gel SDS-PAGE 4-12% des ARNPII mutantes purifiées chez la levure.
- Figure 16 : Phénotypes de croissance des souches mutantes dans différentes conditions.
- Figure 17 : Essai d'élongation *in vitro*.

LISTE DES ABRÉVIATIONS

5FOA	Acide 5 fluoroorotique
6-AU	6-azauracile
Ada	Adaptor
Ade	Adénine
AdML	<i>Adenovirus major late</i>
AdMLP	<i>Adenovirus major late promoter</i>
ADN	Acide désoxyribonucléique
ADNc	ADN codant
ADP	Adénosine diphosphate
ARN	Acide ribonucléique
ARNm	ARN messenger
ARNP	ARN polymérase
ARNr	ARN ribosomaux
ARNsn	<i>Small nuclear RNA</i>
<i>ARNsno</i>	<i>Small nucleolar RNA</i>
ARNt	ARN de transfert
ATP	Adénosine triphosphate
ATPase	ATP hydrolase
BRE ^d	<i>TFIIB recognition element downstream</i>
BRE ^u	<i>TFIIB recognition element upstream</i>
CAK	<i>Cyclin-dependant kinase activating kinase complex</i>
CBP	<i>Calmodulin-binding peptide</i>
CDK	<i>Cyclin dependant kinase</i>
CF	<i>Cleavage Factor</i>
CHD	<i>Chromodomain SWI/SNF-like helicase/ATPase domain and DNA binding domain</i>
ChIP	Immunoprécipitation de la chromatine
CHRAC	<i>Chromatin accessibility complex</i>
CPF	<i>Cleavage / polyadenylation factor</i>
CS	Syndrome de Cockayne
CTD	<i>Carboxy-terminal domain</i>
CTDK-1	<i>CTD kinase 1</i>

CTP	Cytosine triphosphate
DCE	<i>Downstream core element</i>
dNTP	Désoxyribonucléotide triphosphate
DMT	<i>DNA methyltransferase</i>
DPE	<i>Downstream promoter element</i>
DSIF	<i>DRB-sensitivity inducing factor</i>
EcRE	<i>Ecdysone receptor element</i>
ELL	<i>Eleven-nineteen lysine-rich in leukemia</i>
EMSA	<i>Electrophoretic mobility shift assay</i>
FACT	<i>Facilitates chromatin transcription</i>
Fcp1	<i>TFIIF-associated CTD phosphatase</i>
FGT	Facteur général de la transcription
GAL	Galactose
GCN	<i>General control of amino acid synthesis</i>
GTP	Guanosine triphosphate
HAT	<i>Histone acétyltransférase</i>
HBX	<i>Hepatitis B virus X protein</i>
His	Histidine
HTH	Hélice tour hélice
IgG	Immunoglobuline G
Inr	Initiateur
ISWI	<i>Imitation of SWI</i>
KO	<i>Knock-out</i>
LCR	<i>Locus control region</i>
Leu	Leucine
Med	Médiateur
MTE	<i>Motif ten element</i>
MS/MS	<i>Double Tandem mass spectrometry</i>
NELF	<i>Negative elongation factor</i>
NER	<i>Nucleotide excision repair</i>
NoCR	<i>Nucleolar remodeling complex</i>
NTP	Ribonucléotide triphosphate
NuRD	<i>Nucleosome remodeling factor</i>
pb	Paire de base

PCAF	<i>p300/CBP-associated factor</i>
Poly(dI-dC)	<i>Poly desoxyinosine - desoxycytosine</i>
P-TEFb	<i>Positive transcription elongation factor b</i>
Py	<i>Pyrimidine</i>
RAP	<i>RNA polymerase II-associated protein</i>
Rpb	<i>RNA polymerase B</i>
RSC	<i>Remodel the structure of chromatin</i>
RSF	<i>Remodeling and spacing factor</i>
RXR	<i>Retinoid X receptor</i>
SAGA	<i>SPT-ADA-GCN5 acétyltransférase</i>
SANT	<i>SWI3, ADA2, N-CoR, TFIIB</i>
SCP1	<i>Small CTD phosphatase 1</i>
SNF	<i>Sucrose non fermenting</i>
Spt	<i>Supressor of Ty insertion</i>
Srb	<i>Supressor of RNA polymerase B</i>
STAGA	<i>Supressor of Ty insertion-TAF-Gcn5-acetyltransferase</i>
SWI	<i>Switching mating-type</i>
TAF	<i>TBP-associated factor</i>
TAP	<i>Tandem Affinity Purification</i>
TBP	<i>TATA binding protein</i>
TCR	<i>Transcription coupled repair</i>
TEV	<i>Tobacco etch virus</i>
TF	<i>Transcription factor</i>
TFTC	<i>TBP-free TAF-containing complex</i>
TLF	<i>TBP like factor</i>
TLP	<i>TBP like protein</i>
TRF	<i>TBP related factor</i>
Trp	<i>Tryptophane</i>
TTD	<i>Trichothiodystrophie</i>
TTP	<i>Thymidine triphosphate</i>
UAS	<i>Upstream activating sequence</i>
UBLCP1	<i>Ubiquitin-like domain containing CTD phosphatase 1</i>
Ura	<i>Uracile</i>
URS	<i>Upstream repressing sequence</i>

UTP	Uridine triphosphate
UV	Ultraviolet
VP16	<i>Herpes simplex virion protein 16</i>
XP	<i>Xeroderma pigmentosum</i>

REMERCIEMENTS

Je tiens à remercier mon directeur de recherche, Benoit Coulombe, pour son encadrement, son soutien et ses conseils durant ma maîtrise.

J'aimerais aussi remercier Annie Bouchard et Diane Forget pour leur aide pratique, leurs conseils et leur support. Je remercie aussi tous les membres du laboratoire passé et présent avec qui j'ai aimé travailler et échanger.

Enfin, je tiens à remercier mon conjoint, Olivier Jeffroy, ainsi que ma famille pour m'avoir encouragée et soutenue durant mes études.

1 INTRODUCTION

Les ARN polymérases (ARNP) dépendantes de l'ADN sont des enzymes qui catalysent la formation d'un ARN à partir d'une matrice d'ADN. Chez les eucaryotes, il existe trois ARNP I, II et III qui synthétisent chacune des types spécifiques d'ARN. Ainsi, l'ARNPI est responsable de la synthèse des ARN ribosomiaux (ARNr), l'ARNPII, des ARN messagers (ARNm), des ARNsn (« small nuclear RNA ») et des ARNsno (« small nucleolar RNA ») et l'ARNPIII, des ARN de transfert (ARNt) et de l'ARNr 5s.

L'ARNm est l'intermédiaire qui permet la synthèse d'une protéine via le mécanisme de la traduction. Le nombre de gènes codant pour une protéine est estimé à 5500 chez la levure et environ 21000 dans les cellules humaines. La transcription génique est un processus qui est hautement régulé dans la cellule pour lui permettre de répondre à son environnement, de se diviser et de se spécialiser de manière adéquate. La régulation de la transcription par l'ARNPII est donc un élément crucial pour la cellule. La régulation de la transcription se fait à plusieurs niveaux grâce à un grand nombre de protéines et de séquences d'ADN régulatrices.

1-1 Structure d'un gène de classe II

1-1.1 Promoteur basal

Le promoteur basal a été défini comme la région minimale d'ADN suffisante pour permettre l'initiation de la transcription par la machinerie de l'ARNPII. Plusieurs éléments sont retrouvés dans la séquence du promoteur basal comme la boîte TATA, l'initiateur, le DPE, le DCE, le BRE et le MTE (Figure 1).

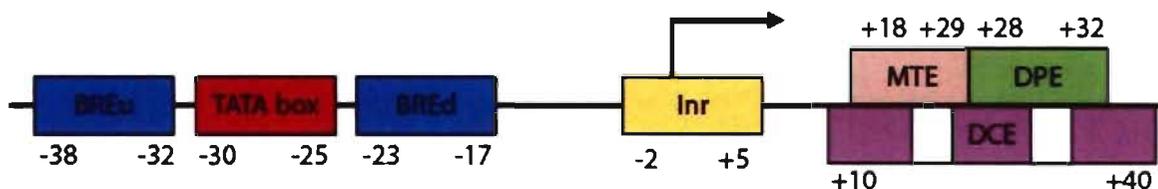


Figure 1 : Promoteur basal d'un gène de classe II. Il faut noter que les éléments identifiés dans cette figure se retrouvent dans certains promoteurs avec des combinaisons variables.

Les éléments composant le promoteur basal interagissent avec les facteurs généraux de la transcription (FGT) pour permettre l'initiation de la transcription. Les différents éléments du promoteur basal, leur séquence consensus et les protéines qui les lient sont résumés dans le tableau 1.

Tableau 1 : Facteurs généraux de la transcription associés au promoteur basal.

Élément du promoteur basal	Séquence consensus	Protéine associée
BRE ^u	(G/C)(G/C)(G/A)CGCC	TFIIB
Boite TATA	TATA(A/T)AA(A/G)	TBP
BRE ^d	(G/A)T(T/G/A)(T/G)(G/T)(T/G)(T/G)	TFIIB
Initiateur	PyPyAN(T/A)PyPy	TAF
MTE	C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C)	
DCE	SI : CTTC	TAF
	SII : CTGT	
	SIII : AGC	
DPE	(A/G)G(A/T)(C/T)(G/A/C)	TAF

1-1.1.1 La boite TATA

La boite TATA est le premier élément du promoteur basal à avoir été découvert en 1979 par Golberg [revu par Smale et Kadonaga, 2003]. Il s'agit d'une séquence du promoteur riche en AT qui est conservée chez la majorité des eucaryotes. Elle est située en amont du site d'initiation de la transcription (+1), généralement localisée entre -25 et -30 pb. Néanmoins, chez la levure *Saccharomyces cerevisiae*, elle est localisée entre -40 et -120 pb [Struhl, 1989]. Sa séquence consensus est la suivante : TATA(A/T)AA(A/G).

La boite TATA est reconnue et liée par la protéine TBP (TATA Binding Protein). TBP fait partie d'un plus large complexe TFIID qui sera décrit par la suite (section 1-3.1). Même si TBP a une très grande affinité pour la séquence consensus de la boite TATA [Wong et Bateman, 1994], elle est aussi capable de lier une grande diversité de séquences riches en AT [Hahn *et al.*, 1989; Patikoglou *et al.*, 1999]. La boite TATA

fut longtemps considérée comme un élément essentiel et immuable du promoteur de tous les gènes eucaryotes. Cependant, au fur et à mesure de la caractérisation des promoteurs eucaryotes, il a été montré que la boîte TATA n'est pas présente au niveau de tous les promoteurs et que sa séquence peut varier. Ainsi, chez *S. cerevisiae*, seuls 20% des promoteurs contiennent une séquence fidèle à la séquence consensus [Basehoar *et al.*, 2004]. Seulement 33% des 1941 promoteurs prédits chez la Drosophile la contiennent [Ohler *et al.*, 2002] et 24 à 32% des promoteurs potentiels chez l'Homme [Suzuki *et al.*, 2001; Yang *et al.*, 2007].

1-1.1.2 L'initiateur

L'initiateur (Inr) est une séquence conservée qui comprend le site +1 d'initiation de la transcription. Dans les premières études comparatives des séquences promotrices, il a été montré que le site +1 était généralement une adénosine (A), le site -1 une cytosine (C) et que les nucléotides voisins de ces deux positions étaient le plus souvent des pyrimidines (Py) [Corden *et al.*, 1980]. Plusieurs études ont montré que des mutations de ces bases entraînaient une réduction de l'efficacité de la transcription et une hétérogénéité du choix du site d'initiation [Grosschedl et Birnstiel, 1982; Corden *et al.*, 1980; Talkington et Leder, 1982; Dierks *et al.*, 1983; Concino *et al.*, 1984; Tokunaga *et al.*, 1984].

La séquence consensus de l'Inr est la suivante : PyPyA+1N(T/A)PyPy [Javahery *et al.*, 1994; Lo et Smale, 1996.]. Elle a été définie à l'origine chez l'Homme mais elle est identique chez la Drosophile. L'Inr est présent dans 69% des 1941 promoteurs prédits chez la Drosophile [Ohler *et al.*, 2002] et 46% des promoteurs chez l'Homme [Yang *et al.*, 2007].

L'Inr a un rôle semblable à la boîte TATA et il est capable de fonctionner indépendamment de celle-ci. Leur action est synergique s'ils sont séparés de 25 à 30 pb mais ils agissent indépendamment s'ils sont séparés de plus de 30 pb. Quand ils sont séparés de 15 à 20 pb, ils ont toujours une action synergique. Cependant, le site +1 est déplacé car le choix de celui-ci est dicté par la position de la boîte TATA. L'Inr, quant à lui, dicte la direction de la transcription [O'Shea-Greenfield et Smale, 1992; Ren *et al.*,

2005]. L'action synergique de ces deux éléments est confirmée par leur liaison commune à TFIID [Emami *et al.*, 1997]. L'Inr est, en effet, reconnu et lié spécifiquement par le complexe TFIID via les sous-unités TAF (TBP Associated Factors). TBP ne peut à lui seul reconnaître l'Inr [Smale, 1997]. L'Inr est également capable d'interagir avec d'autres protéines, notamment l'ARNPII.

Les promoteurs qui ne contiennent pas de boîte TATA mais un Inr peuvent être liés par TBP via une séquence riche en A/T divergente de la boîte TATA en -25 -30 pb. Dans le cas d'autres promoteurs sans boîte TATA, la liaison de TBP à l'ADN n'est pas nécessaire [Martinez *et al.*, 1995].

1-1.1.3 Le DPE

Le DPE (Downstream Promoter Element) est un élément du promoteur basal conservé de la Drosophile à l'Homme. Il est principalement retrouvé dans les promoteurs sans boîte TATA. Le DPE a été identifié dans ce type de promoteur grâce à sa capacité de liaison à TFIID. Le promoteur type contenant un DPE contient également un Inr. Le DPE est situé en aval du site +1 entre +28 et +32 pb [Kutach et Kadonaga, 2000.]. Sa séquence consensus est la suivante : (A/G)G(A/T)(C/T)(G/A/C) [Smale et Kadonaga, 2003].

Le DPE agit en synergie avec l'initiateur. En effet, des mutations dans le DPE ou dans l'Inr entraînent une perte de liaison avec TFIID et de l'activité transcriptionnelle [Burke et Kadonaga, 1996.]. De plus, la distance entre ces deux éléments est primordiale, une seule insertion ou délétion entre les deux éléments affectent considérablement la transcription. Différentes études ont montré que le DPE seul n'est pas capable d'assurer la transcription d'un gène, c'est la principale distinction entre la boîte TATA et le DPE [Kutach et Kadonaga, 2000.]. Cependant, une étude récente [Lewis *et al.*, 2005] montre qu'un promoteur contenant un DPE et l'Inr ne peut pas activer la transcription en présence du seul TFIID. L'activation de la transcription nécessite la présence d'autres protéines, ce qui suggère un niveau de régulation supérieur de la transcription.

1-1.1.4 Le DCE

Le DCE (Downstream Core Element) est un élément qui a été récemment découvert en aval du site +1 dans le promoteur du gène de la β -globuline humaine [Lewis *et al.*, 2000]. Cet élément a été caractérisé par des mutations dans la région +10 et +40 pb. Il est composé de trois sous-éléments appelés SI, SII et SIII. Une mutation dans un de ses sous-éléments entraîne des réductions drastiques de la transcription. Les séquences consensus de ces trois sous-éléments sont respectivement les suivantes pour SI, SII et SIII: CTTC, CTGT et AGC.

Le DCE, comme le DPE, est reconnu et lié par TFIID via les TAF. [Lewis *et al.*, 2000; Lee *et al.*, 2005]. Le DCE diffère cependant du DPE en plusieurs choses. Le DPE est centré autour de +30 alors que le DCE s'étend de +10 à +40. De plus, contrairement au DPE, il agit dans un contexte plus large de promoteur. En effet, il agit avec à la fois une boîte TATA et un Inr mais aussi au sein de promoteur sans boîte TATA. Enfin, le DPE et le DCE ne sont pas reconnus par les mêmes TAF.

1-1.1.5 Le BRE

Le BRE (TFIIB Recognition Element) est un élément du promoteur qui est reconnu et lié par TFIIB. Il est composé de deux éléments localisés immédiatement en amont BRE^u (BRE upstream), et en aval BRE^d (BRE downstream) de la boîte TATA. Les séquences consensus des BRE^u et BRE^d sont respectivement les suivantes : (G/C)(G/C)(G/A)CGCC et (G/A)T(T/G/A)(T/G)(G/T)(T/G)(T/G).

Le BRE^u, le premier découvert, est un élément conservé chez les *Archae* avec des séquences non similaires mais non conservé chez la Levure et chez les Plantes. Il peut avoir une action positive ou négative sur la transcription [Lagrange *et al.*, 1998; Qureshi et Jackson, 1998; Evans *et al.*, 2001; Chen et Manley, 2003].

Plus récemment, le BRE^d a été découvert, et son action dépend de la composition du promoteur. Il a un effet négatif quand un BRE^u est présent dans le promoteur mais il a un effet positif en l'absence de BRE^u [Deng et Robert, 2005].

Les promoteurs sans boîte TATA ont plus souvent un BRE que les autres. Le BRE^d a la même prévalence que le BRE^u mais ils ont des distributions indépendantes au sein des promoteurs [Deng et Robert, 2005].

1-1.1.6 Le MTE

Le MTE (Motif Ten Element) est un élément du promoteur basal qui a été récemment découvert par l'analyse informatique des séquences promotrices chez la Drosophile [Ohler *et al.*, 2002]. Il est conservé de la Drosophile à l'Homme. Il est localisé entre +18 et +29 pb et possède la séquence consensus suivante :

C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C).

L'étude de mutations a permis de caractériser cet élément. Ainsi il a été montré qu'il s'agit d'un élément différent du DPE même si leurs séquences se chevauchent. En effet, le MTE peut compenser la perte du DPE ainsi que celle de la boîte TATA. De plus, il a été montré que la distance entre l'Inr et le MTE était primordiale pour l'action de ce dernier. Le MTE peut interagir avec TFIID [Lim *et al.*, 2004].

1-1.2 Séquences régulatrices

En plus du promoteur basal, d'autres séquences sont essentielles à la régulation de la transcription par l'ARNPII. Les séquences régulatrices permettent aux activateurs ou aux répresseurs de se lier à l'ADN. On retrouve dans ces séquences : les UAS, les enhancers, les URS, les silencers et les LCR.

1-1.2.1 Séquences activatrices

1-1.2.1.1 Les UAS

Les UAS (Upstream Activating Sequence) sont des éléments auxquels se lient des activateurs. Elles agissent sur la transcription quand elles sont situées près du site d'initiation de la transcription et en amont de celui-ci [Lee et Young, 2000; Guarente et Hoar, 1984].

1-1.2.1.2 Les enhancers

Les enhancers sont, comme les UAS, liés par des activateurs de la transcription. Contrairement aux UAS, ces séquences sont capables d'influencer la transcription même à de grandes distances et ce, indépendamment de leur orientation [Blackwood et Kadonaga, 1998; Lee et Young, 2000].

1-1.2.1.3 Les LCR

Les LCR (Locus Control Region) sont semblables aux enhancers dans la mesure où ils servent de sites de liaison pour les activateurs. Cependant, ils sont souvent arrangés en complexes. Leur action est indépendante de leur site d'intégration dans la chromatine mais leur effet est limité par leur orientation et leur distance au promoteur [Bulger et Groudine, 1999; Lee et Young, 2000].

1-1.2.2 Séquences répressives

1-1.2.2.1 Les URS

Les URS (Upstream Repressing Sequence) sont des séquences spécifiques reconnues par les répresseurs de la transcription. Les répresseurs transcriptionnels sont capables d'inhiber la transcription de différentes manières : interférence avec les activateurs, remodelage de la chromatine ou blocage du recrutement de la machinerie de la transcription [Struhl, 1995; Lee et Young, 2000; Hanna-Rose et Hansen, 1996; Gaston et Jayaraman, 2003].

1-1.2.2.2 Les silencers

Les silencers sont des éléments qui ont été définis par opposition aux enhancers. Leur action répressive est indépendante de leur position et de leur orientation [Ogbourne et Antalis, 1998]. L'ADN subissant l'effet d'un silencer semble être inhibé à la suite d'interactions entre les histones et des protéines ciblant la désacétylation [Lee et Young,

2000].

1-1.2.3 L'insulator

L'insulator est un élément qui a dans un premier temps été défini par sa capacité à bloquer les effets des séquences régulatrices définies ci-dessus [Fourel *et al.*, 2004]. Il isole ainsi les régions chromosomiques ce qui évite de moduler de manière indésirable la transcription d'autres gènes. Il bloque l'action des séquences régulatrices quand il est situé entre la séquence régulatrice et le promoteur. Son mode d'action n'est pas encore clairement établi mais il semble que son effet soit lié à la structure de l'ADN. Un deuxième type d'insulator a été défini. Il agit comme une barrière pour empêcher la diffusion de l'hétérochromatine. En effet, l'insulator sert de point d'ancrage pour le recrutement de facteurs remodelant la chromatine. Des insulators possédant ces 2 fonctions ont déjà été décrits [Gaszner et Felsenfeld, 2006].

1-2 Chromatine : obstacle à la transcription

Dans le noyau des cellules eucaryotes, l'ADN se retrouve sous forme condensée et il est associé à diverses protéines. C'est cet ensemble formé par l'ADN et des protéines qui forme la chromatine. Il existe plusieurs niveaux de compaction de la chromatine. À la base de l'organisation de la chromatine, se trouve le nucléosome. Le « cœur » du nucléosome est composé d'un octamère d'histones qui comprend deux copies des histones H2A, H2B, H3 et H4. Autour de cet octamère est enroulé environ 147pb d'ADN. Il s'agit de la structure dite en « collier de perles », elle a une épaisseur d'environ 10nm [revu par Horn et Peterson, 2002].

Les histones sont des petites protéines basiques contenant deux domaines fonctionnels : un domaine C-terminal nommé « histone fold » qui interagit avec les autres histones et l'ADN et un domaine N-terminal riche en acides aminés basiques. Les domaines N-terminal et C-terminal sont le site de nombreuses modifications post-traductionnelles. Le domaine N-terminal des histones est situé à l'extérieur du

nucléosome où il peut interagir avec des protéines, l'ADN situé entre deux nucléosomes et l'ADN d'autres nucléosomes. Les interactions entre nucléosomes et l'interaction des nucléosomes avec les histones « linker », H1 ou H5, permettent de passer à un niveau de condensation supérieur de la chromatine en fibre de 30nm de diamètre [revu par Woodcock et Dimitrov, 2001]. Le degré de condensation de la chromatine varie au cours du cycle cellulaire et en fonction de l'ADN qui la compose. Ainsi la chromatine peut atteindre un haut niveau de compaction quand les fibres interagissent entre elles pour former les chromatides durant la métaphase de la mitose. La manière dont les fibres de 30nm se replient *in vivo* pour former des chromatides de 100 à 300nm d'épaisseur n'a pas encore été élucidée. Deux modèles ont été proposés : un repliement en solénoïde (hélices de nucléosomes) ou en zigzag (empilement des nucléosomes). À l'opposé, durant l'interphase, où les gènes sont activement transcrits et l'ADN se réplique pour la mitose, la chromatine est alors décondensée et dispersée dans le noyau, il s'agit de l'euchromatine. Cependant, environ 10% de la chromatine interphasique reste très condensée, il s'agit de l'hétérochromatine. Elle se compose essentiellement de séquences d'ADN répétées et se situe majoritairement au niveau des centromères des chromosomes. Elle contient très peu de gènes et il a été montré que si un gène actif était déplacé dans une zone hétérochromatique il perdait son activité [Eberl *et al.*, 1993].

La structure de la chromatine est dynamique, elle est constamment remodelée au cours du cycle cellulaire. Il existe plusieurs types de complexes qui régulent sa structure. Ainsi, l'ADN n'est pas facilement accessible. La chromatine est donc une barrière physique à la transcription, la régulation de la structure chromatinienne permet un niveau supplémentaire de régulation de la transcription.

1-2.1 Modifications des histones

Aussi bien les queues que les domaines globulaires des histones sont l'objet de nombreuses modifications post-traductionnelles: acétylation (lysine), méthylation (lysine, arginine), phosphorylation (sérine, thréonine), ubiquitination (lysine), sumoylation (lysine) ou ADP-ribosylation (glutamate) [revu par Khorasanizadeh, 2004].

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. On parle alors de modifications de l'euchromatine. On parlera de modifications de l'hétérochromatine pour les modifications qui inactivent des gènes ou de plus grandes régions chromosomiques. Pour la plupart, ces modifications sont distribuées selon un "pattern" distinct au niveau de la région en amont d'un gène.

1-2.2 Facteurs de remodelage

Les complexes utilisant l'hydrolyse de l'ATP pour altérer les contacts entre l'ADN et les histones sont de manière générale appelés complexes de remodelage de la chromatine [revu par Lusser et Kadonaga, 2003]. Ces complexes possèdent tous une sous-unité ATPasique qui appartient à la famille *Snf2-like*. Les sous-familles de complexes de remodelage sont définies par rapport à la présence de domaines protéiques spécifiques distincts de l'ATPase. Ainsi les membres de la sous-famille SNF2 (SWI/SNF, RSC et hSWI/SNF) possèdent un bromodomaine, ceux de la sous-famille ISWI (RSF, hCHRAC, NoCR) un domaine SANT et la sous-famille CHD1 (NuRD, Mi-2) un chromodomaine et un domaine de liaison à l'ADN.

L'action de ces différents complexes a pour conséquences un changement d'accessibilité à l'ADN, ce qui module l'activité transcriptionnelle. Il peut s'agir d'un désenroulement passager de l'ADN situé au bout d'un nucléosome, la formation d'une boucle d'ADN, ou une translation du nucléosome ("sliding").

1-3 Facteurs généraux de la transcription

L'ARNPII catalyse la synthèse de l'ensemble des ARNm chez les eucaryotes. Cependant, elle est incapable d'initier seule la transcription au niveau d'un promoteur ou de répondre aux protéines régulatrices de la transcription en l'absence d'autres facteurs. La capacité de l'ARNPII d'initier la transcription est restaurée en présence d'autres facteurs, les facteurs généraux de la transcription (FGT). Les FGT ont été définis comme

les protéines minimales nécessaires à l'initiation de la transcription par l'ARNPII *in vitro* (Tableau 2).

Tableau 2 : Les facteurs généraux de la transcription et leur rôle.

Facteur	Sous-unité	Fonction
TFIID	TBP	Liaison à la boîte TATA
	TAF	Liaison à l'Inr et au DPE, Réponse aux activateurs
TFIIA	α	Liaison à TBP, Stabilisation de la liaison TBP-ADN, Activation de la transcription
	β	
	γ	
TFIIB		Liaison à TBP, Stabilisation de la liaison TBP-ADN, Liaison aux BRE, Choix du site +1
TFIIF	RAP30/Tfg1	Recrutement de l'ARNPII, Enroulement du promoteur,
	RAP74/Tfg2	Dégagement du promoteur, Stimulation de l'élongation
TFIIE	α	Recrutement de TFIIH, Régulation des activités kinases et hélicases de TFIIH, Ouverture du promoteur, Dégagement du promoteur
	β	
TFIIH	XPB	Hélicase, Ouverture du promoteur, Dégagement du promoteur, NER
	XPB	Hélicase, Ouverture du promoteur, NER, Stimulation de la transcription
	p62	NER
	p52	NER
	p44	Ubiquitine ligase, NER, Stimulation de XPD
	cycline H	Phosphorylation du CTD, Dégagement du promoteur
	p34	NER
	MAT1	Interaction avec Cdk7 et la cycline H
	Cdk7	Kinase, Phosphorylation du CTD, Dégagement du promoteur
	p8	NER, Stabilité TFIIH
TFIIS		Stimulation de l'activité de clivage de l'ARNPII, Sortie des sites d'arrêt

1-3.1 TFIID

TFIID est le complexe qui se fixe le premier au promoteur et permet la formation du complexe de préinitiation. Il est conservé entre les espèces. Ce complexe est composé de TBP associé à environ 14 TAF [revu par Hampsey, 1998; Davidson, 2003].

1-3.1.1 TBP

TBP est la protéine qui reconnaît et s'associe à la boîte TATA. C'est une protéine bipartite avec une région aminoterminal variable et un domaine carboxyterminal très conservé (80% de conservation de la Levure à l'Homme). Il lie la boîte TATA et favorise la liaison d'autres facteurs de transcription au promoteur. En se liant à la boîte TATA, il adopte une structure en «selle de cheval» avec une face concave hydrophobe interagissant avec l'ADN et une face convexe qui interagit avec TFIIA, TFIIB et les TAF [Kim *et al.*, 1993a; Kim *et al.*, 1993b]. Le domaine aminoterminal présente une importante variabilité de séquence et de taille en fonction des espèces. Chez la levure, ce domaine est non essentiel à la viabilité cellulaire et à la transcription *in vitro*. De plus, il a été montré que le domaine carboxyterminal seul a une plus grande affinité pour l'ADN *in vitro* et que le domaine aminoterminal inhibe la liaison à l'ADN *in vivo*. Il apparaît donc que la conformation de la protéine TBP influence sa capacité de liaison à l'ADN.

1-3.1.2 TAF

Les TAF ont été définies comme des cofacteurs associés à TBP au sein du complexe TFIID. Les TAF 1 à 15 (auparavant nommées TAFII) sont les protéines qui s'associent à TBP pour former TFIID. Seul TBP est requis à la transcription basale *in vitro* mais TFIID semble nécessaire pour répondre aux activateurs. Ainsi, il a été montré que plusieurs activateurs lient directement les TAF et que certaines TAF sont capables de reconnaître des éléments du promoteur comme le DPE ou l'Inr [Sauer et Tjian, 1997].

Les TAF ne sont pas seulement les composantes de TFIID. Elles sont aussi présentes dans quatre autres complexes, chez les mammifères: TFTC (TBP-free TAF-

containing complex) [Wieczorek *et al.*, 1998], PCAF (p300/CBP associated factor) [Ogryzko *et al.*, 1998], les complexes SAGA (Spt-Ada-Gcn5 acetylase) [Grant *et al.*, 1998] et STAGA (SPT3-TAF9-GCN5 containing complex) [Ogryzko *et al.*, 1998]. TFTC est structurellement similaire à TFIID, il peut remplacer *in vitro* TFIID au niveau des promoteurs avec et sans boîte TATA. Les complexes TFTC, PCAF, SAGA et STAGA ont tous une activité HAT (Histone AcetylTransférase).

1-3.1.3 TRF

Pendant longtemps, TBP fut considéré comme un facteur universel de la transcription pour les trois ARNP eucaryotes. Néanmoins la découverte de protéines similaires à TBP, les TRF (TBP Related Factor), a remis en question ce dogme. Il existe trois TRF. La première protéine découverte, TRF1, semble présente uniquement chez la Drosophile. Le domaine carboxyterminal de TRF1 a une très forte homologie avec celui de TBP. Il a été montré que TRF1 remplace TBP dans la transcription par l'ARNPIII et qu'il était capable de se lier aux promoteurs sans boîte TATA [Crowley *et al.*, 1993]. TRF2, aussi nommée TLF (TBP Like Factor), TLP (TBP Like Protein) ou TRP (TBP Related Protein), a une homologie de 40% avec le domaine carboxyterminal de TBP. Elle est retrouvée chez la plupart des eucaryotes [Dantonel *et al.*, 1999]. Pour le moment, la séquence reconnue par TRF2 n'a pas été décrite mais il ne s'agit pas de la boîte TATA [Rabenstein *et al.*, 1999]. Il apparaît que TRF2 est nécessaire à l'expression spécifique de gènes durant l'embryogenèse [Veenstra *et al.*, 2000]. Enfin, la protéine TRF3 présente une très forte homologie avec le domaine carboxyterminal de TBP (plus de 90%). Elle est retrouvée chez les vertébrés mais pas chez les eucaryotes inférieurs. Tout comme TBP, elle fait partie d'un complexe à plusieurs sous-unités mais de taille inférieure à TFIID [Persengiev *et al.*, 2003].

1-3.1.4 Structure

La structure de TFIID en forme de fer à cheval à trois lobes a été définie par microscopie électronique [Leurent *et al.*, 2002; Leurent *et al.*, 2004]. Il apparaît que TBP

est situé au centre des lobes ce qui lui permet d'interagir avec l'ADN. Les TAF1 et 5 ont un rôle structural de liaison des trois lobes et les autres TAF sont réparties dans les trois lobes. Deux conformations, ouverte et fermée de TFIID ont été identifiées ce qui suggère qu'il pourrait agir comme une pince pour lier l'ADN [Andel *et al.*, 1999; Brand *et al.*, 1999].

1-3.2 TFIIA

TFIIA est une protéine qui se lie à TBP et augmente son affinité à la boîte TATA *in vitro* [Yokomori *et al.*, 1994]. TFIIA peut ainsi se lier de manière non spécifique à l'ADN [Tan *et al.*, 1996; Geiger *et al.*, 1996]. L'association de TBP et de TFIIA stabilise la liaison de TBP à l'ADN et facilite ainsi la formation du complexe de préinitiation. TFIIA n'est pas nécessaire *in vitro* pour la transcription basale quand TBP est seul, mais en présence de TFIID il s'avère indispensable [Sun *et al.*, 1994]. Le fait que TFIIA ne soit pas toujours requis dans des systèmes de transcription *in vitro* remet en cause son rôle de facteur général de la transcription, il s'agirait plutôt un activateur de la transcription. Cette idée est supportée par le fait que TFIIA interagit avec un grand nombre de facteurs comme TFIID, le complexe SAGA, des activateurs et des répresseurs de la transcription [Hoiby *et al.*, 2007].

TFIIA est composé de deux sous-unités chez la Levure et de trois chez la Drosophile et l'Homme (α , β , γ). Les deux plus grands polypeptides de TFIIA (α et β) sont codés par un seul gène dont le produit est clivé. Les 2 sous-unités ainsi clivées sont ciblées par la dégradation protéosomale, le niveau de TFIIA dans la cellule est donc régulé par ce clivage [Hoiby *et al.*, 2004]. De plus, il a été montré que lorsque le niveau d'expression de TFIIA est faible, le taux de transcription est faible aussi. Il apparaît donc que la régulation du taux de TFIIA cellulaire peut contribuer à réguler l'expression génique.

1-3.3 TFIIB

Le facteur TFIIB est une protéine monomérique composée d'un domaine carboxyterminal contenant un motif HTH (Hélice Tour Hélice) et d'une région aminoterminal liant le zinc. TFIIB interagit avec TBP pour faciliter le recrutement de l'ARNPII et de TFIIF au promoteur [Buratowski *et al.*, 1989]. Il est indispensable pour la transcription *in vitro*. Des études de mutagenèse ont montré que TFIIB avait un rôle dans le choix du site +1 [Berroteran *et al.*, 1994; Pinto *et al.*, 1994; Pinto *et al.*, 1992]. Son domaine carboxyterminal se lie à TBP et interagit directement avec l'ADN, via son domaine HTH, à la fois en amont et en aval de la boîte TATA (au niveau du BRE^u et/ou du BRE^d dans les promoteurs les contenant) [Deng et Roberts, 2005]. La région aminoterminal de TFIIB est en forme de doigt de zinc. Elle interagit avec l'ARNPII et les deux sous-unités de TFIIF [Ha *et al.*, 1993; Pardee *et al.*, 1998].

Les domaines aminoterminal et carboxyterminal sont capables d'interagir entre eux en une conformation dite "fermée". Cette interaction intramoléculaire permet la régulation de l'activité de TFIIB [Elsby et Roberts, 2004].

1-3.4 TFIIF

Le facteur TFIIF est un facteur essentiel à la transcription *in vitro*. Il est composé de plusieurs sous-unités, chez la Levure il contient trois sous-unités Tfg1, Tfg2 et Tfg3 alors qu'il n'en contient que deux chez l'Homme RAP30 et RAP74. Il n'existe pas chez l'Homme d'équivalent à Tfg3 néanmoins elle présente une forte homologie avec la sous-unité TAF14 de TFIID. Chez l'Homme, TFIIF se lie au complexe de préinitiation sous la forme d'un hétérotétramère (2 molécules de RAP30 et 2 de RAP74). Les deux plus grandes sous-unités (Tfg1, Tfg2 et RAP74, RAP30) présentent des homologies de séquences et de fonction avec la sous-unité σ de l'ARNP bactérienne [Garrett *et al.*, 1992; McCracken et Greenblatt, 1991; Sopta *et al.*, 1989; Sun et Hampsey, 1995; Yonaha *et al.*, 1993]. De la même manière que σ , TFIIF joue un rôle dans l'initiation de la transcription, il favorise l'arrivée de l'ARNPII dans le complexe de préinitiation grâce à sa forte affinité avec celle-ci [Sopta *et al.*, 1985]. Tout comme σ , TFIIF, via sa sous-

unité RAP30, empêche les liaisons non spécifiques de l'ARNPII à l'ADN. TFIIF, via sa sous-unité RAP74, joue également un rôle avec TFIIB dans le choix du site +1 [Freire-Picos *et al.*, 2005]. La sous-unité RAP74 a un autre rôle important dans l'initiation de la transcription, il permet l'enroulement du promoteur autour du complexe de préinitiation. En effet, il a été montré que RAP74 induit un changement de topologie au niveau du complexe d'initiation contenant TBP, TFIIB, TFIIF, TFIIE et l'ARNPII [Robert *et al.*, 1998; revu par Coulombe et Burton, 1999]. L'entrée de RAP74 dans ce complexe induit une courbure de l'ADN près du site +1 ce qui permet l'enroulement du promoteur autour de l'ARNPII. Cette hypothèse a été confirmée par des images de microscopie électronique de l'ARNPII avec les FGT [Forget *et al.*, 1997]. TFIIF permet le dégagement du promoteur en arrêtant le cycle des initiations avortées et prévient avec TFIIH l'arrêt des complexes d'élongation précoces.

Le facteur TFIIF est aussi impliqué dans la phase d'élongation de la transcription. Il augmente l'efficacité d'élongation et stimule la phosphatase Fcp1 du domaine carboxyterminal de la sous-unité Rpb1 (CTD) de l'ARNPII. Son activité stimulatrice sur l'élongation est différente de celle de TFIIS. Il peut interférer *in vitro* avec celui-ci. TFIIF n'est pas associé au complexe d'élongation de façon stable mais de façon transitoire [Zawel *et al.*, 1995; Sims *et al.*, 2004]. Il supprime les pauses transitoires que l'ARNPII fait durant l'élongation. Récemment, il a été montré que TFIIF seul ne peut pas aider l'ARNPII mais que la présence du facteur p-TEFb (positive Transcription Elongation Factor b) est nécessaire pour supprimer les pauses [Cheng et Price, 2007].

1-3.5 TFIIE

Le facteur TFIIE est composé de deux sous-unités riches en acides aminés chargés, TFIIE α et TFIIE β [Ohkuma *et al.*, 1990]. Ces deux sous-unités sont essentielles à la transcription *in vitro*. Tout comme TFIIF, il se lie au complexe de préinitiation sous la forme d'un hétérotétramère [Ohkuma *et al.*, 1990; Hayashi *et al.*, 2005; Chen *et al.*, 2007]. Il arrive dans le complexe de préinitiation après le recrutement de l'ARNPII et interagit avec la forme non phosphorylée de l'ARNPII, TFIIB et les deux sous unités de TFIIF. La présence de TFIIE est requise pour le recrutement au complexe de

préinitiation de TFIIF et la régulation des activités kinase et hélicase de celui-ci [Flores *et al.*, 1992; Ohkuma et Roeder, 1994]. La sous-unité α de TFIIE via son domaine carboxyterminal va se lier de manière spécifique à TFIIF [Chen *et al.*, 2007]. La sous-unité β lie l'ADN double et simple brin(s), TFIIB et RAP30. TFIIE se lie au promoteur de manière non spécifique au niveau de la région -10/+10 pb [Kim *et al.*, 2000]. L'action de TFIIE et TFIIF permet l'ouverture du complexe d'initiation. Par la suite, avec TFIIF et TFIIF, il facilite le dégagement du promoteur, l'arrêt du cycle d'initiations avortées et le passage à la phase d'élongation productive [Goodrich et Tjian, 1994; Yamamoto *et al.*, 2001; Kumar *et al.*, 1998; Watanabe *et al.*, 2003].

1-3.6 TFIIF

Le facteur TFIIF est un complexe multimérique composé de 10 sous-unités d'une masse moléculaire totale équivalente à celle de l'ARNPII. C'est un facteur essentiel de la transcription. Il joue un rôle important dans l'initiation de celle-ci. Il est divisé en deux sous-complexes le noyau (core) et le CAK (Cyclin-Activating Kinase complex). Le noyau de TFIIF chez l'Homme contient les sous-unités suivantes: XPB, p62, p52, p44, p34 et p8 [Coin *et al.*, 2006]. La protéine XPD relie le noyau de TFIIF au sous-complexe CAK. Le CAK comprend cdk7, la cycline H et MAT1. C'est le seul des FGT ayant une activité enzymatique. Il a une activité kinase via sa sous-unité cdk7, deux activités hélicases via ses sous-unités XPB et XPD et une activité ubiquitine ligase par p44 [Takagi *et al.*, 2005]. Le rôle de TFIIF n'est pas restreint à la transcription. En effet, le noyau de TFIIF est un composant essentiel de la machinerie de réparation de l'ADN par excision nucléotidique (NER) et de la réparation de l'ADN couplée à la transcription (TCR). Le sous-complexe CAK, quand il est seul, joue un rôle dans le contrôle du cycle cellulaire [Laroche *et al.*, 1998]. Durant la transcription, TFIIF permet la formation du complexe ouvert à travers l'activité hélicase de XPB et XPD. Cependant, seule XPB est essentielle à la transcription. Des mutations dans les protéines XPB, XPD et p8 sont responsables de maladies chez l'Homme: le *Xeroderma Pigmentosum* (XP), le syndrome de Cockaine (CS) et la trichothiodystrophie (TTD). Il s'agit dans les trois cas d'une extrême photosensibilité mais dans le cas du XP la maladie est due à des défauts du

NER alors que pour le CS et le TTD, il s'agit de défauts dans le TCR et la transcription [Giglia-Mari *et al.*, 2004].

Après l'ouverture de la bulle de transcription entre -9 et +2, l'ARNPII commence la synthèse de l'ARN. La bulle de transcription s'étend au fur et à mesure de l'ajout des nucléotides. Après l'ajout du quatrième nucléotide, le complexe transcriptionnel change de conformation, il devient plus stable et TFIIH n'est plus nécessaire. Le sous-complexe CAK est aussi impliqué dans la transcription. Quand il fait partie de TFIIH, le CAK via *cdk7* permet la phosphorylation du CTD de l'ARNPII [Rossignol *et al.*, 1997]. L'état de phosphorylation de l'ARNPII joue un rôle important durant la transcription. En effet, l'ARNPII qui est recrutée au promoteur est hypophosphorylée (forme I₀). La phosphorylation du CTD de Rpb1 (forme I₂a) va permettre le dégagement de l'ARNPII du promoteur et l'entrée en phase d'élongation, possiblement en modifiant les interactions de l'ARNPII avec ses partenaires.

1-3.7 TFIIS

Le facteur TFIIS est essentiel à la transcription, mais contrairement aux précédents FGT, il n'est pas impliqué dans l'initiation de la transcription, mais principalement dans la phase d'élongation.

In vitro, TFIIS est capable de stimuler le clivage de l'ARN naissant de l'ARNPII en arrêt. L'ARNPII a tendance à reculer sur la matrice d'ADN quand l'hybride ADN-ARN est instable au niveau du site catalytique (« backtracking ») jusqu'à l'atteinte d'un registre plus stable et à s'arrêter [Nudler *et al.*, 1997]. Afin de sortir de l'arrêt et de poursuivre la transcription, elle doit cliver l'extrémité 3' de l'ARNm. Ceci génère une nouvelle extrémité 3' de l'ARN qui pourra être utilisée par l'ARNPII pour reprendre la transcription. *In vivo*, de nombreuses études ont montré que TFIIS joue un rôle en élongation [Exinger et Lacroute F, 1992; Pokholok *et al.*, 2002; Hartzog *et al.*, 1998; Orphanides *et al.*, 1999; Costa et Arndt, 2000; Mason et Struhl, 2005].

TFIIS est composé de trois domaines dont les conformations sont indépendantes les unes des autres [Morin *et al.*, 1996; Awrey *et al.*, 1998; Olmsted *et al.*, 1998; Booth *et al.*, 2000]. Le domaine I qui comprend les 130 premiers acides aminés n'est pas requis

pour les fonctions connues de TFIIS même si des interactions biochimiques et génétiques entre ce domaine et des facteurs d'initiation ont été décrites [Pan *et al.*, 1997; Wery *et al.*, 2004; Malagon *et al.*, 2004; Davie et Kane, 2000; Fish *et al.*, 2006]. Le domaine II est relié au domaine III par un domaine de liaison (linker) non structuré. Le domaine II (aa 131-240) et le linker sont requis pour la liaison à l'ARNPII. Le domaine III (aa 265-309) est essentiel pour la stimulation du clivage d'ARN. Les domaines II et III à eux seuls supportent toutes les activités connues de TFIIS *in vitro* et *in vivo*. Cependant, il a été montré que le domaine aminoterminal de TFIIS est essentiel pour une bonne interaction avec l'holoenzyme [Pan *et al.*, 1997]. De plus, le domaine I est capable d'interagir avec le médiateur et le complexe SAGA [Wery *et al.*, 2004]. Une autre étude a mis en évidence la nécessité de TFIIS pour recruter TBP et l'ARNPII au promoteur ainsi que pour recruter SAGA et le médiateur au niveau des UAS [Prather *et al.*, 2005]. Enfin plusieurs études génétiques ont montré qu'il existait des interactions entre les gènes de TFIIS et ceux des facteurs d'initiation de la transcription comme des composants du Médiateur, Swi/Snf et TFIIF [Malagon *et al.*, 2004; Davie et Kane, 2000; Fish *et al.*, 2006].

Dernièrement, des études ont mis en évidence par spectrométrie de masse la présence de TFIIS au sein du complexe de préinitiation au niveau du promoteur [Ranish *et al.*, 2003; Kim *et al.*, 2007]. TFIIS semble recruté de manière stable au promoteur mais dépendante de TBP, STAGA et le médiateur. De plus, il apparaît que TFIIS est requis pour la formation et la stabilité du complexe de préinitiation. Le domaine II et le linker suffisent pour stimuler la formation du complexe de préinitiation bien que le domaine I soit également essentiel à cette activité. Le domaine III, quant à lui, n'est pas nécessaire [Ranish *et al.*, 2003; Kim *et al.*, 2007]. Ces résultats sont confirmés par une autre étude qui a mis en évidence le recrutement de TFIIS au niveau de certains promoteurs, la nécessité de sa présence pour un recrutement efficace de l'ARNPII en l'absence de médiateur [Guglielmi *et al.*, 2007].

1-3.8 Le médiateur

Le médiateur est un complexe polypeptidique. Ce n'est pas un facteur général de la transcription. Il est requis pour la réponse aux régulateurs de la transcription (activateurs et répresseurs) à tous les promoteurs de classe II [revu par Myers et Kornberg, 2000]. C'est la raison pour laquelle le médiateur fait partie de la machinerie basale de la transcription. Il a été découvert par Kornberg et collaborateurs qui ont montré que, chez *S. cerevisiae*, ce complexe était essentiel *in vitro* à l'activation de la transcription par les protéines Gcn4 et GAL4-VP16 dans un système reconstitué contenant l'ARNPII et les FGT [Kim, *et al.*, 1994].

Chez la Levure, le complexe est composé d'une vingtaine de polypeptides comprenant les protéines Srb (Suppressor of RNA polymerase B) et Med, ainsi que d'autres protéines découvertes par de nombreux criblages génétiques [Nonet et Young, 1989]. Chez les mammifères, le médiateur a également été mis en évidence. Il est constitué d'une trentaine de polypeptides mais seulement huit ont des homologues chez *S. cerevisiae*. La purification du médiateur a montré, selon les méthodes et les outils, une grande variabilité dans la composition du complexe aussi bien chez la Levure que chez l'Homme [Myers et Kornberg, 2000; Lee et Young, 2000; Lewis et Reinberg, 2003]. Cette variabilité met en évidence son rôle d'interface entre les régulateurs de l'expression génique et la machinerie transcriptionnelle de l'ARNPII.

Le médiateur peut être purifié en association avec l'ARNPII ou sans elle. Le terme d'holoenzyme est utilisé quand le médiateur et l'ARNPII sont copurifiés ensemble. Cependant, le concept d'holoenzyme chez les eucaryotes n'est pas le même que chez les procaryotes. En effet, l'holoenzyme bactérienne est suffisante à elle seule pour permettre l'initiation de la transcription ce qui n'est pas le cas chez les eucaryotes. Il a été montré qu'il n'y a pas un pré-assemblage de l'holoenzyme avant sa liaison au promoteur comme c'est le cas chez les procaryotes [Featherstone, 2002; Kuras *et al.*, 2003]. De plus, la quantification des composants de la machinerie transcriptionnelle n'appuie pas l'existence d'un pré-assemblage d'une holoenzyme [Borggreffe *et al.*, 2001]. Enfin, il a été montré que la liaison de l'ARNPII au médiateur n'est pas

nécessaire *in vitro* et *in vivo* à la fonction activatrice du médiateur [Bhoite *et al.*, 2001. ; Park *et al.*, 2001].

Le médiateur est composé de trois domaines : la « tête », le « milieu » et la « queue ». Ces trois domaines constituent le « petit » complexe qui active la transcription. En présence d'un quatrième module, le module CDK, le médiateur (sous forme d'un « grand » complexe) réprime la transcription. Le médiateur interagit avec l'ARNPII au niveau du CTD via la « tête ». La « queue » sert de plateforme de liaison pour les activateurs [revu par Casamassimi et Napoli, 2007; Conaway *et al.*, 2005]. Il sert d'adaptateur entre les modulateurs liés aux séquences régulatrices et le complexe de préinitiation de la transcription lié au promoteur. Grâce à sa capacité de liaison à l'ARNPII, il peut aussi stimuler la transcription basale. Mais contrairement à l'ARNPII et à certains FGT, il est incapable de se lier à des séquences spécifiques d'ADN. Il reste encore à déterminer par quel mécanisme le médiateur transmet le signal des régulateurs à la machinerie transcriptionnelle. Le médiateur est donc un facteur essentiel de la transcription des gènes de classe II. Il semble influencer à la fois le recrutement de la machinerie transcriptionnelle et la formation du complexe d'initiation de la transcription.

1-4 L'ARN polymérase II

1-4.1 Structure

L'ARNPII est composée de douze sous-unités, nommées Rbp 1 à 12, dont les séquences sont conservées chez les eucaryotes. Leur poids moléculaire varie de 10 à 220 kDa (Tableau 3). Les deux plus grandes sous-unités de l'ARNPII, Rpb1 et Rpb2, respectivement 220 et 140 kDa, forment le centre catalytique de l'enzyme. La comparaison de séquences entre l'ARNPII et l'ARNP procaryote a montré une forte homologie entre les deux enzymes. La majorité des acides aminés conservés se situent au niveau du site actif. En effet, il a été montré que les sous-unités Rpb1 et Rpb2 sont homologues aux sous-unités β' et β de l'ARNP procaryote. L'ARNP procaryote contient également un homodimère constitué de deux sous-unités α . Les sous-unités Rpb3 et Rpb11 présentent des homologies avec α et s'assemblent sous la forme d'un

hétérodimère, semblable à l'homodimère procaryote. Enfin, la sous-unité Rpb6 présente une homologie moindre avec la sous-unité ω de l'ARNP bactérienne [Minakhin *et al.*, 2001].

1-4.2 Composition et rôle des sous-unités

Cinq des sous-unités de l'ARNPII sont communes aux trois ARNP eucaryotes : Rpb 5, 6, 8, 10 et 12. De plus, plusieurs sous-unités sont homologues entre les ARNPI, II et III, y compris les sous-unités catalytiques, même si elles ne sont pas identiques.

1-4.2.1 Les sous-unités communes

Bien que les rôles des sous-unités communes soient peu étudiés, il est connu que Rpb5 lie la protéine virale HBx (Hepatitis B virus X protein), qui est une protéine activatrice de la transcription [Cheong *et al.*, 1995]. Rpb5 et Rpb9 interagissent avec Rpb1 et il semblerait que ces interactions aient un rôle dans le maintien de la structure de Rpb1 au niveau de la bulle de transcription [Zaros *et al.*, 2007]. Rpb6 est impliqué dans l'assemblage de l'ARNPII et stabilise le complexe de 12 sous-unités [Minakhin *et al.*, 2001]. Elle contribue à l'association du sous-complexe Rpb4-Rpb7 à l'ARNPII en interagissant avec Rpb7 [Tan *et al.*, 2003]. L'étude de Rpb8 a mis en évidence le fait que cette sous-unité est capable de lier l'ADN simple brin de manière non spécifique. [Kang *et al.*, 2006].

L'ARNPII présente une forte identité de séquences de la Levure à l'Homme, et cela au niveau de l'ensemble des sous-unités. Il a été démontré que six sous-unités de l'ARNPII humaine, Rpb5, 7, 8, 9, 10, 12, peuvent remplacer les sous-unités équivalentes chez la Levure [Mckune *et al.*, 1995]. De plus, l'étude de délétions géniques chez la Levure a mis en évidence le fait que les sous-unités Rpb4 et Rpb9 ne sont pas essentielles à la viabilité cellulaire.

1-4.2.2 Rpb9

Même si elle est non essentielle chez la Levure, Rpb9 est essentielle à la survie chez la Drosophile. Rpb9 joue un rôle important dans différentes étapes de la transcription. Ainsi, il a été montré que durant l'initiation de la transcription Rpb9 influence le choix du site +1 [Hull *et al.*, 1995; Sun *et al.*, 1996]. Le déplacement du +1 en amont du site « normal » serait dû à une interaction déficiente entre TFIIF et l'ARNPII sans Rpb9 [Ziegler *et al.*, 2003]. Rpb9 est important durant l'élongation car sa délétion de Rpb9 chez *S. cerevisiae* entraîne une sensibilité au 6-azauracile (6-AU) et lorsque cette délétion est couplée à une délétion du gène codant pour TFIIS cela produit un phénotype létal [Hemming *et al.*, 2000]. Il a également été démontré que Rpb9 stimule avec TFIIS le clivage de l'ARN au niveau des sites d'arrêts de l'ARNPII et donc la reprise de l'élongation [Awrey *et al.*, 1997]. Rpb9 est aussi impliqué au niveau du processus de TCR [Li et Smerdon, 2002]. Enfin, il a été démontré que la délétion de Rpb9 entraîne une diminution du taux de fidélité de l'ARNPII [Nesser *et al.*, 2006].

1-4.2.3 Le sous-complexe Rpb4/Rpb7

La sous-unité Rpb4 forme avec Rpb7 un sous-complexe dissociable du reste de l'ARNPII. Ce complexe est retrouvé en quantité sous-stœchiométrique dans les préparations purifiées d'ARNPII. De la même manière que Rpb9, même si Rpb4 n'est pas essentielle, elle a un rôle important dans la transcription. En effet, il a été montré que le complexe Rpb4/Rpb7 est nécessaire *in vitro* à la transcription dépendante du promoteur bien que l'ARNPII contenant seulement 10 sous-unités soit capable de transcrire une matrice d'ADN dépourvue de promoteur. De récentes études ont montré que le complexe Rpb4/Rpb7 reste lié à l'ARNPII durant l'intégralité de la transcription du gène [Runner *et al.*, 2008; Cojocar *et al.*, 2008]. Ces études invalident donc les hypothèses selon lesquelles le complexe Rpb4/Rpb7 ne serait lié à l'ARNPII qu'à certaines étapes de la transcription comme l'initiation.

La structure cristallographique de l'ARNPII complète a révélé que Rpb7 interagit avec Rpb1 et Rpb6. Le complexe Rpb4/Rpb7 est localisé près de la sortie de l'ARNm en

cours de synthèse et du CTD. Rpb4 et Rpb7 interagissent avec plusieurs facteurs de transcription et pourraient donc recruter des protéines interagissant avec le CTD.

Rpb4 est aussi impliquée dans différentes étapes post-transcriptionnelles. En effet, il a été montré que Rpb4 est essentielle pour l'exportation des ARNm au cytoplasme en conditions de stress. Cette fonction de Rpb4 est indépendante de son rôle en transcription [Farago *et al.*, 2003]. De plus, il a été montré que Rpb4 est impliqué dans le processus de maturation des ARNm en 3'. Elle permet le couplage des étapes de transcription et de maturation des ARNm en liant les facteurs de maturation comme Rna14 et Rna15 [Runner *et al.*, 2008]. Enfin, Rpb4 et Rpb7 sont aussi impliquées dans la dégradation d'ARNm spécifiques [Lotan *et al.*, 2005; Lotan *et al.*, 2007].

Tableau 3 : Les sous-unités de l'ARNPII.

Sous-unité	Poids moléculaire en kDa	Essentielle	Caractéristiques
Rpb1	220	+	Catalytique, CTD
Rpb2	140	+	Catalytique
Rpb3	33	+	Assemblage
Rpb4	16	-	Sous-complexe avec Rpb7, Maturation ARNm en 3'
Rpb5	25	+	Commune
Rpb6	14,4	+	Commune, Assemblage
Rpb7	19	+	Sous-complexe avec Rpb4
Rpb8	17	+	Commune
Rpb9	14,5	-	Élongation, Choix du +1, TCR, Fidélité
Rpb10	7,6	+	Commune
Rpb11	13,3	+	Assemblage
Rpb12	7	+	Commune

1-4.3 Données cristallographiques

Au cours des dernières années, la publication de structures cristallographiques à haute résolution de l'ARNPII de levure a permis d'émettre de nouvelles hypothèses sur le fonctionnement de l'enzyme et sur le mécanisme transcriptionnel au niveau

moléculaire [Gnatt *et al.*, 2001; Cramer *et al.*, 2001; Bushnell et Kornberg, 2003; Bushnell *et al.*, 2002; Bushnell *et al.*, 2004; Westover *et al.*, 2004a; Westover *et al.*, 2004b; Kettenberger *et al.*, 2003; Armache *et al.*, 2003; Armache *et al.*, 2005]. Les meilleures résolutions ont été obtenues avec l'ARNPII à 10 sous-unités (déplétion du sous-complexe Rpb4/Rpb7) seule [Cramer *et al.*, 2001] ou en élongation sur l'ADN [Gnatt *et al.*, 2001; Westover *et al.*, 2004a]. La structure cristallographique de l'ARNPII à 12 sous-unités a également été obtenue par deux équipes malgré les difficultés liées à la sous-stœchiométrie du sous-complexe Rpb4/Rpb7 au sein des préparations d'ARNPII [Bushnell et Kornberg, 2003; Armache *et al.*, 2003; Armache *et al.*, 2005]. Cependant, la résolution de la structure de l'ARNPII à 12 sous-unités est plus faible que celle obtenue avec l'ARNPII à 10 sous-unités.

1-4.3.1 Le site actif

La structure globale de l'ARNPII est celle d'une mâchoire avec une pince mobile qui permet de maintenir les acides nucléiques au niveau du site actif (Figure 2). Les différentes structures cristallographiques ont permis d'identifier différents domaines de l'ARNPII qui ont été nommés en fonction de leur rôle présumé ou de leur localisation. Le sillon aussi nommé canal primaire est un domaine de Rpb1 chargé positivement qui forme le cœur de l'enzyme. Son entrée est bordée par les mâchoires inférieure (Rpb5) et supérieure (Rpb1 et Rpb9). Ensuite, le sillon est entouré par la pince mobile (Rpb1) et le lobe (Rpb2). Enfin, il est fermé par un domaine de Rpb2 : le mur (« wall »). La sortie de l'ARNm se fait par un canal localisé entre le mur de Rpb2 et la pince de Rpb1. C'est au fond du sillon que le site catalytique de l'ARNPII est localisé. Le site actif est composé de résidus acides des sous-unités Rpb1 et Rpb2. Ces acides aminés permettent la coordination des deux ions métalliques Mg^{2+} (métaux A et B) essentiels à la catalyse. Les acides nucléiques prennent place dans le sillon où ils sont maintenus en place par les mâchoires supérieure et inférieure. Le pore, appelé également canal secondaire, s'ouvre dans le canal principal, près du site actif. Il a la forme d'un entonnoir inversé. Le pore permettrait la circulation des NTP entrant et des pyrophosphates rejetés. Il permettrait

aussi d'accommoder l'extrémité 3' de l'ARNm lorsque l'ARNPII recule sur la matrice d'ADN durant des arrêts de la transcription [Cramer *et al.*, 2001; Gnatt *et al.*, 2001].

Les cristallographies de l'ARNPII à 10 sous-unités ont révélé que l'enzyme a des conformations distinctes en présence ou non d'ADN. C'est la position de la pince qui définit ces conformations. Sans ADN, la pince est dans une conformation dite ouverte, car elle peut accueillir l'ADN dans le sillon. Alors qu'en élongation, la conformation de la pince est fermée. Dans un premier temps, il a été supposé que la pince était un élément mobile qui permettait de stabiliser le complexe durant l'élongation [Cramer *et al.*, 2001; Gnatt *et al.*, 2001]. Néanmoins, les structures de l'ARNPII avec ses 12 sous-unités ont entraîné une modification de ce modèle [Armache *et al.*, 2003; Bushnell et Kornberg, 2003]. En effet, en présence du sous-complexe Rpb4/Rpb7, la pince est en position fermée même en l'absence d'acides nucléiques ce qui rend l'entrée directe de l'ADN impossible. Deux hypothèses ont été formulées pour expliquer l'entrée de l'ADN. Selon la première hypothèse, le sous-complexe Rpb4/Rpb7 se dissocierait de l'ARNPII, la pince retrouverait une conformation ouverte pour permettre la liaison à l'ADN. Rpb4/Rpb7 se réassocierait à l'ARNPII après la formation de la bulle de transcription entraînant la fermeture de la pince. [Armache *et al.*, 2003]. La deuxième hypothèse suggère que l'ADN lierait d'abord la pince fermée [Armache *et al.*, 2003; Bushnell et Kornberg, 2003]. Puis après l'ouverture de la bulle de transcription par TFIIF, l'ADN « descendrait » dans le canal principal où il pourrait interagir avec les boucles « switch » 1, 2 et 3 [Armache *et al.*, 2003].

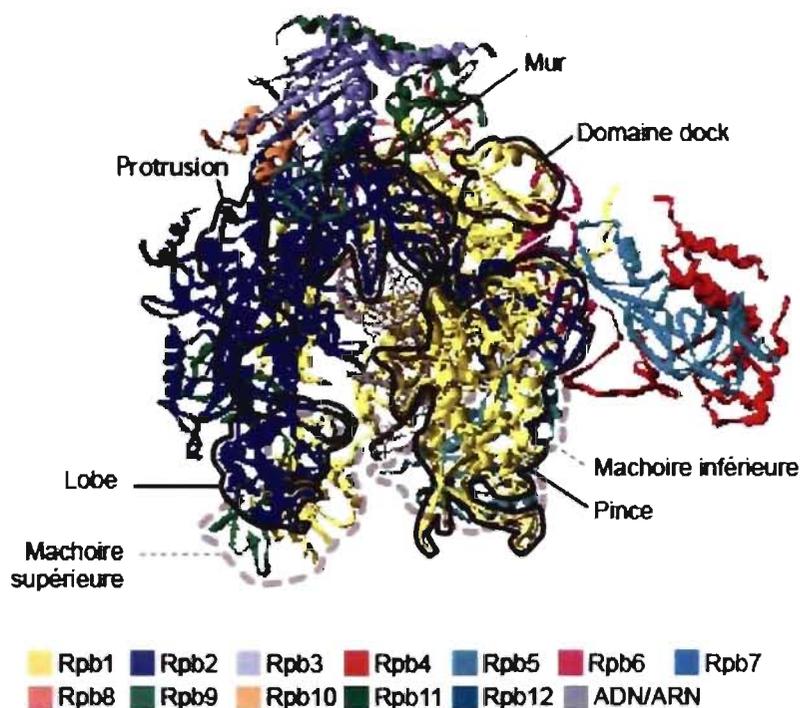


Figure 2 : Structure de l'ARNPII à 12 sous-unités. Les sous-unités sont colorées selon le code indiqué sous la figure. Figure modifiée à partir de Domecq *et al.*, 2008.

Les structures de l'ARNPII liée à une matrice d'ADN et à un ARNm ont révélé la présence d'un hybride ADN-ARN de 8 à 9 paires de bases dans le sillon [Gnatt *et al.*, 2001; Westover *et al.*, 2004a]. Ces structures ont aussi permis de définir des boucles et des hélices de Rpb1 et Rpb2, conservées entre les espèces eucaryotes, qui sont localisées près du site catalytique dans la région de l'hybride (Figure 3, Tableau 4). Ainsi, l'hélice de pontage est une hélice de Rpb1 qui traverse le sillon et qui se situe près du site +1. Elle a deux conformations distinctes : droite [Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Westover *et al.*, 2004a] ou courbe [Zhang *et al.*, 1999]. L'oscillation entre ces deux conformations suggère un rôle dans la translocation des acides nucléiques durant la polymérisation [Gnatt *et al.*, 2001]. La boucle « trigger » ou déclencheur est un élément mobile de Rpb1, situé juste sous l'hélice de pontage. Il a été proposé que le déclencheur contrôle les mouvements de l'hélice de pontage chez les procaryotes [Bar-Nahum *et al.*, 2005]. Ces résultats ont été confirmés chez *S. cerevisiae* par une étude cristallographique [Wang *et al.*, 2006]. Dernièrement, plusieurs études ont mis en évidence l'importance de cette boucle dans la catalyse de l'ARNm chez *E. coli* et *S. cerevisiae*. Les changements

de conformation du déclencheur permettraient la reconnaissance et la translocation des nucléotides au niveau du site actif [Wang *et al.*, 2006; Vassylyev *et al.*, 2007]. Chez les bactéries, il aurait également un rôle dans le mécanisme de pause de l'ARNP et dans la terminaison de la transcription [Epshtein *et al.*, 2007; Touloukhonov *et al.*, 2007].

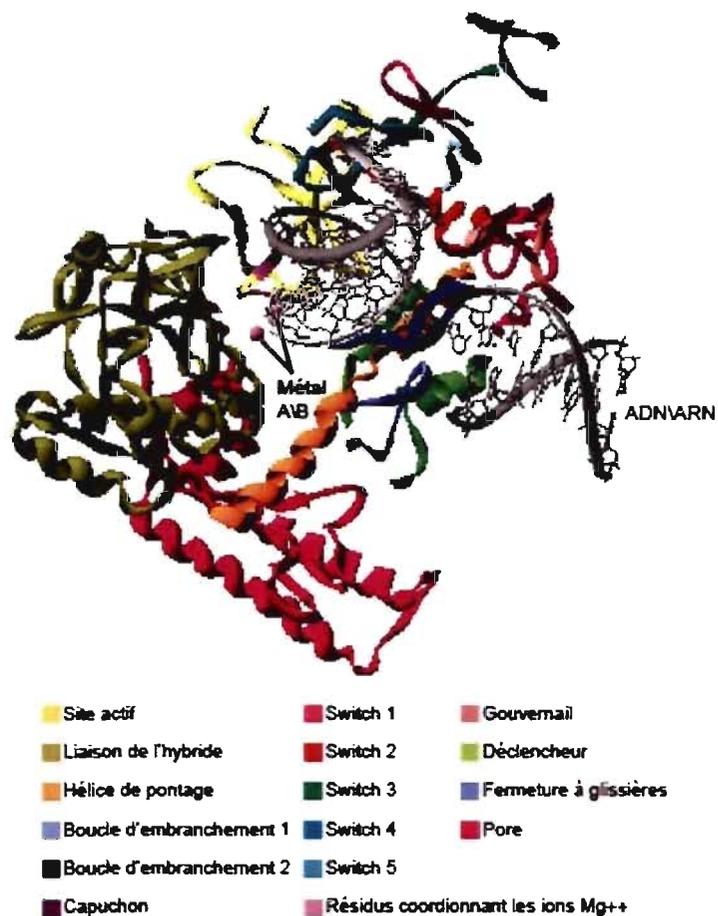


Figure 3 : Éléments structuraux du site actif de l'ARNPII. Les domaines sont colorés selon le code indiqué sous la figure. Figure modifiée à partir de Domecq *et al.*, 2008.

Les quatre boucles nommées boucle d'embranchement 1 (« fork loop 1 »), capuchon (« lid »), gouvernail (« rudder »), et fermeture à glissière (« zipper ») sont situées près de l'extrémité amont de l'hybride ADN/ARN. Leur position suggère qu'elles participent à la séparation de l'ARN de l'ADN afin de maintenir la longueur de l'hybride de 8 à 9 pb [Westover *et al.*, 2004a]. La boucle d'embranchement 1, qui interagit avec les bases -5 à -7 de l'ARN, empêcherait la formation d'un hybride étendu.

Le capuchon interagit avec l'ARN aux positions -8 à -10 et pourrait servir à guider celui-ci vers le canal de sortie. Le gouvernail contacte le brin d'ADN matrice aux positions -9 à -11 et pourrait ainsi empêcher sa réassociation à l'ARN. Néanmoins chez l'ARNP procaryote, une étude fonctionnelle du gouvernail a montré que cette boucle était impliquée dans l'initiation de la transcription en supportant l'ouverture de la bulle de transcription, ainsi que dans la stabilité du complexe d'élongation, et non dans le maintien de la longueur de l'hybride ADN-ARN [Kuznedelov *et al.*, 2002a]. La fermeture à glissière aurait des rôles de dissociation de l'hybride, sortie de l'ARNm et maintien de l'extrémité amont de la bulle. La boucle d'embranchement 2 (« fork loop 2 ») est localisée près de l'extrémité aval de la bulle de transcription et elle pourrait être impliquée dans le maintien de celle-ci. En interagissant avec l'ADN en + 3, elle aurait un rôle dans le désenroulement de l'ADN en aval [Wang *et al.*, 2006].

Cinq boucles de Rpb1 et Rpb2 situées près de la pince ont été appelées les boucles « switch » 1 à 5. Les boucles « switch » 1, 2 et 3 sont désordonnées dans le cristal de l'ARNPII seule [Cramer *et al.*, 2001], mais ordonnées dans celui de l'enzyme en élongation [Gnatt *et al.*, 2001]. Elles interagissent avec les acides nucléiques au niveau de l'hybride ADN-ARN. Comme cela a été mentionné ci-dessus, les boucles « switch » serviraient à coupler la fermeture de la pince à la présence d'un hybride ADN-ARN dans le canal principal [Gnatt *et al.*, 2001]. Une étude récente chez *S. cerevisiae* a mis en évidence le rôle de la boucle « switch » 2 dans la stabilisation de l'hybride durant initiation. Des mutations dans cet élément entraînent un accroissement du nombre d'initiation avortées et un déplacement en aval du site +1 [Majovski *et al.*, 2005].

La boucle « flap » est un domaine de Rpb2 situé près du canal de sortie de l'ARNm au niveau du domaine du mur. Chez l'ARNP procaryote, il a été montré que le « flap » interagit avec les ARN ayant une structure secondaire en épingle à cheveux [Toulokhonov *et al.*, 2001; Toulokhonov et Landick, 2003]. L'interaction du « flap » avec ces structures en épingle à cheveux permettent l'induction des pauses transcriptionnelles [Toulokhonov *et al.*, 2001; Toulokhonov et Landick, 2003 ; Mooney *et al.*, 1998]. Le « flap » a également un rôle dans l'initiation de la transcription

bactérienne [Kuznedelov *et al.*, 2002b]. Pour l'instant, il n'y a pas eu d'étude sur ce domaine chez l'ARNPII.

Tableau 4 : Les domaines du site catalytique de l'ARNPII

Sous-unité	Domaine	Fonction proposée
Rpb1	Hélice de pontage	Translocation des acides nucléiques
	Déclencheur	Translocation et reconnaissance des nucléotides
	Capuchon	Séparation de l'hybride ADN-ARN, Maintien de la bulle en aval
	Gouvernail	Séparation de l'hybride ADN-ARN
	Fermeture à glissière	Séparation de l'hybride ADN-ARN, Sortie de l'ARNm et Maintien de la bulle en amont
	Switch 1	Liaison du brin matrice d'ADN au niveau de l'hybride ADN-ARN, Maintien de la bulle en aval, Couplage fermeture de la pince à la présence d'un hybride ADN-ARN dans le sillon central
	Switch 2	
Switch 5		
Rpb2	Boucle d'embranchement 1	Séparation de l'hybride ADN-ARN
	Boucle d'embranchement 2	Maintien de la bulle en aval, Désenroulement de l'ADN
	Switch 3	Liaison du brin matrice d'ADN au niveau de l'hybride ADN-ARN, Maintien de la bulle en aval, Couplage fermeture de la pince à la présence d'un hybride ADN-ARN dans le sillon central
	Switch 4	
	Flap	Reconnaissance du promoteur et des pauses (procaryote)

1-4.3.2 Liaison des FGT

Des structures de l'ARNPII à 12 sous-unités liée à différents facteurs généraux de la transcription (TFIIB, TFIIF, TFIIS) ont été obtenues par cristallographie ou par microscopie chez *S. cerevisiae*. Ces données ont permis de révéler les sites de liaison de ces facteurs et de mieux comprendre leur mécanisme d'action.

Ainsi la structure cristallographique de l'ARNPII complexée au facteur TFIIB a montré que le domaine N-terminal de TFIIB se liait à l'ARNPII au niveau du domaine « dock » de Rpb1 [Bushnell *et al.*, 2004]. De plus, il apparaît qu'un élément, nommé « finger domain », au niveau N-terminal de TFIIB entre dans le site actif par le canal de sortie de l'ARNm en cours de synthèse. Ce domaine de TFIIB se retrouve alors près de l'hybride ADN/ARN ce qui pourrait expliquer les différents rôles de TFIIB dans l'initiation. Comme TFIIB est en contact à la fois avec le promoteur, via son domaine C-terminal, et l'hybride, cela permet d'expliquer son rôle dans le choix du site +1 et dans la stabilisation des complexes de préinitiation précoces. Le fait que TFIIB bloque la sortie de l'ARNm pourrait expliquer la raison pour laquelle le départ de TFIIB est nécessaire pour le passage à l'élongation. En effet, le blocage de la sortie de l'ARN pourrait entraîner une initiation avortée et le relâchement de TFIIB permettrait le dégagement du promoteur [Cramer, 2004].

La structure de l'ARNPII en complexe avec TFIIF a été réalisée par cryo-microscopie électronique [Chung *et al.*, 2003]. Elle a mis en évidence le fait que ces deux protéines entrent en contact sur une zone assez étendue. La plus grande des sous-unités de TFIIF (Tfg1, RAP74) lie l'ARNPII au niveau du sous-complexe Rpb4/Rpb7 et de la pince. La sous-unité Tfg2 (RAP30 chez l'Homme) se lie à l'ARNPII en plusieurs points de contact le long du sillon. L'interaction de Tfg2 avec l'ARNPII est très semblable à la liaison de la sous-unité σ de l'ARNP procaryote à l'holoenzyme bactérienne. La structure du complexe ARNPII/TFIIF confirme donc l'homologie de structure de σ et TFIIF.

La structure cristallographique de l'ARNPII libre liée à TFIIS a montré une zone étendue d'interactions entre TFIIS et la surface de la polymérase dans la région du pore [Kettenberg *et al.*, 2003]. Une boucle de TFIIS, contenant un ruban de zinc, entre dans le

pore et interagit avec le site actif de l'ARNPII. Des résidus très conservés de cette boucle sont positionnés près du site de liaison du deuxième ion métallique, ces résidus peuvent coordonner cet ion et une molécule d'eau ce qui stimulerait l'activité de clivage de l'ARN [Jeon *et al.*, 1994]. La liaison de TFIIS au site catalytique entraîne aussi un changement de conformation de l'enzyme qui pourrait favoriser le repositionnement des acides nucléiques et donc aider à sortir des sites d'arrêt via le clivage de l'ARN. Cette hypothèse a été confirmée par la structure cristallographique d'une ARNPII en élongation complexée avec TFIIS. En effet, l'ajout de TFIIS à un complexe en élongation a entraîné un réalignement de l'ARN dans le site actif.

1-4.4 Le CTD

Le domaine C-terminal de la sous-unité Rpb1 est spécifique à l'ARNPII. Il se compose de répétitions en tandem d'un heptapeptide : Y-S-P-T-S-P-S. Cette séquence est très conservée chez les eucaryotes, mais le nombre de répétitions varie selon les espèces. Ainsi, chez *S. cerevisiae*, les répétitions sont au nombre de 26, 42 chez la Drosophile et 52 chez l'Homme [Nonet *et al.*, 1987]. Le CTD n'est pas essentiel *in vitro* à la transcription dépendante du promoteur mais il est essentiel *in vivo*. En effet, il a été démontré que la délétion totale du CTD chez la Drosophile et la Levure était létale. Le CTD joue un rôle très important dans les différentes étapes de la transcription. En plus de son rôle dans la réaction transcriptionnelle, le CTD de Rpb1 est un site d'amarrage pour de nombreuses protéines impliquées dans les différentes étapes de la transcription et la maturation des ARNm. Le CTD permet ainsi le couplage de la transcription à la maturation des ARNm mais il apparaît également que le CTD permettrait de coupler la transcription des ARNm au remodelage de la chromatine, à la réparation de l'ADN, à l'édition des ARNm et à l'export nucléaire des ARNm [Reed, 2003; Sims *et al.*, 2004; Ares et Proudfoot, 2005]. L'état de phosphorylation du CTD de Rpb1 régule son activité. Le CTD contient cinq sites potentiels de phosphorylation ; cependant les sérines 2 et 5 sont les principaux résidus phosphorylés.

Le CTD libre est un long domaine flexible non structuré qui s'étend de la base de la pince à l'extérieur de l'ARNPII. Les données cristallographiques n'ont pas permis de

définir sa structure à cause de sa mobilité. Cependant, il a été montré que le CTD non-phosphorylé adopte une forme compactée et qu'il est situé à proximité de la surface de l'enzyme. Lorsque le CTD est phosphorylé (forme IIa), il adopte une conformation plus étendue à cause des interactions de charges.

L'état de phosphorylation du CTD de Rpb1 durant la transcription est assez bien défini. Ainsi, l'ARNPII ayant un CTD hypophosphorylé (IIo) est préférentiellement recrutée au promoteur pour former le complexe de pré-initiation. Suite à la formation du complexe de préinitiation, les sérines 5 sont phosphorylées par la sous-unité cdk7 de TFIIH durant l'initiation. Lors du passage de l'initiation à l'élongation, les sérines 2 du CTD sont phosphorylées par TFIIH et par la sous-unité cdk9 du facteur d'élongation pTEF-b chez l'Homme [Price *et al.*, 2000]. Chez la Levure, la kinase CTDK-1 (CTD kinase 1) phosphoryle le CTD sur les sérines 2 [Cho *et al.*, 2001]. Il est nécessaire que l'ARNPII soit déphosphorylé à la fin du processus transcriptionnel. La déphosphorylation va de nouveau permettre le recrutement au promoteur et un autre cycle de transcription. Fcp1 est la plus connue des phosphatases du CTD. Elle est conservée parmi les eucaryotes et elle est essentielle à la viabilité cellulaire [Archambault *et al.*, 1997; Archambault *et al.*, 1998]. Elle se lie à l'ARNPII via le sous-complexe Rpb4/Rpb7 et déphosphoryle le CTD préférentiellement au niveau des sérines 2 [Kamenski *et al.*, 2004]. Son activité est stimulée par la sous-unité RAP74 de TFIIF alors que TFIIB l'inhibe [Chambers *et al.*, 1995]. Dernièrement de nouvelles phosphatases du CTD ont été identifiées : SCP1 (Small CTD Phosphatase 1), Ssu72 et UBLCP1 (Ubiquitin-Like domain containing CTD Phosphatase 1). Ces trois phosphatases déphosphorylent le CTD au niveau des sérines 5 [Krishnamurthy *et al.*, 2004 ; Zheng *et al.*, 2005 ; Yeo *et al.*, 2003].

1-4.5 Mécanisme

1-4.5.1 Polymérisation de l'ARN

Un mécanisme catalytique commun à deux ions métalliques pour toutes les polymérases a été proposé [Steitz, 1998]. Par ce mécanisme, un premier ion Mg^{2+} , appelé métal A, facilite l'attaque nucléophile de l'oxygène en 3' du premier

nucléotide sur le phosphate α en 5' du nucléotide entrant. Le second ion métallique, appelé métal B, facilite la relâche du pyrophosphate. Chez l'ARNP procaryote et l'ARNPII, le métal A est coordonné par trois aspartates de β /Rpb1 strictement conservés contenus dans le motif NADFDGD [Cramer *et al.*, 2001; Gnatt *et al.*, 2001, Zhang *et al.*, 1999]. Le métal B semble avoir une faible affinité pour l'ARNP libre. Il entrerait dans le site actif avec le nucléotide entrant par le pore. Ainsi, il est coordonné par trois résidus acides (E836 et D837 de Rpb2 et D481 de Rpb1) et le nucléotide entrant le stabiliserait au niveau du site actif [Langelier *et al.*, 2005; Westover *et al.*, 2004b]. La formation de la liaison phosphodiester entraîne une translocation des nucléotides pour permettre l'addition d'un nouveau nucléotide et la relâche d'un pyrophosphate. Le métal B partirait avec le pyrophosphate par le pore.

1-4.5.2 Clivage de l'ARN

En plus, de son activité de polymérisation, l'ARNPII peut également catalyser, dans certaines circonstances, le clivage endonucléotidique en 3' des ARNm en cours de synthèse. En effet, pour sortir des sites d'arrêt et de pause, l'ARNPII doit cliver l'extrémité 3' du transcrit [Fish et Kane., 2002; Wind et Reines, 2000]. L'activité de clivage de l'enzyme est facilitée par TFIIS. La structure de TFIIS complexé à l'ARNPII a révélé que TFIIS pouvait coordonner le métal B dans le site actif ce qui laisse suggérer que l'activité de clivage est aussi un mécanisme à deux ions métalliques [Jeon *et al.*, 1994]. Des études similaires chez les procaryotes ont été réalisées avec un homologue de TFIIS, GreB, et ont menés aux mêmes conclusions [Opalka *et al.*, 2003; Sosunova *et al.*, 2003].

1-5 La réaction transcriptionnelle

La transcription se déroule en trois étapes principales : l'initiation, l'élongation et la terminaison. L'étape d'initiation de la transcription peut se diviser en plusieurs sous-

étapes : la formation du complexe de préinitiation, l'ouverture du complexe, les cycles d'initiations avortées et le dégagement du promoteur.

1-5.1 Initiation

1-5.1.1 Le complexe de préinitiation

Le complexe de préinitiation est composé des FGT et l'ARNPII qui s'assemblent au promoteur. L'assemblage du complexe de préinitiation a dans un premier temps été décrit comme séquentiel. Dans ce modèle, la protéine TBP du complexe TFIID lie le promoteur au niveau de la boîte TATA et induit une courbure de 80° de l'ADN [Kim *et al.*, 1993a, Kim *et al.*, 1993b]. Le facteur TFIIA vient stabiliser le complexe TFIID-ADN en liant TBP et l'ADN. La protéine TFIIIB s'ajoute au complexe déjà formé en liant TBP et l'ADN au niveau des BRE. TFIIIB facilite le recrutement de l'ARNPII. L'ARNPII entre dans le complexe de préinitiation complexée à TFIIIF. L'entrée du facteur TFIIIF dans le complexe entraîne la formation d'une deuxième courbure de l'ADN près du site +1 ce qui permet l'enroulement du promoteur autour de l'ARNPII [Forget *et al.*, 1997, Robert *et al.*, 1998, revu par Coulombe et Burton, 1999]. Enfin, à ce complexe s'ajoutent TFIIIE et TFIIH.

Cependant, le modèle séquentiel d'assemblage a été revu par la découverte chez la Levure et chez l'Homme de différents complexes préassemblés en solution [Myer *et al.*, 1998; Wilson *et al.*, 1996; Jeronimo *et al.*, 2004]. Même s'il est peu probable que l'assemblage du complexe de préinitiation complet se fasse en solution avant la liaison au promoteur chez les eucaryotes comme c'est le cas chez les procaryotes (holoenzyme), une alternative associant un assemblage en solution de certaines protéines et un assemblage partiellement séquentiel est envisageable.

1-5.1.2 Complexe ouvert, cycles d'initiations avortées, dégagement du promoteur

Lorsque le complexe de préinitiation est assemblé au promoteur, le facteur TFIIH permet l'ouverture de la bulle de transcription des positions -9 à +2. L'ouverture de la

double hélice d'ADN est réalisée par les activités hélicases de TFIIF grâce à l'énergie fournie par l'hydrolyse de l'ATP. Selon un modèle mis de l'avant par notre laboratoire, l'ouverture de la bulle de transcription est rendue plus aisée par le fait que l'enroulement serré de l'ADN autour de la polymérase entraîne une torsion de l'hélice d'ADN qui se retrouve ainsi « sous-enroulée » [Forget *et al.*, 1997; Robert *et al.*, 1998; revu par Coulombe et Burton, 1999].

Après l'ouverture de la bulle de transcription, la première liaison phosphodiester de l'ARNm en cours de synthèse initie la réaction transcriptionnelle. Lorsque l'ARN atteint une longueur de 4 nucléotides, le complexe ouvert change de conformation ce qui le stabilise. Le complexe transcriptionnel entre alors dans un cycle d'initiations avortées. De courts transcrits de 2 à 10 nucléotides sont synthétisés puis relâchés [Holstege *et al.*, 1997]. Enfin, le complexe se dégage du promoteur et entre en phase d'élongation productive. La manière dont le complexe sort du cycle d'initiations avortées et se dégage du promoteur n'est pas connue avec certitude. Cependant, des études récentes chez les procaryotes ont permis de proposer un nouveau modèle : le « scrunching model » [Kapanidis *et al.*, 2006; Revyakin *et al.*, 2006]. Selon ce modèle, le cycle d'initiations avortées est dû au fait que l'ARNP reste fixée au promoteur alors qu'il n'y a pas de translocation de l'ADN. L'ARNP se trouve dans une conformation intermédiaire qui lui permet de compacter et dérouler l'ADN dans le site actif sans bouger du promoteur. L'absence de translocation de l'ADN permet de réaliser une économie d'énergie, cette énergie serait accumulée et permettrait à l'ARNP de se dégager du promoteur.

1-5.2 Élongation

À la suite du dégagement du promoteur, l'ARNPII entre en phase d'élongation productive. À cette étape, les facteurs généraux de la transcription impliqués dans l'initiation ne sont plus liés à l'enzyme excepté TFIIF qui est aussi un facteur d'élongation. D'autres facteurs d'élongation vont se lier à l'ARNPII ainsi que les facteurs impliqués dans la maturation des ARNm et les facteurs de remodelage de la chromatine [Sims *et al.*, 2004]. Quelques facteurs d'élongation ainsi que leur fonction sont décrits dans le tableau suivant.

Tableau 5 : Les facteurs d'élongation de l'ARNPII et leurs rôles.

Facteurs d'élongation	Fonction
TFIIF	Stimulation de l'efficacité de l'ARNPII, Modulation de l'activité de TFIIIS, Diminution des pauses
TFIIS	Stimulation de l'activité de clivage de l'ARNPII, Sortie des sites d'arrêt
p-TEFb	Phosphorylation de la sérine 2 du CTD, Sortie des pauses induites par DSIF/NELF
Complexe elongin	Stimulation de l'efficacité de l'ARNPII, Diminution des pauses
ELL	Stimulation de l'efficacité de l'ARNPII, Diminution des pauses
DSIF	Stimulation de l'élongation, Suppression des terminaisons précoces, Effet négatif sur l'élongation quand associé à NELF
NELF	Effet négatif sur l'élongation en favorisant les pauses
FCP1	Stimulation de l'élongation, Déphosphorylation du CTD et recyclage de l'ARNPII
FACT	Facilite l'élongation en remodelant de la chromatine

1-5.3 Terminaison

La terminaison est l'ultime étape de la transcription par l'ARNPII, son mécanisme est peu connu. À cette étape, le transcrit et la matrice d'ADN sont relâchés par l'enzyme. La terminaison est couplée à la maturation en 3' de l'ARNm. Le signal de polyadénylation est essentiel à la terminaison, l'efficacité de la terminaison dépend de sa force. La polyadénylation du transcrit permet le recrutement du complexe de clivage et polyadénylation qui comprend trois facteurs : le facteur de clivage et de polyadénylation CPF et les facteurs de clivage CFIA et CFIB. L'ARNPII synthétise l'ensemble des ARNm mais aussi les ARNsn et les ARNsno. Des études récentes ont mis en évidence

que la terminaison des ARNsno se fait par un autre mécanisme utilisant le complexe protéique Nrd1p. Il apparaît que l'ARNPII interagit avec ces deux complexes de terminaison et que l'enzyme peut choisir entre ces deux modes pour la terminaison de la transcription [Bucheli *et al.*, 2007; Lykke-Andersen *et al.*, 2007].

1-6 Objectif et hypothèses

L'objectif général de notre laboratoire est d'analyser l'organisation moléculaire de l'ARNPII afin de mieux comprendre le processus transcriptionnel. L'objectif de ma maîtrise était d'étudier le mécanisme catalytique de l'ARNPII en fonction de sa structure. Le site catalytique de l'enzyme, composé des deux plus grosses sous-unités de la polymérase, est hautement conservé à travers les espèces. Les structures cristallographiques à haute résolution de l'ARNPII obtenues chez *S. cerevisiae* ont permis de définir l'importance de certains domaines de l'enzyme dans le mécanisme transcriptionnel. À partir de ces données, au sein du laboratoire une première étude de mutagenèse dirigée de l'ARNPII en cellules humaines a été réalisée.

Dans un premier temps, mon projet de maîtrise a consisté à poursuivre cette étude. Pour ce faire, grâce à un système mis au point au laboratoire, des cellules humaines exprimant des sous-unités Rpb1 sauvage ou mutantes portant une étiquette TAP ont été cultivées pour permettre la purification d'ARNPII sauvage ou mutantes assemblées *in vivo*. Les enzymes purifiées par la méthode TAP ont ensuite été testées dans différents essais biochimiques fonctionnels *in vitro*. Ce travail nous a permis de mieux comprendre le rôle de plusieurs domaines de Rpb1 dans le processus de transcription, notamment au niveau des boucles « switch » 1, « switch » 2, et « zipper ».

Dans un second temps, mon projet a consisté à caractériser l'ARNPII chez la levure, *Saccharomyces cerevisiae*. Nous avons choisi d'utiliser un nouvel organisme car le choix des mutations les plus informatives s'avérait difficile à faire avec des cellules humaines. *S. cerevisiae* est un organisme de choix pour ce projet car il nous a permis de créer et de caractériser facilement et rapidement un grand nombre de mutants. Nous avons choisi de caractériser les autres éléments structuraux clés du site catalytique au

niveau de la sous-unité Rpb2. Pour ce faire, il a été nécessaire d'adapter nos protocoles de purification de protéines ainsi que nos expériences biochimiques. De nouveaux mutants de l'ARNPII ont donc été créés, purifiés et caractérisés. L'étude de l'ARNPII chez la levure nous a permis de mieux comprendre le rôle de la boucle d'embranchement 2 et les acides aminés du site actif même.

2 MATÉRIEL ET MÉTHODES

2-1 La méthode TAP (Tandem Affinity Purification)

Cette méthode permet de purifier des complexes protéiques en conditions natives. Pour cela, il est nécessaire de créer une protéine de fusion contenant la protéine d'intérêt et l'étiquette TAP et de l'introduire dans une cellule hôte humaine ou de levure. La protéine de fusion, une fois exprimée, pourra former des complexes *in vivo* avec ses partenaires. Ces complexes pourront être purifiés par la méthode TAP [Rigaut *et al.*, 1991]. Dans le cadre de mon projet de recherche, la méthode de purification TAP a été utilisée pour purifier l'ARNPII qui, comme cela a été décrit ci-dessus, est une protéine multimérique à haut poids moléculaire.

La méthode peut être divisée en quatre étapes principales aussi bien dans les cellules humaines que dans les cellules de levure :

- (1) La construction d'un plasmide contenant la protéine d'intérêt fusionnée à l'étiquette TAP.
- (2) L'introduction dans les cellules du plasmide contenant l'ADNc d'intérêt.
- (3) La purification de l'ARNPII par double affinité.
- (4) L'identification de l'ARNPII sur gel SDS-PAGE.

2-1.1 Cellules humaines

2-1.1.1 Construction du plasmide contenant le gène d'intérêt fusionné à l'étiquette TAP

Les clones contenant l'ADNc d'intérêt proviennent de la compagnie Invitrogen. L'ADNc d'intérêt a ensuite été sous-cloné dans le vecteur pMZI. Ce vecteur contient des éléments de réponse à l'ecdysone modifié (EcRE) ainsi que l'étiquette TAP en C-terminal. Toutes les étapes de construction des plasmides contenant les gènes des sous-unités Rpb1 et Rpb2 (mutées ou non) fusionnées à l'étiquette TAP ont été réalisées au sein du laboratoire par Marie-France Langelier et Dania Baali.

2-1.1.2 Transfection stable des cellules avec l'ADNc d'intérêt.

Les cellules humaines transfectées sont les cellules EcR-293. Le plasmide pMZI contenant la protéine de fusion est inséré de manière aléatoire dans le génome. La transfection stable de ces cellules a été réalisée par la méthode utilisant du phosphate de calcium. Cette étape ainsi que la sélection des clones exprimant la protéine d'intérêt ont été réalisées par Marie-France Langelier et Dania Baali. Les cellules EcR-293 ont été choisies car elles permettent d'induire de manière spécifique l'expression des protéines d'intérêt. Les cellules EcR-293 contiennent un plasmide codant le récepteur de l'ecdysone (VgEcR dérivé de la *Drosophila* et modifié pour contenir le domaine de transactivation VP16) et le récepteur X rétinoïde (RXR dérivé des cellules mammifères). En présence d'ecdysone ou d'un analogue (Ponastérone A), un hétérodimère VgEcR-RXR se forme et va se lier aux EcRE, contenus dans le vecteur pMZI, ce qui entraîne l'expression de la protéine d'intérêt.

2-1.1.3 Purification par double affinité

L'étiquette TAP est composée d'un double site de liaison pour les immunoglobulines G (IgG) de la protéine A de *Staphylococcus aureus*, d'un site de clivage pour la protéase TEV (Tobacco Etch Virus) et d'un peptide de liaison à la calmoduline (CBP) (Figure 4).

pH 7,9; 1,5mM MgCl₂; 0,5mM DTT; 0,5mM AEBSF; 1,26M KAc; 37,5% glycérol) est ajouté au mélange qui est homogénéisé 10 fois.

Après une ultracentrifugation pour enlever les membranes et les précipités cellulaires, l'extrait de protéines est dialysé pour la nuit à 4°C dans le tampon de dialyse 100G (10mM Hépès pH 7,9; 0.1mM EDTA pH 8,0; 0.1mM DTT; 0.1mM KAc, 10% de glycérol).

2-1.1.3.2 Méthode TAP

Pour purifier l'ARNPII, la méthode TAP est réalisée à une concentration de 500mM en sels. En purifiant avec une telle quantité de sels, la préparation ne contient plus de FGT (Figure 6) comme c'est le cas avec une purification à 100mM en sels.

2-1.1.3.2.1 La colonne d'IgG

L'extrait protéique est mis en présence des billes d'IgG (50μL/g de culot cellulaire) qui ont été équilibrées au préalable avec le tampon IPP (10mM Tris pH8; 500mM NaCl; 0,1% Triton X100; 10% glycérol). Les billes sont ensuite lavées 4 fois sur une colonne avec le tampon IPP et une fois avec le tampon TEV (10mM Tris pH8; 500mM NaCl; 0,1% Triton X100; 0,5mM EDTA pH 8,0; 1mM DTT; 10% glycérol). Les billes sont traitées avec la protéase TEV (30 unités par gramme de culot cellulaire) dans le tampon TEV durant 16 heures à 4°C. La protéase va permettre la libération et l'élution des protéines portant l'étiquette TAP et des protéines avec lesquelles elles sont liées.

2-1.1.3.2.2 La colonne de calmoduline

L'éluat de la première colonne d'affinité est récupéré et mis en présence des billes de calmoduline (50μL/g de culot cellulaire) qui ont été équilibrées au préalable avec le tampon CBB (10mM Tris pH8; 500mM NaCl; 1mM Imidazole; 1mM MgAc; 2mM CaCl₂; 10mM β-mercaptoéthanol; 0,1% Triton X100; 10% glycérol). Les billes

sont incubées 2 heures à 4°C en présence de CaCl₂ (2mM). Les billes sont ensuite lavées sur une colonne avec le tampon CBB. L'éluat est réalisé en deux fois (soit 5 volumes de la colonne) avec le tampon CEB (10mM Tris pH8; 500mM NaCl; 1mM Imidazole; 1mM MgAc; 2mM CaCl₂; 10mM β-mercaptoéthanol; 0,1% Triton X100; 10% glycérol).

2-1.1.4 Concentration et identification des complexes purifiés

L'éluat est concentré par dialyse à 4°C dans le tampon F (10mM Hépès pH 7,9; 100mM NaCl; 0,1mM EDTA pH 8,0; 20% PEG8000; 0,5mM DTT; 20% glycérol). La dialyse prend fin quand le volume de l'éluat atteint un volume de 100 à 200 μL.

Un aliquot de 10 μL d'éluat concentré est migré sur un gel SDS-PAGE à gradient 4-12% (NuPAGE, Invitrogen) pour confirmer la présence de l'ensemble des sous-unités de l'ARNPII. L'activité de l'enzyme est ensuite testée par transcription ou élongation *in vitro*.

2-1.2 Levures

2-1.2.1 Construction du plasmide contenant le gène d'intérêt fusionné à l'étiquette TAP

La création des plasmides contenant l'ADNc de la protéine d'intérêt fusionnée à l'étiquette TAP a été réalisée en plusieurs étapes (Figure 5).

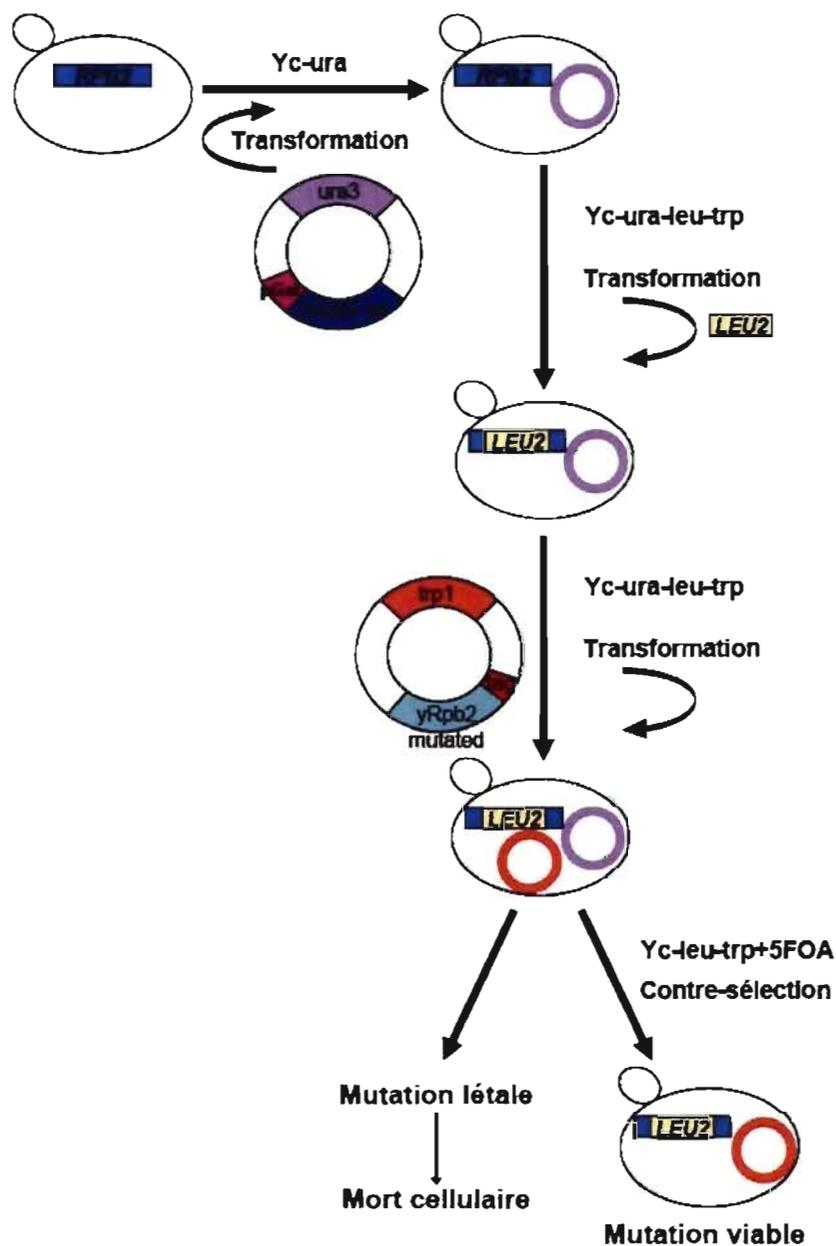


Figure 5 : Schéma résumant la construction des souches de levures.

2-1.2.1.1 PCR

L'ADNc de la sous-unité Rpb2, ainsi que son promoteur endogène, avait été incorporé dans un plasmide centromérique (*TRP1*, *CEN*, *ARS*), nommé pFL39, par Vincent Trinh et Marie-France Langelier. Pour ajouter l'étiquette TAP, une PCR inverse

de ce plasmide a été faite, ce qui a permis l'insertion du site de restriction de l'enzyme NotI en 3' du gène. La réaction de PCR contient 2,5 ng de matrice d'ADN; 0,5 μ M de chaque amorce; 200 μ M de dNTP; 0,02 U/ μ L de Phusion Hot Start (New England Biolabs) et le tampon Phusion GC (1X). Le programme comprend les étapes suivantes : 98°C / 1 min; 98°C / 10 sec, (T_m -5°C) / 30 sec, 68°C / (1min par kb), les trois dernières étapes sont répétées en boucle 35 fois et finalement 68°C / 10 min.

Parallèlement, une PCR de l'étiquette TAP est réalisée pour ajouter en 5' le site de restriction de l'enzyme NotI. La réaction de PCR contient 2 ng de matrice d'ADN; 0,35 μ M de chaque amorce; 150 μ M de dNTP; 0,05 U/ μ L de Pfu turbo (Stratagene) et le tampon Pfu (1X). Le programme comprend les étapes suivantes : 95°C / 2 min; 95°C / 30 sec, (T_m -5°C) / 30 sec, 72°C / (1min par kb), les trois dernières étapes sont répétées en boucle 30 fois et finalement 72°C / 10 min.

2-1.2.1.2 Digestion enzymatique

Les produits PCR sont migrés sur gel d'agarose 1%, purifiés avec le kit de purification de Qiagen et digérés par l'enzyme NotI (New England Biolabs) pendant 2 heures dans le tampon recommandé par le fabricant. Les produits de digestion sont migrés 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 260nm.

2-1.2.1.3 Ligation

L'introduction des sites de restriction de NotI en 5' de l'étiquette TAP et en 3' du gène Rpb2 au sein du vecteur pFL39 permet d'avoir des extrémités cohésives d'un côté, et de l'autre côté les extrémités sont franches et phosphorylées. Ainsi la ligation ne peut se faire que dans un sens. La ligation est réalisée à 14°C pendant 3 heures. Le volume réactionnel de 40 μ L comprend 3 unités de T4 DNA ligase (USB), 30 fmol d'ADN du vecteur (pFL39 contenant le gène d'intérêt), 180 fmol d'ADN de l'insert (étiquette TAP), 1 mM d'ATP et le tampon OPA 1X.

2-1.2.1.4 Transformation des bactéries compétentes

Les bactéries compétentes qui ont été utilisées sont des bactéries XL-1 rendus compétentes au laboratoire. Elles ont été transformées avec 10 μ L du produit de ligation pour 100 μ L de bactéries. Ce mélange est incubé 30 min sur glace, puis il subit un choc thermique de 90 sec à 42°C et après il repose 2 min sur glace. À la réaction sont ajoutés 900 μ L de milieu LB, puis le mélange est mis à incuber 1 heure à 37°C. L'ensemble des bactéries est étalé sur un milieu sélectif contenant 100 μ g/mL d'ampicilline et incubé pour la nuit à 37°C. Les colonies bactériennes ayant poussées sont repiquées et les clones sont sélectionnés par digestion enzymatique après purification de leur ADN plasmidique avec un kit Qiagen. Les clones positifs sont ensuite séquencés et le plasmide est produit en grande quantité.

2-1.2.1.5 Mutagénèse

La mutagénèse dirigée du gène *RPB2* chez la levure a été réalisée par PCR inverse à partir de la copie sauvage du gène *RPB2* portant l'étiquette TAP en C-terminal clonée dans le vecteur pFL39. La réaction de PCR contient 2,5 ng de matrice d'ADN; 0,5 μ M de chaque amorce; 200 μ M de dNTP; 0,02 U/ μ L de Phusion Hot Start (New England Biolabs) et le tampon Phusion GC (1X). Le programme comprend les étapes suivantes : 98°C / 1 min; 98°C / 10 sec, (T_m -5°C) / 30 sec, 68°C / 10 min, les trois dernières étapes sont répétées en boucle 35 fois et finalement 68°C / 10 min. Le produit de la PCR est migré sur gel d'agarose 1% puis purifié avec un kit de purification Qiagen. La ligation est réalisée à 14°C pendant 3 heures dans un volume de 40 μ L dans les conditions suivantes : 3 unités de T4 DNA ligase (USB), 250 ng d'ADN, 1 mM d'ATP et le tampon OPA 1X. La transformation des bactéries est réalisée dans les mêmes conditions que celles décrites au paragraphe 1.2.1.4.

2-1.2.2 Transformation des levures

Les plasmides contenant les gènes d'intérêt TAP étiquetés sont introduits dans des souches de levure qui avaient été préalablement créées au laboratoire par Vincent Trinh. Ces souches W303-1A (*MATa can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 ade2-1*) contiennent le plasmide pYC2/CT (*URA3, CEN, ARS*) une copie du gène *RPB2* sauvage sous le contrôle du promoteur p*GAL*. De plus, dans ces souches, la copie génomique du gène *RPB2* a été inactivée en l'interrompant avec le gène *LEU2*.

À partir d'une culture de 10mL de la souche de levure à transformer, un millilitre de culture est centrifugé et resuspendu dans de l'eau distillée, cette étape est répétée 2 fois. Le culot obtenu est remis en suspension dans 50 μ L de TELiAc 1X (0,01M Tris pH 7,5; 1mM EDTA pH 8,0, 0,1M LiAc). À ce mélange sont ajoutés 1 μ g d'ADN transformant, 50 μ g d'ADN de sperme de saumon (préalablement bouilli) et 300 μ L de PEG LiAc 40% (40% PEG, 1X TELiAc). Le mélange est incubé 30 min à 30°C avec agitation puis 15 min à 42°C. Après une centrifugation de 15 sec, le mélange est resuspendu dans 100 μ L d'eau distillée. L'ensemble est étalé sur un milieu sélectif (milieu minimum YNB (Yeast Nitrogen Base) avec 2% agar, 2% glucose, 40 μ g/mL d'adénine et 20 μ g/mL d'histidine) puis incubé 3 jours à 30°C. Les clones positifs sont validés par PCR.

2-1.2.3 Échange de plasmide : « plasmid shuffling »

Le « plasmid shuffling » est une méthode qui permet par contre-sélection avec l'acide 5 fluoroorotique (5FOA) d'enlever les plasmides contenant comme marqueur de sélection le gène *URA3*. Le 5FOA est un composé toxique pour les levures qui contiennent le gène *URA3*; ainsi seules les levures ayant perdues le gène *URA3* survivent. Le plasmide pYC2/CT contient le gène *URA3*, ce qui a permis d'obtenir des souches contenant uniquement le gène *RPB2* muté ou non avec l'étiquette TAP.

Les pétris sont composés du milieu minimum sélectif adéquat pour les souches (YNB, 2% agar, 20 μ g/mL histidine, 40 μ g/mL adénine et 12 μ g/mL uracile) et de 1mg/mL de 5FOA.

2-1.2.4 Sélection sur 6-azauracile

L'étude des phénotypes en présence de la drogue 6-AU est faite sur des pétris composés du milieu minimum sélectif adéquat pour les souches (YNB, 2% agar, 2% glucose, 20 $\mu\text{g}/\text{mL}$ histidine et 40 $\mu\text{g}/\text{mL}$ adénine) et de 50 $\mu\text{g}/\text{mL}$ de 6-AU.

2-1.2.5 Purification par double affinité

2-1.2.5.1 Extraction des protéines cellulaires

L'extraction de l'ensemble des protéines cellulaires ne se fait pas de manière identique dans les cellules humaines et les levures. En effet, la résistance de la paroi cellulaire des levures nécessite une action mécanique plus importante. Les levures sont comme les cellules humaines lysées dans des conditions natives pour extraire l'ensemble des protéines cellulaires.

À température ambiante, un volume de tampon A (10mM Hépès pH 7,9; 1,5mM MgCl_2 ; 10mM KCl; 0,5mM DTT; 0,5mM AEBSF et 1 tablette de « Complete EDTA-free » (Roche)) est ajouté aux levures récoltées pour les resuspendre. Les étapes suivantes sont réalisées sur glace pour prévenir toute dégradation protéique. La suspension de levures est passée deux fois dans une « French-Press » à une pression de 1200 psi (8,27 MPa).

Après la lyse mécanique, la concentration en chlorure de potassium de l'extrait est ajustée à 200mM. Après deux ultracentrifugations, l'extrait de protéines est dialysé durant 3 heures à 4°C dans le tampon de dialyse D (20mM Hépès pH 7,9; 50mM KCl; 0,2mM EDTA pH 8,0; 0,5mM DTT; 20% de glycérol). L'extrait peut alors être congelé dans de l'azote liquide puis stocké à -80°C.

2-1.2.5.2 Méthode TAP et identification des complexes

Les étapes de purification sur les colonnes d'affinité sont les mêmes pour les extraits de cellules humaines et les extraits de levure.

2-2 Création des souches *RPB2*-TAP et KO *DST1*

Pour étudier l'impact des mutations du gène *RBP2* sur l'élongation, des souches de levures contenant le gène *DST1* (qui code pour la protéine TFIIS) interrompu par la cassette Kanamycine ainsi que les mutations de *RBP2* ont été créées. Pour cela, au sein du laboratoire, Marie-France Langelier et Vincent Trinh avaient auparavant créé une souche de levure W303-1A (*MATa can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 ade2-1*) *RPB2* :: *LEU2* et *DST1* :: *kanMX* ainsi que le plasmide p*GAL-RPB2*. Cette souche a été transformée avec les plasmides pFL39 contenant une copie sauvage ou mutante du gène *RBP2* avec l'étiquette TAP. La méthode de transformation utilisée est la même que celle détaillée au paragraphe 2-1.2.2.

2-3 Essai de transcription *in vitro*

Les essais de transcription *in vitro* ont été réalisés pour vérifier la capacité de transcrire des ARNPII humaines purifiées par TAP-tag. Ces essais sont réalisés avec un système reconstitué de protéines purifiées.

L'assemblage des protéines se fait sur glace avec TBP (30ng), TFIIB (30ng), RAP30 (30ng), RAP74 (65ng), TFIIE34 (40ng), TFIIE56 (60ng), TFIIH (4ng) et 165ng d'ARN polymérase II de fœtus de veau ou humaine portant l'étiquette TAP. La purification des facteurs généraux de la transcription a été réalisée telle que précédemment décrite [Robert *et al.*, 1996; Robert *et al.*, 1998]. La réaction de transcription se fait dans un volume de 20µL dans une solution tampon contenant 24mM d'Hépès pH 7,9; 100mM KCl; 20µM EDTA pH 8,0; 20µM EGTA pH8,0; 20% glycérol; 2mM DTT et H₂O DEPC. Ce mélange de protéines est incubé avec 500ng de matrice d'ADN linéarisée, contenant le promoteur AdML de -50 à +10 pb, et fusionné à une cassette sans guanine [Sawadogo & Roeder, 1985]. La réaction est initiée avec l'addition de 750µM d'ATP, 750µM de CTP, 10µM d'UTP et 2.5µCi de [α -32P]UTP, puis incubée 30 minutes à 30°C.

L'arrêt de la réaction est réalisée par l'addition de 20 μ L de solution Stop (0.1M de NaAc pH5.0; 0.5% SDS; 2mM d'EDTA et 100 μ g/ml d'ARNt). La précipitation des ARN est réalisée avec du phénol acide, du chloroforme et d'alcool isoamyl (ratio 25 :24 :1) puis à l'aide d'éthanol. La réaction est migrée à 300V durant 1h30 sur un gel dénaturant de 4% de polyacrylamide 20 :1 contenant 8.3M d'urée. Le gel est ensuite séché 45 minutes à 80°C et révélé par autoradiographie sur Phospho-Imager et sur film Biomax MS (Kodak).

Dans ces conditions, la réaction de transcription *in vitro* permet la synthèse d'un transcrit de 391 nucléotides.

2-4 Essai d'élongation *in vitro*

Les essais d'élongation *in vitro* ont été menés dans le but de tester la capacité de transcription des ARNPII de levure qui ont été purifiées par TAP-tag.

La matrice d'ADN a été préparée en hybridant deux oligonucléotides de séquences aléatoires, dont l'un contient une extension de 15 CMP [Pan & Greenblatt, 1994]. Les brins transcrits et non-transcrits (ratio 1:1,1) ont été incubés pour permettre leur hybridation dans une solution tampon (25mM Tris pH 8,0; 50mM KCl, et 8mM MgCl₂) chauffée à 90°C pendant 10 minutes et refroidie par la suite à la température ambiante sur une période de 5 heures.

Pour l'essai d'élongation, 10ng de matrice ont été incubés avec 165ng d'ARN polymérase II de fœtus de veau ou de levure portant l'étiquette TAP. La réaction est initiée par l'addition de 750 μ M d'ATP, 750 μ M de CTP, 750 μ M de GTP, 7,5 μ M d'UTP et 6,25 μ Ci de [α -32P]-UTP et est incubée 30 minutes à 30°C. La réaction est ensuite arrêtée avec l'addition de 20 μ L de solution Stop. Les ARN sont ensuite précipités avec du phénol acide, du chloroforme et d'alcool isoamyl (ratio 25 :24 :1) puis à l'aide d'éthanol. La réaction est migrée à 400V durant 3h30 sur un gel dénaturant de 18% de polyacrylamide 40:2 contenant 9% d'urée. Le gel est ensuite révélé (sans séchage) par autoradiographie sur Phospho-Imager et sur film Biomax MS (Kodak).

Dans ces conditions, la réaction de transcription *in vitro* permet la synthèse d'un transcrit d'une soixantaine de nucléotides.

2-5 Essai de retard sur gel (EMSA)

La méthode de retard sur gel permet de visualiser sur gel l'interaction d'un complexe protéique à l'ADN. Cette méthode nous a permis d'étudier deux types d'interaction : (1) la liaison du complexe de pré-initiation au promoteur et (2) la capacité de l'ARNPII à lier non spécifiquement un ADN.

2-5.1 Synthèse des sondes radioactives

La préparation des sondes se fait sur glace. La première réaction consiste à digérer 1500ng d'un plasmide contenant le promoteur d'adénovirus majeur tardif (AdMLP), modifié ou non, avec 8-20 unités des enzymes de restriction DraI et BamHI (Invitrogen) pendant 2 heures à 37°C. Ensuite, la sonde d'ADN de 110 pb est complétée à l'aide de 5 unités d'ADN polymérase I Klenow avec 25 μ Ci de [α -³²P]dGTP et 3mM de dATP, dCTP et dTTP.

2-5.2 Assemblage

2-5.2.1 Au promoteur

La reconstitution du complexe de pré-initiation se fait sur glace avec les protéines purifiées TBP (80ng), TFIIB (80ng), RAP30 (300ng), RAP74 (600ng), TFIIE34 (160ng), TFIIE56 (380ng) et 150ng d'ARNPII de fœtus de veau, ou 200ng d'ARNPII humaine portant l'étiquette TAP. Ce mélange de protéines est incubé avec 3000cpm de la sonde marquée et d'oligonucléotide compétiteur poly(dI-dC) (2,5mg/mL) dans une solution de tampon G (12mM Hépès pH 7.9; 60mM KCl; 8mM MgCl₂; 0,12mM EDTA pH 8.0; 5mM β -mercaptoéthanol; 50ng/mL de BSA et 12% de glycérol)

durant 30 minutes à 30°C puis 5 minutes à 4°C. L'ajout du poly(dI-dC) permet de prévenir la formation de complexes non spécifiques.

2-5.2.2 ADN non spécifique

L'assemblage de l'ARNPII à un ADN non spécifique se fait sur la glace. À 3000 cpm de sonde marquée est ajouté 150ng d'ARNPII de fœtus de veau ou 200ng d'ARNPII humaine portant l'étiquette TAP dans une solution de tampon G. Le mélange incube 30 minutes à 30°C puis 5 minutes à 4°C. Dans ce type d'assemblage, l'ajout d'un compétiteur comme le poly(dI-dC) n'est pas nécessaire; aucun complexe non-spécifique ne peut se former du fait que seul l'ARNPII purifiée est ajoutée.

2-5.3 Gel de retardement

Les complexes assemblés sont déposés sur un gel refroidi de 4,5% de polyacrylamide 40 :1; 2mM MgCl₂; 0,5X TBE (445 mM Tris, 445 mM Borate, 10 mM EDTA pH 8,0). La migration est réalisée à 300V pendant une heure dans une solution tampon refroidie de 2mM MgCl₂; 0,5X TBE. Le gel est ensuite séché et révélé par autoradiographie sur Phospho-Imager et sur film Biomax MS (Kodak).

3 RÉSULTATS

3-1 Purification par double affinité de l'ARNPII

La purification par la méthode TAP a été adaptée au sein de notre laboratoire pour permettre l'isolation de complexes protéiques dans des cellules humaines. Le système de production utilisé dans les cellules humaines est inductible afin d'exprimer la protéine de fusion à un niveau physiologique.

De routine au laboratoire, la méthode TAP est réalisée à basse concentration de sels (100mM NaCl) pour garder les complexes purifiés dans des conditions natives. Nous avons modifié le protocole en augmentant la concentration de sels dans les tampons. La procédure pour la lyse des extraits cellulaires est toujours réalisée avec une faible concentration de sels pour conserver l'ensemble des protéines contenues dans les culots de cellules. Il en est de même pour l'étape de concentration du complexe par dialyse. La concentration de sels est augmentée dans les tampons de purification utilisés lors de la double purification sur colonne d'affinité et pour l'élution du complexe.

La purification de l'ARNPII à 100mM en sels entraîne une copurification de certains FGT tels TFIIB et TFIIF [Jeronimo *et al.*, 2004]. En purifiant avec une forte quantité de sels, la préparation ne contient plus de FGT qui pouvaient interférer dans certaines expériences *in vitro*. Ainsi, nos préparations d'ARNPII ne contiennent pas la sous-unité RAP30 de TFIIF (Figure 6) et nous avons pu également observé l'absence apparente des autres FGT (résultats non montrés). La purification à forte concentration en sels (500mM NaCl) nous permet d'avoir des préparations d'ARNPII plus pures. Comme il va l'être montré par la suite, l'utilisation de hautes concentrations de sels ne nuit ni à la composition de l'ARNPII (pas de perte de sous-unité) ni à la fonction de l'enzyme.

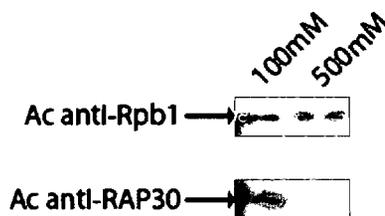


Figure 6 : Immunobuvardage de préparations d'ARNPII humaine sauvage purifiées avec des concentrations différentes de sels. Dans cette figure, les préparations utilisées ont été obtenues par purification à partir de cellules contenant la sous-unité Rpb2 étiquetée.

Par la suite, nous avons adapté la méthode TAP pour purifier le complexe de l'ARNPII chez les levures. De la même manière que pour l'ARNPII humaine, le protocole TAP est réalisé avec de fortes concentrations de sels afin d'obtenir des préparations d'enzyme les plus pures possibles.

3-2 Étude structure / fonction du site catalytique de l'ARNPII humaine

L'étude structure / fonction de l'ARNPII humaine vise à mieux caractériser le mécanisme moléculaire précis de l'enzyme à travers des essais fonctionnels *in vitro*. Cette étude a porté sur la sous-unité catalytique Rpb1. Pour ce faire, des lignées cellulaires ont été créées afin de purifier l'ARNPII.

Ainsi trois domaines de *RPB1* ont été mutés, la fermeture à glissière et les boucles « switch » 1 et 2. Le mutant Zip correspond à la délétion complète du domaine de la fermeture à glissière Trois acides aminés consécutifs de la boucle « switch » 1 (S1431, F1432 et E1433) ont été mutés en alanine, il s'agit du mutant Sw1-1431. Enfin, le mutant Sw2-348 correspond à une sextuple mutation dans la boucle « switch » 2 de six résidus (K346, E347, G348, R349, V350 et R351) en alanine.

Le choix des sites de mutations a été fait sur la base de l'homologie entre les séquences de plusieurs espèces (*E. coli*, *T. aquaticus*, *S. cerevisiae*, *D. melanogaster*, *C. elegans*, ...) et en fonction du rôle proposé de ces différents domaines dans la réaction de transcription (voir section 1-4.3.1). Ces acides aminés très conservés de la bactérie à l'Homme sont situés au cœur du site actif près des acides nucléiques et peuvent interagir

avec eux (Figure 7). Les mutations ont été réalisées dans le but d'étudier le rôle de ces résidus dans la réaction de catalyse de l'ARNPII, notamment durant l'initiation et l'élongation.

```

MHGGPPSGDSACPLRTIKRVQFGVLSPELKRMSVTEGGIKYPETTEGGRPKIGGLMDPRO
GVIERTGRCQTCAGNMTECPGHFVGHIELAKPVFHVGFVLRKTMKVLRCVCFCSKLLVDSSNP
KIKDILAKSRGQPKRRLTHVYDLCKGKNICEGGEEMDNKFGVEQPEGDEDLTKERGHGGCGR
YQPRIRRSGLELYAEWKHVNEDSQEKKILLSPERVHEIFKRISDEECFVLGMEPRYARPEWM
IVTVLPVPPPLSVRPAVVMQGSARNQDDLTHKLADIVKINNQLRRNEQNGAAAHVIAEDVKLL
QFHVATMVDNELPGLPRAMQKSGRPLKSLKQFLKKGKEGRVVRGNIMGRVDFEARTVITPDPN
LSIDQVGVPRSIAANMTFAEIVTFFNIDRLOELVRRNGSOYPGAKEYIIRDNGDRI DLRFHPK
PSDLHLQOTGYKVERHMCDDGDIVIFNRQPTLHKMSMMGHRVRI LPWSTFRLNLSVTPYNADF
DGDEMNLHL PQSLETRAEIQELAMVPRMIVTPQSNRPVMGIVQDTLTAVRKFKTRDVFLERG
EVMNLLMFLSTWDGKVPQPAI LKPRPLWTGKQIFSLIIPGHINCIRTHSTHPDDEDSGPYKH
ISPGDTKVVVENGELIMGILCKKSLGTSAGSLVHISYLEMGHDITRFLFYSNIQTVINNWLLI
EGHTIGIGDSIADSKTYODIQNTIKKAKQDVEVIEKAHNELEPTPGNTLRQTFENQVNRRI
LNDARDKTGSSAOKSLSEYNNFKSMVVS GAKGSKINISQVIAVVGQONVEGKRI PFGFKHRT
LPHFIKDDYGPESRGFVENSYLAGLTPTEFFHAMGGREGLIDTAVKTAETGYIQRRLIKSM
ESVMVKYDATVRNSINQVQLRYGEDGLAGESVEFQNLATLKPSNKA FEKKFRFDYTNERAL
RRTLOEDLVKDVLSNAHIQNELEREFERMREDREVLRFVIFPTGDSKVVLPCNLLRMIWNAOK
IFHINPRLPSDLHPIKVVVEGVKELSKKLVIVNGDDPLSRQAQENATLLEFNHILRSTLC SRRM
AEFFRLSGEAFDWLLGEIESKFNQAI AHPGEMVGALAAQSLGEPATQMTLNTFHYAGVSAKN
VTLGVPRKELINISKKPKTPSLTVFLLGQSARDAERA KDILCRLEHTLRKVTANTAIYYD
PNPQSTVVAEDQEWVNVYEMPFDVARIS PWLLRVELDRKHMTDRKLTMEQIAEKINAGFG
DDLNCIFNDDNAEKLVLRI RIMNSDENKMQEEEEVVDKMDDDVFLRCIESNMLTDMTLOGIE
QISKVYMHLPQTDNKKKI IITEDGEFKALQEWILETDGVSLMRVLSKDVDPVRTTSNDIVE
IFTVLGIEAVRKALERELYHVISFDGSYVNYRHLALLCDTMTCRGHLMAITRHGVNRQDTGP
LMKCSFEETVDVLMEEAAAHGESDPMKGVSENI MLGQLAPAGTGCDFLLDLDAEKCYGMEIPT
NIPGLGAAGPTGMFFGSAPS PMGGISPAMTPWNQGATPAYGAWSPSVGSGMTPGAAGFSPSA
ASDASGFSPGYSPAWSPTPGSPGSPGSSPYIPSPGGAMSPSYSPTS PAYEPRSPGGYTPQS
PSYSPTSPSYSPTSPSYSPTSPNYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPT
SPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPT
TSPSYSPTSPNYSPTSPNYTPTSPSYSPTSPSYSPTSPNYTPTSPNYSPTSPSYSPTSPSYS
PTSPSYSPPRYTPQSPTYTPSSPSYSPSSPSYSPTSPKYTPTSPSYSPSSPEYTPSTPKY
SPTSPKYSPTSPKYSPTSPYSPPTPKYSPTSPYSPVYTPSTPKYSPTSPYSPYTPSTPKY
YSPTSPYSPYSPKYSTYSPYSPYSPYSPYSPYSPYSPYSPYSPYSPYSPYSPYSPYSPYSPY

```

Fermeture à glissière :
 Délétion des 10 résidus
Boucle switch 1 :
 S1431, F1432, E1433 → A
Boucle switch 2 :
 K346, E347, G348, R349, V350, R351 → A

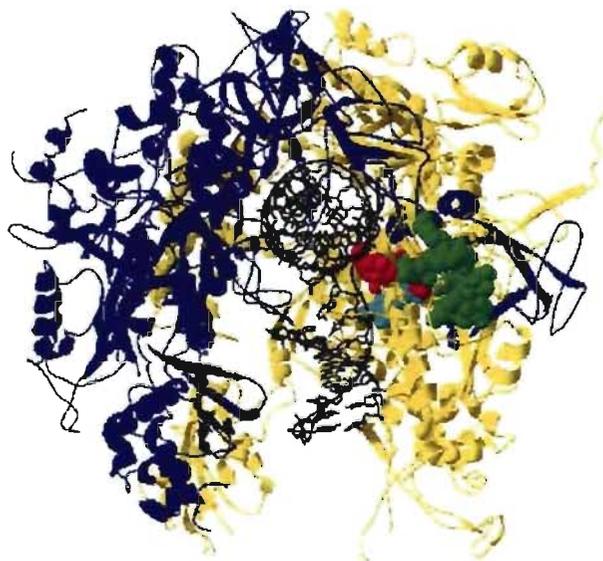


Figure 7 : Représentation schématique des mutations de la protéine Rpb1 humaine montrées relativement à la séquence et la structure. La séquence de la fermeture à

glissière est indiquée par un cadre vert, par un cadre bleu pour la boucle switch1 et par un cadre rose pour la boucle « switch » 2. Dans la structure cristallographique, les acides aminés mutés sont indiqués en vert pour la fermeture à glissière, en bleu pour la boucle « switch » 1 et en rose pour la boucle « switch » 2.

3-2.1 Purification des ARNPII humaines

La purification par la méthode TAP des ARNPII sauvage et mutantes a été réalisée comme décrite précédemment (section 3-1). Les éluats des enzymes purifiées sont déposés sur gel SDS-Page 4-12% pour visualiser l'ensemble des sous-unités de l'ARNPII en solution (Figure 8).

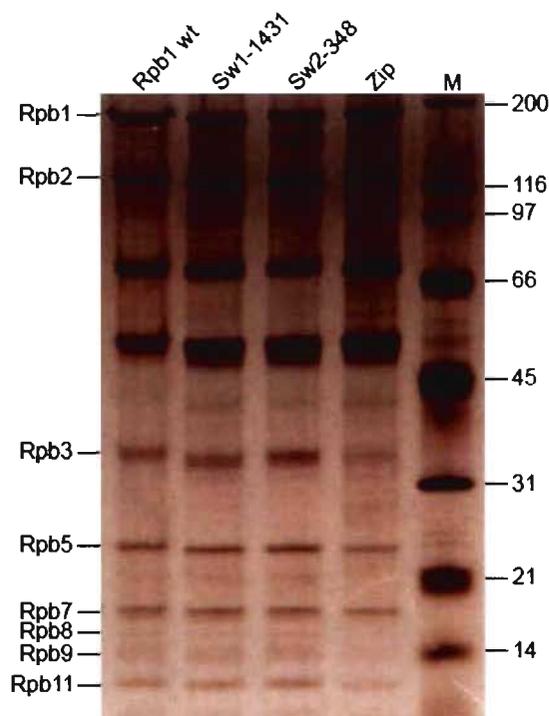


Figure 8 : Gel SDS-PAGE 4-12% des ARNPII humaines mutantes purifiées.

La Figure 8 montre que les ARNPII mutantes Sw1-1431 et Zip ont une composition en sous-unités similaire à l'enzyme sauvage. Ainsi ces deux mutations ne semblent pas affecter la capacité d'assemblage de l'ARNPII. Par contre, la Figure 8 montre également que le mutant Sw2-348 a une composition différente comparé à l'enzyme sauvage. En effet, Sw2-348 contient une bande supplémentaire entre les sous-

unités Rpb1 et Rpb2 (Figure 9). Cette bande peut correspondre à une protéine supplémentaire qui lie l'ARNPII mutante Sw2-348 ou à un produit de dégradation de Rpb1.

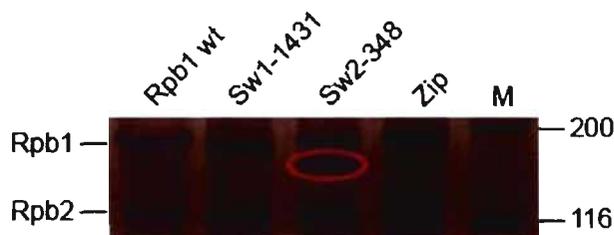


Figure 9 : Gel SDS-PAGE 4-12% des ARNPII humaines mutantes purifiées.

L'analyse par MS/MS de cette bande a révélé qu'il s'agissait de la sous-unité Rpb1 dont le domaine N-terminal était tronqué. Malheureusement, il n'est pas possible de séparer cette sous-unité tronquée du reste de la préparation. En effet, l'étiquette TAP est située à l'extrémité C-terminal de Rpb1, la sous-unité tronquée est donc retenue sur les colonnes de purification. De plus, nous ne pouvons pas savoir si cette sous-unité tronquée est capable de s'associer aux autres sous-unités pour former un complexe actif ou non. Nous ne savons également pas si la présence de cette sous-unité tronquée dans nos préparations peut affecter le comportement du mutant. Pour cette raison, ce mutant n'a pas été étudié plus à fond.

3-2.2 Caractérisation des ARNPII humaines mutées

Pour tester l'activité des ARNPII humaines purifiées, un essai de transcription *in vitro* a été réalisé. La figure 10 montre que les ARNPII mutantes n'ont aucune activité transcriptionnelle détectable *in vitro* dans notre système. D'autres essais de transcription ont été réalisés avec des quantités plus grandes d'ARNPII mutantes mais aucune activité transcriptionnelle n'a pu être mise en évidence (données non montrées).

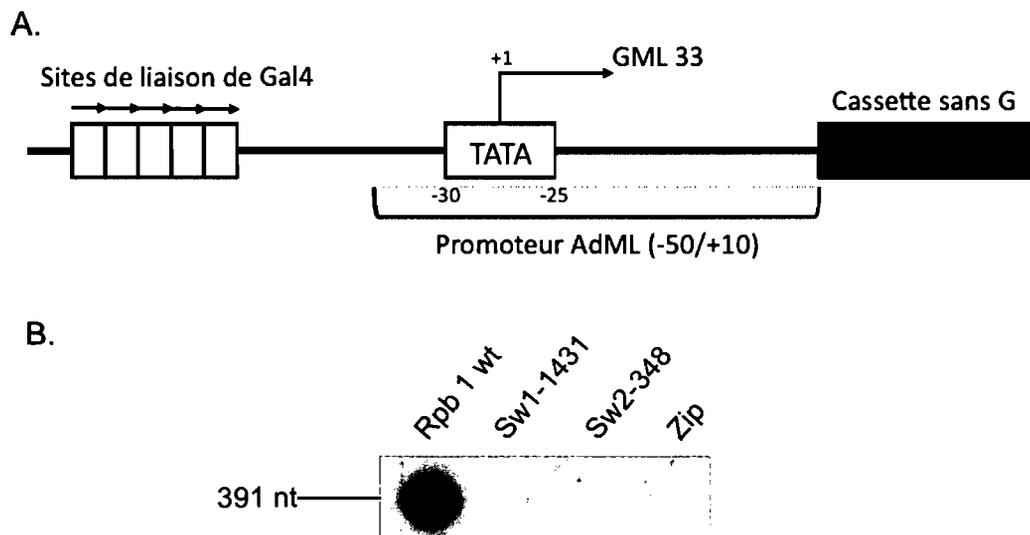


Figure 10 : Essai de transcription *in vitro*.

A. Représentation schématique de la matrice d'ADN utilisée pour les essais de transcription *in vitro*. À 37 pb en amont de la boîte TATA, sont situés cinq sites de liaison Gal4 répétés en tandem.

B. Réaction de transcription *in vitro* des ARNPII humaines sauvage et mutantes.

Comme aucune activité enzymatique n'a été mise en évidence avec les protéines mutantes, des essais *in vitro* ont été réalisés afin de tester la capacité des ARNPII mutantes de lier l'ADN ou d'entrer dans un complexe de pré-initiation de la transcription au promoteur. Pour cela, des essais de retard sur gel ont été menés dans différentes conditions.

Dans un premier temps, un essai de retard sur gel a été réalisé dans le but de tester la capacité de liaison de ces mutants à l'ADN. En effet, l'ARNPII est capable, seule, de se lier à l'ADN de manière non spécifique. Pour réaliser cet essai, les ARNPII sauvage et mutantes purifiées ont été incubées avec de l'ADN double brin marqué radioactivement ne contenant aucun élément du promoteur (séquence non spécifique).

La figure 11 montre que le mutant Zip ne lie pas l'ADN dans cet essai. Le mutant Sw1-1431 lie l'ADN manière comparable à l'ARNPII sauvage, alors que le mutant Sw2-348 lie très fortement l'ADN.

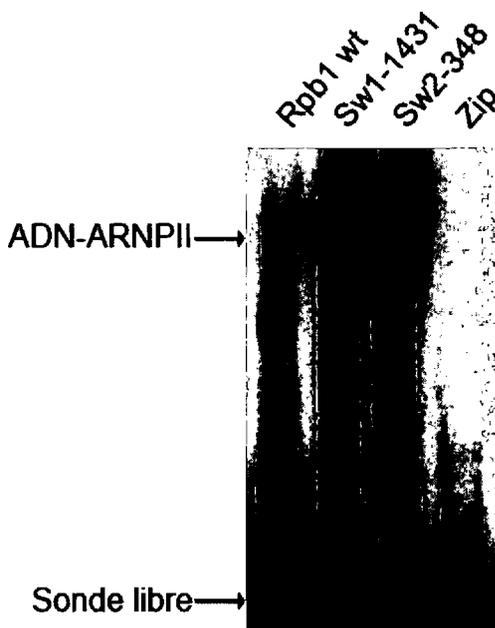


Figure 11 : Essai de retard sur gel de liaison non spécifique des RNAPII humaines à l'ADN.

Dans un second temps, un autre essai de retard sur gel a été réalisé avec cette fois un promoteur génique et les FGT (Figure 12). Dans cet essai, les contrôles négatifs sont réalisés en omettant du mélange réactionnel la protéine TBP. Cette protéine est essentielle à la formation du complexe de pré-initiation, ce qui permet de vérifier la spécificité de liaison au promoteur des complexes formés. Sur la Figure 12, deux types de complexes spécifiques sont présents. Le complexe situé le plus haut sur l'autoradiographie du gel correspond au complexe de pré-initiation. Il comprend l'ARNPII purifiée, les facteurs généraux de la transcription et le promoteur AdML. Le deuxième complexe situé plus bas dans le gel correspond à un complexe spécifique intermédiaire qui ne contient pas tous les éléments du complexe de pré-initiation.

Cet essai a permis de confirmer le fait que le mutant Zip ne possède pas d'activité détectable de liaison à l'ADN *in vitro* puisqu'il ne peut pas s'associer au complexe de pré-initiation. Le mutant Sw1-1431 présente une capacité d'association au complexe de pré-initiation semblable à celle de l'ARNPII sauvage. Plus surprenant, le mutant Sw2-348 agit également de manière similaire à l'enzyme sauvage.

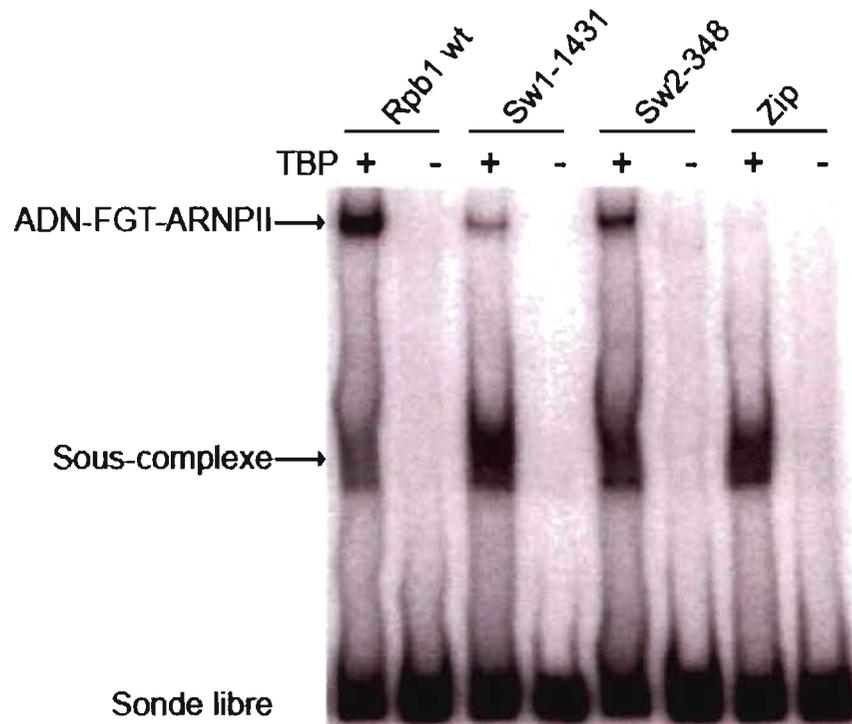


Figure 12 : Essai de retard sur gel des ARNPII humaines au promoteur.

3-3 Étude structure / fonction du site catalytique de l'ARNPII chez *S. cerevisiae*

L'étude structure-fonction de l'ARNPII chez la levure *S. cerevisiae* vise à mieux caractériser le fonctionnement moléculaire précis de l'enzyme à travers des essais fonctionnels *in vivo* et *in vitro*. Cette étude a porté sur la sous-unité catalytique Rpb2. Pour ce faire, des souches de levure ont été créées pour purifier l'ARNPII. La stratégie globale de la méthode employée est présentée dans la figure 13.

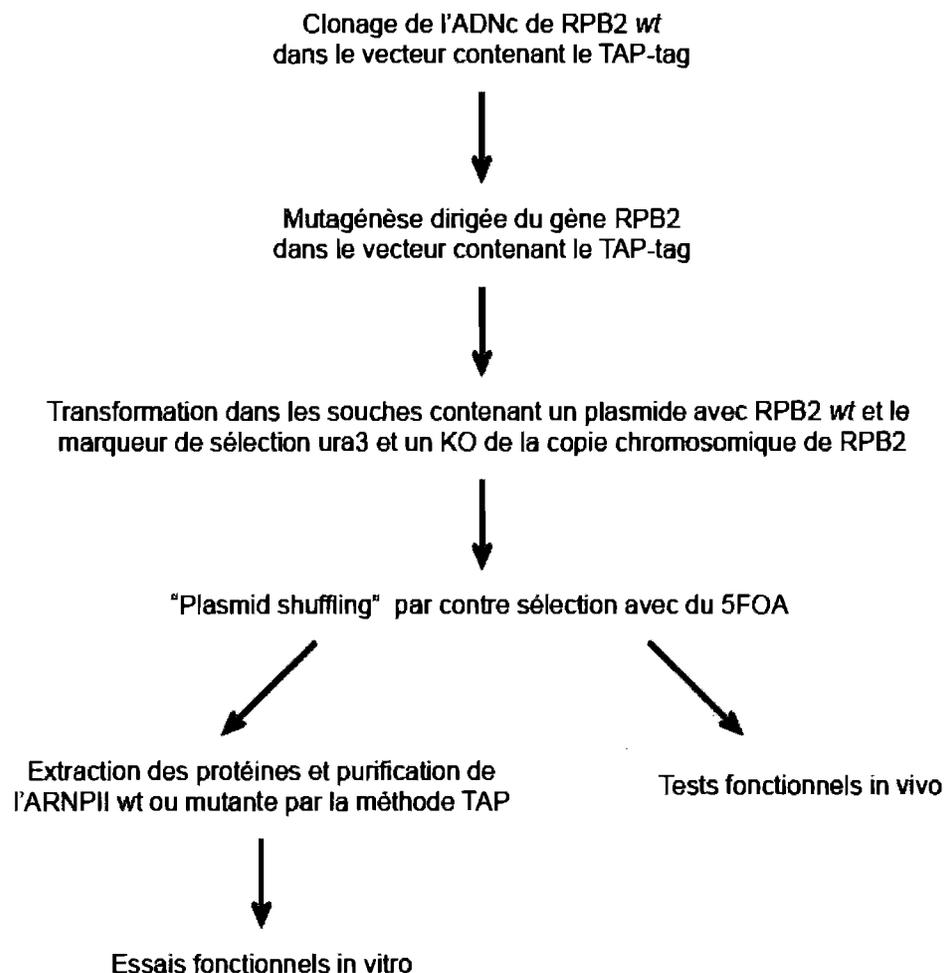


Figure 13 : Représentation schématique de la méthode ayant permis de générer les souches de levures et de purifier les ARNPII sauvage et mutantes chez la levure.

La figure 14 montre la représentation schématique des mutations qui ont été générées dans le gène *RPB2* au niveau de la séquence protéique primaire et de la structure cristallographique de la sous-unité Rpb2. La mutagenèse a été effectuée dans deux éléments du site catalytique de l'ARNPII : la boucle d'embranchement 2 et le site actif.

Ainsi cinq acides aminés, R504, D505, K510, R512 et Q513, ont été mutés en alanine au niveau de la boucle d'embranchement 2. Au niveau du site actif trois résidus, R983, R1020 et E1028, ont été modifiés en glutamine. Le choix des sites de mutations a été fait en considérant l'homologie avec la séquence d'*E. coli*. Ces acides aminés très conservés de la bactérie à la levure sont situés au cœur du site actif près des acides

nucléiques et peuvent interagir avec eux. Les mutations ont été réalisées dans le but d'étudier le rôle de ces résidus dans la réaction de catalyse de l'ARNPII, notamment durant l'élongation.

```
MSDLANSEKYYDEDPYGFEDSAPITAEDSWAVISAFFREKGLVSQQLDSFNQFVDYTLQ
DIICEDSTLILEQLAQHTTESDNI SRKYEISFGKIYVTKPMVNESDGVTHALYPQEARLR
NLTYSSGLFVDVKKRTYEAIDVPGRELKYELIAEESDDSES GKVF IGRLPIMLRSKNCY
LSEATESDLYKKECPFDMGYFIINGSEKVLIAQERSAGNIVQVFKKAAPSPI SHVAEI
RSALERGSRFISTLQVKLYGREGSSARTIKATLPYIKQDIP I V I I FRALGIIPDGEILEH
ICYDNDWQMLEMLKPCVEDGFVIQDRETALDFIGRRGTALGIKKEKRIQYAKDILQKEF
LPHITQLEGFESRKAFFLG YMINRLLLCALDRKDQDDR DHFGKRLDLAGPLLAQLFKTL
FKKLTKDIFRYMORTVEEAHDFNMKLA INAKTITSGLYALATGNWGEQKKAMSSRAGVS
QV LNRYTYSSTLSHLRRTNTPIGRDGKLAKPRQLHNTHWGLVCPAETPEGOACGLVKNLS
LMSCISVGTDPMP IITFLSEWGMEPLEDYVPHQSPDATRVFVNGVWHGVHRNPARLME TL
RTLRRKGDINPEVSMIRDIREREKELKIFTDAGR VYRPLFIVEDDES LGHKELKVRKGHIAK
LMATEYQDIEGGFEDVEEYTWSSLLNEGLVEYIDAEEEE SILIAMQPEDLEPAEANEEND
LDVDPAKRIRVSHHATTFTHCEIHPSMILGVAAS IIPFDHNQSPRNTYQSAMGQAMGV
FLTNYNVRMDTMANILYYPQKPLGTTRAMEYLKFRELPAQONAIVA IACYS GYNQEDSMI
MNQSSI DRGLFRSLFFRSYMDQEKY GMSITETF EK PQR TNL R MKHGT YDKLDDGLIA
PGVRVSGEDVIIGKTPISPDEEELGORTAYH SKRDASTPLRSTENGIVDQVLTNNDG
LRFVKVRVRTTKIPQIGDKFASRHGQKGTIGIT YRREDMPFTAEGIVPDLIINPHAI PSR
MTVAHLIECLLSKVAALSGNEG DASPFTDITVEGISKLLREHGYQSRGFVEMYNGHTGKK
LMAQIFFGPTY YQRLRHMVDDKI HARARGPMQVLTROPVEGRSRDGGGLRF GEMERDCMIA
HGAASFLKERLMEASDAFRVHICGICGLMTVIAKLNHNQFECKGCDNKID IYQIHIPYAA
KLLFQELMAMNITPRLYTDRSRDF-
```

Fork loop 2 :

R504, D505, K510, R512, Q513 → A

Active Site :

R983, R1020, E1028 → Q

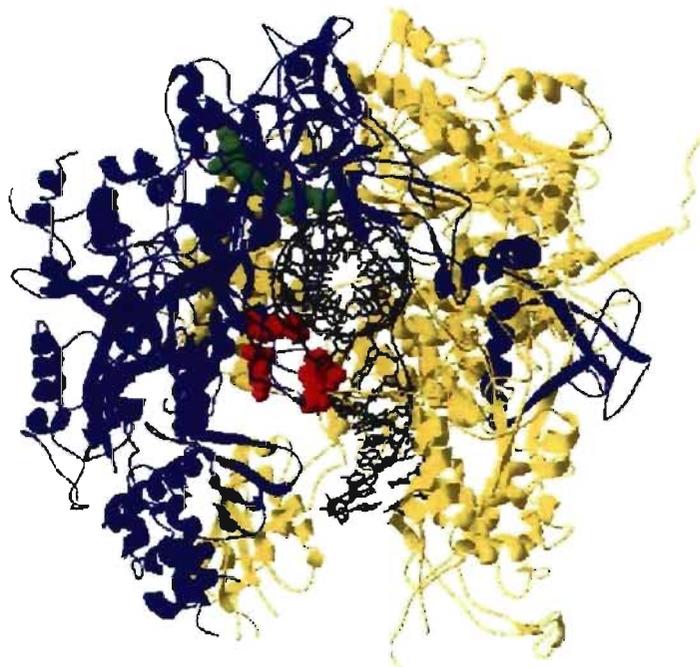


Figure 14 : Représentation schématique des mutations de la protéine Rpb2 de *S. cerevisiae* montrées relativement à la séquence et la structure. La séquence de la boucle

d'embranchement 2 est indiquée par un cadre rouge et par un cadre jaune pour le site actif. Dans la structure cristallographique, les acides aminés mutés sont indiqués en rouge pour la boucle d'embranchement 2 et en jaune pour le site actif.

3-3.1 Purification des ARNPII de levures

La purification par la méthode TAP des mutants non létaux, c'est à dire capables de pousser en présence de 5FOA, a été réalisée comme décrite au chapitre 2. Les éluats des enzymes purifiées sont déposés sur gel SDS-Page 4-12% pour visualiser l'ensemble des sous-unités de l'ARNPII purifiées (Figure 15). Aucune anomalie dans la composition des enzymes n'est révélée par ce gel.

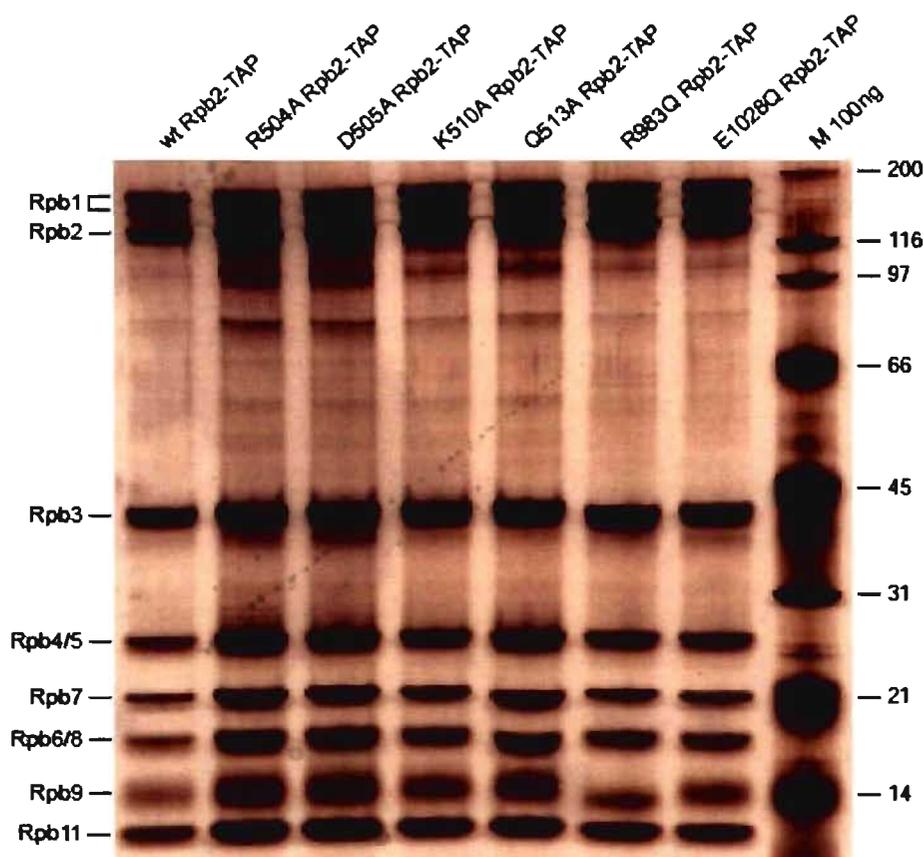


Figure 15 : Gel SDS-PAGE 4-12% des ARNPII mutantes purifiées chez la levure.

3-3.2 Caractérisation des ARNP_{II} mutées de levures

Les souches de levures contenant un KO pour la copie génomique du gène *RPB2*, un plasmide pYC2/CT avec le gène *RPB2* sauvage sous le contrôle du promoteur pGal et un plasmide pFL39 contenant le gène *RPB2* sauvage ou muté avec l'étiquette TAP en C-terminal sous le contrôle du promoteur endogène du gène *RPB2* ont été créées. La méthode de « plasmid shuffling » a été utilisée pour étudier l'effet de l'étiquette TAP, la létalité des mutations et pouvoir ensuite purifier l'ARNP_{II} par la méthode TAP.

Dans la figure 16, il apparaît que l'ajout de l'étiquette TAP en C-terminal n'affecte pas la croissance de la souche sauvage. Ainsi la fonction de la protéine *in vivo* n'est pas affectée par cet ajout. De plus, dans cette même figure, la croissance sur le milieu contenant la drogue 5FOA a mis en évidence le fait que les mutations R512A et R1020Q sont létales.

Le 6-azauracile (6-AU) est une drogue qui conduit à une réduction des taux intracellulaires de GTP et d'UTP. Ce composé n'est pas létal en soi, mais conjugué à des mutations de l'ARNP_{II}, il peut ralentir ou bloquer la croissance cellulaire des levures. La culture des souches sur 6-AU a mis en évidence une difficulté de croissance pour les mutants D505A et R983Q (Figure 16). Il semble que ces mutations ont un effet sur la capacité de polymérisation de l'ARNP_{II}.

Enfin les souches de levures dont les copies génomiques des gènes *RPB2* et *DST1* (codant pour la protéine TFIIS) sont inactivées, ont été transformées avec le plasmide pGAL-*RPB2* (*URA3*) et le plasmide pFL39 contenant le gène *RPB2* sauvage ou muté avec l'étiquette TAP en C-terminal sous le contrôle du promoteur endogène du gène *RPB2*. En faisant croître les levures en présence de glucose, seule la protéine Rpb2 ayant l'étiquette TAP est exprimée. En effet, le gène placé sous le contrôle du promoteur pGAL n'est transcrit qu'en présence de galactose. Le facteur TFIIS, comme il a été vu précédemment, est un facteur d'élongation important qui permet à l'ARNP_{II} en arrêt de reprendre son activité de polymérisation de l'ARN. En l'absence de cette protéine, les mutations de l'ARNP_{II} affectant la capacité d'élongation de l'enzyme peuvent être

mises en évidence. La culture de ces souches sur milieu minimum avec glucose a montré une croissance plus lente pour le mutant E1028Q.

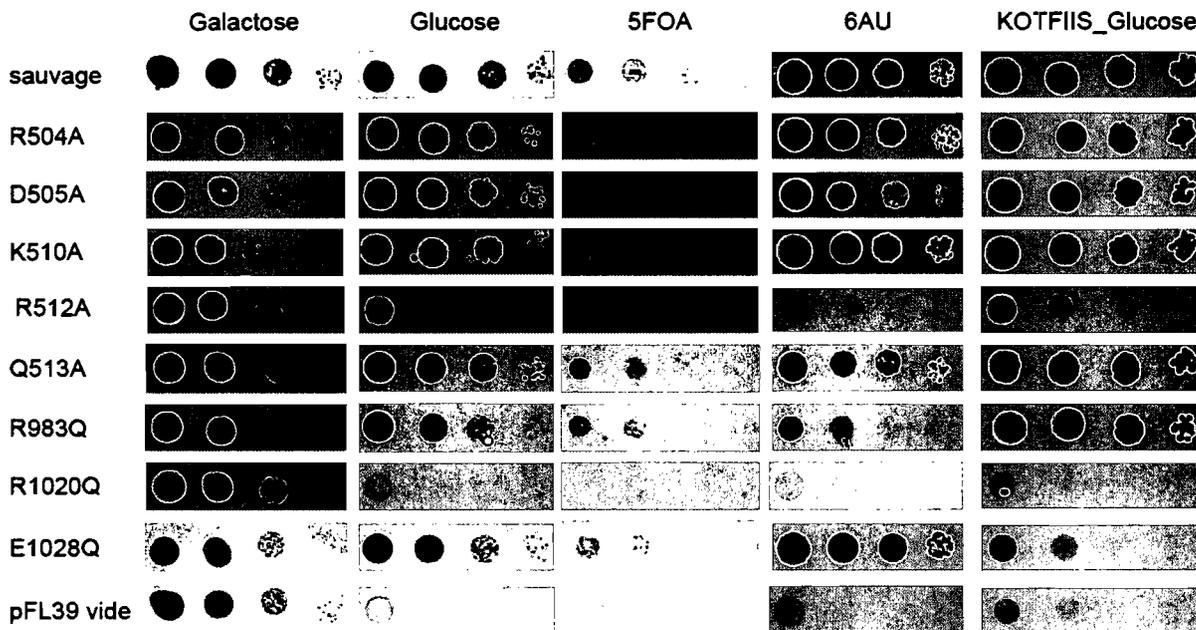


Figure 16 : Phénotypes de croissance des souches mutantes dans différentes conditions. Les tests de croissance ont été faits par dilutions en séries sur des pétris avec le milieu adéquat.

Pour tester la capacité de transcription des ARNPII purifiées, un essai d'élongation *in vitro* a été réalisé (Figure 17). Cette figure montre que les ARNPII mutantes fonctionnent qualitativement comme l'ARNPII sauvage; ces mutants sont donc fonctionnels dans cet essai.

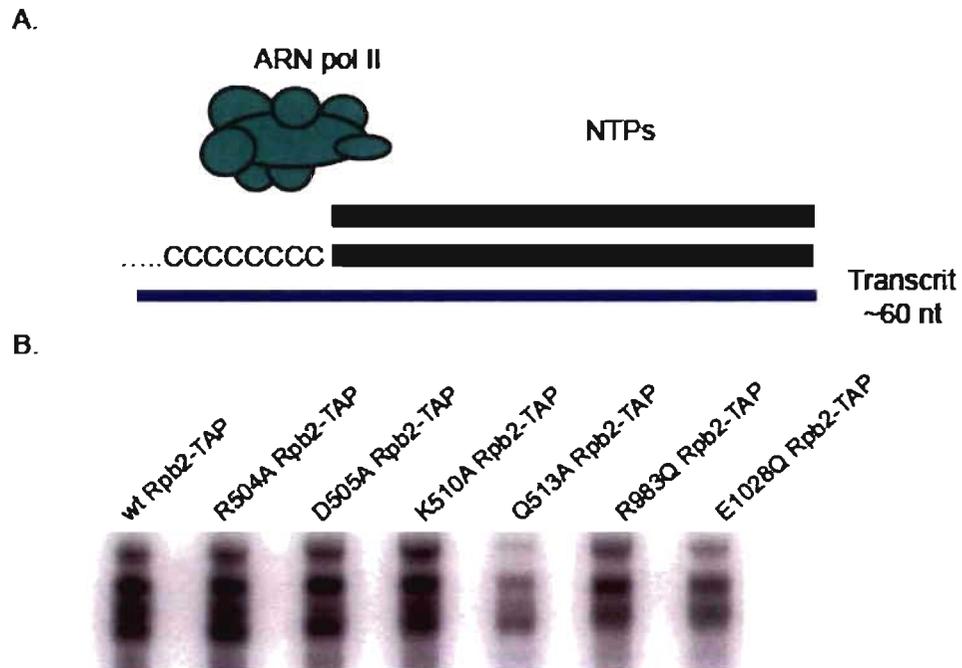


Figure 17 : Essai d'élongation in vitro.

A. Représentation schématique d'un essai d'élongation in vitro. L'ARNPII est incubée avec une matrice d'ADN composée de deux brins appariés et laissant une extrémité 3' libre. L'enzyme se fixe sur la queue de polyC en 3' et transcrit la matrice en l'absence des FGT.

B. Élongation in vitro des ARNPII purifiées de levure.

Après cette première analyse fonctionnelle, les ARNPII mutées de levures sont maintenant disponibles pour réaliser des essais biochimiques plus poussés. Ces essais nous permettront de mieux caractériser la fonction des résidus mutés ainsi que le rôle des éléments structuraux du site actif de l'ARNPII.

4 DISCUSSION

4-1 Purification simple et efficace de l'ARNPII

La purification de l'ARNPII est classiquement réalisée en plusieurs étapes de chromatographie d'échange d'ions et en utilisant des tampons à fortes concentrations en sels. La méthode de purification TAP développée par le laboratoire de Séraphin [Rigaut, *et al.*, 1999] permet de purifier à partir d'une protéine étiquetée des complexes protéiques en conditions natives grâce à deux étapes de chromatographie d'affinité. Dans notre laboratoire, cette méthode a été adaptée pour la purification de complexes protéiques à partir de culture de cellules humaines. L'utilisation en cellules humaines d'un système d'expression inductible à l'ecdysone a permis de limiter la formation d'interactions protéiques non-spécifiques en empêchant la surexpression de la protéine étiquetée. L'utilisation de la méthode de purification TAP, dans le laboratoire, a permis de purifier de manière simple et efficace à partir d'une sous-unité étiquetée de l'ARNPII un complexe transcriptionnel contenant les 12 sous-unités de l'ARNPII ainsi que les FGT, TFIIF et TFIIB, la phosphatase Fcp1 et la protéine à fonction inconnue RPAP1 [Jeronimo *et al.*, 2004].

Dans le but d'étudier le mécanisme d'action de l'ARNPII, nous avons modifié le protocole de purification TAP mis au point au laboratoire. Afin d'obtenir des préparations très pures de l'enzyme, nous avons choisi d'augmenter la concentration de sels. Ainsi, les tampons utilisés durant la purification de l'ARNPII contiennent 500mM de NaCl. Cette modification de la méthode nous a permis de purifier l'enzyme complète (12 sous-unités) seule. Les protéines copurifiées avec l'ARNPII avec des tampons à faible concentration en sels sont perdues avec cette adaptation de la méthode de purification (Figure 6).

Pour poursuivre notre étude de l'ARNPII chez *S. cerevisiae*, nous avons adapté le protocole de purification TAP aux levures. Pour cela, nous avons modifié le protocole d'extraction des protéines mis au point en cellules humaines. Les étapes de purification sont les mêmes pour les deux types de cellules. Pour les levures, nous avons également

choisi d'utiliser une forte concentration en sels dans nos tampons afin de ne purifier que l'ARNPII.

L'utilisation dans la purification des ARNPII humaines ou de levures d'une forte concentration de sels n'a pas gêné la purification de l'ensemble des sous-unités de l'ARNPII de type sauvage (Figures 8 et 15). Ce changement du protocole nous a permis d'étudier le comportement de l'ARNPII seule en s'affranchissant du biais que pouvait représenter les cofacteurs de l'enzyme dans nos préparations.

En cellules humaines, la présence de l'étiquette TAP à l'extrémité C-terminale de la sous-unité Rpb1 de l'ARNPII n'affecte pas l'activité de l'enzyme. Les ARNPII humaines sauvages fusionnées à l'étiquette TAP sont parfaitement actives *in vitro*. En effet, l'enzyme de type sauvage purifiée est capable d'initier la transcription en présence des FGT et d'allonger les transcrits d'une matrice d'ADN contenant un promoteur AdML (Figure 10). Ces complexes peuvent également reconnaître une matrice d'ADN de manière non spécifique (Figure 11) et former un complexe de pré-initiation en présence des FGT sur un promoteur AdML (Figure 12).

De la même manière, l'étiquette TAP en C-terminal n'affecte pas les capacités de l'ARNPII de type sauvage de levure *in vitro* et *in vivo*. *In vivo*, les souches de levures qui ne contiennent plus que la sous-unité Rpb2 étiquetée poussent de la même manière que les souches contenant la sous-unité Rpb2 de type sauvage (Figure 16). Ce résultat est confirmé *in vitro* par l'essai d'élongation qui met en évidence la capacité de l'enzyme d'allonger un transcrit à partir d'une matrice d'ADN (Figure 17).

4-2 Choix de la levure comme système modèle d'étude

Le choix d'utiliser la levure *S. cerevisiae* comme modèle d'étude alternatif de l'ARNPII a été fait car il s'agit d'un eucaryote unicellulaire. C'est un organisme très utile pour étudier la transcription qui est un mécanisme très conservé dans l'évolution.

En effet, même si la levure est un eucaryote inférieur, elle possède tous les constituants essentiels communs aux eucaryotes. Le choix de *S. cerevisiae* s'avère très pertinent pour l'étude de l'ARNPII car l'identité de séquence primaire pour l'ensemble des sous-unités l'enzyme chez l'homme et la levure est de 53%.

De plus, la levure *S. cerevisiae* est aussi facilement manipulable qu'une bactérie ce qui rend la mutagenèse et l'étude des mutations plus rapides et aisées qu'avec des cellules humaines. Enfin, l'utilisation de la levure comme organisme d'étude de l'ARNPII nous a permis d'étudier *in vivo* l'effet de l'étiquette TAP ainsi que des mutations de l'enzyme. En effet, contrairement à notre système d'étude en cellules humaines, chez *S. cerevisiae*, la délétion du gène d'intérêt a été possible ce qui nous a permis d'analyser le comportement des cellules contenant seulement un gène muté.

4-3 Étude structure / fonction du site catalytique de l'ARNPII

Les données obtenues par cristallographie de l'ARNPII ont permis d'émettre des hypothèses concernant la fonction de plusieurs éléments structuraux du site actif. Dans le but de comprendre le rôle de plusieurs de ces éléments, nous avons réalisé une mutagenèse dirigée des sous-unités catalytiques Rpb2 et Rpb1 respectivement chez la levure *S. cerevisiae* et chez l'Homme. Ces sous-unités étaient fusionnées avec une étiquette TAP en C-terminal pour permettre la purification de l'ARNPII.

4-3.1 Chez l'Homme

En cellules humaines, notre étude de la sous-unité Rpb1 de l'ARNPII nous a permis de poursuivre la caractérisation de mutants précédemment créés au laboratoire. Ainsi, nos résultats indiquent que la délétion complète de la fermeture à glissière

(mutant Zip) ne semble pas affecter le repliement de Rpb1 ni l'assemblage de l'ARNPII (Figure 8). La caractérisation de ce mutant par un essai de retard sur gel a mis en évidence l'incapacité de ce mutant de l'ARNPII à lier une matrice d'ADN de manière non spécifique (Figure 11). Par ce fait même, il est incapable de former le complexe de préinitiation en présence des facteurs généraux de transcription au niveau d'un promoteur (Figure 12) et de réaliser une transcription dans un système reconstitué (Figure 10). De plus, des essais d'élongation, indépendants du promoteur et des FGT, réalisés précédemment (données non présentées) ont montré que ce mutant était incapable de mener toute activité transcriptionnelle.

Les structures cristallographiques ont suggéré que la fermeture à glissière jouait un rôle dans la séparation des brins codant et non-codant de l'ADN. Cette hypothèse est consistante avec les résultats que nous avons obtenus. La délétion de cette boucle pourrait entraîner une incapacité de liaison de l'ARNPII à l'ADN, ce qui entraînerait un mauvais positionnement de la matrice d'ADN dans le site actif et une incapacité de l'enzyme à initier la transcription. L'absence de la fermeture à glissière pourrait ainsi conduire à un défaut au niveau de l'ouverture de la bulle de transcription et la séparation des brins d'ADN en amont du site actif.

Les résultats que nous avons obtenus en étudiant la mutation des trois acides aminés 1431, 1432 et 1433 de la boucle « switch » 1 en alanines ont mis en évidence l'importance de ces trois résidus dans la synthèse d'ARNm par l'ARNPII. Cette triple mutation ne semble avoir aucun effet sur l'assemblage et la stabilité de l'ARNPII (Figure 8). Par contre, l'ARNPII mutée est incapable de supporter une réaction de transcription (Figure 10). Nous nous sommes donc intéressés aux capacités de cette enzyme mutée à lier l'ADN de manière non spécifique et à former un complexe de préinitiation. Ce mutant semble s'assembler normalement au promoteur (Figure 12) ainsi que lier non spécifiquement une matrice d'ADN comparativement à l'enzyme de type sauvage (Figure 11). La boucle « switch » 1, d'après les données issues de la cristallographie de l'ARNPII, est non structurée quand l'enzyme est libre alors qu'elle devient structurée quand l'ARNPII est complexée à des acides nucléiques [Cramer *et al.*, 2001]. La boucle « switch » 1 ainsi que les quatre autres boucles « switch » permettrait

le couplage de la fermeture de la pince à la présence d'un hybride ADN-ARN dans le canal principal de l'enzyme [Gnatt *et al.*, 2001]. Nos résultats ne nous permettent pas de confirmer cette hypothèse ; néanmoins la boucle « switch » 1 semble jouer un rôle important après la formation du complexe de préinitiation qui permet le début de la transcription. Néanmoins, les défauts de transcription de notre mutants de la boucle « switch » 1 et sa capacité à former un complexe de préinitiation sont cohérents avec l'hypothèse selon laquelle la boucle « switch » 1 aurait un rôle dans le maintien de la bulle de transcription en aval du site +1 [Gnatt *et al.*, 2001]. Pour confirmer nos résultats, des essais d'initiation de la transcription pourraient être réalisés avec des matrices d'ADN fermées et pré-ouvertes afin de voir si notre mutant est capable d'initier la transcription avec une bulle de transcription artificiellement ouverte.

L'étude que nous avons réalisé d'un mutant de la boucle « switch » 2 dont 6 acides aminés K346, E347, G348, R349, V350 et R351 ont été remplacés par des alanines ne nous permet pas de conclure sur le rôle de cette boucle car dans notre préparation fonctionnelle de cette ARNPII mutée la sous-unité Rpb1 est présente sous deux formes, entière et tronquée en N-terminale (Figures 8 et 9). Nous avons toutefois poursuivi notre étude sur ce mutant car il est peu probable que la sous-unité Rpb1 tronquée soit capable de s'assembler avec les autres sous-unités de l'ARNPII. Ainsi ce mutant semble incapable de transcrire une matrice d'ADN (Figure 10) cependant il peut former un complexe de préinitiation et de s'associer de manière spécifique au promoteur (Figure 12). Néanmoins, cette mutation de Rpb1 paraît affectée la liaison non spécifique de l'ARNPII à une matrice d'ADN (Figure 11). La boucle « switch » 2, tout comme la boucle « switch » 1, n'est pas structurée quand l'ARNPII est libre alors qu'elle est structurée quand l'enzyme lie des acides nucléiques [Cramer *et al.*, 2001]. Nos résultats sont en accord avec une interaction de la boucle « switch » 2 avec les acides nucléiques. La capacité de notre mutant à lier très fortement l'ADN de manière non spécifique comparativement à l'enzyme de type sauvage est consistante avec une étude récente sur la boucle « switch » 2. Cette étude, réalisée chez *S. cerevisiae*, a montré le rôle de cette boucle dans la stabilisation de l'hybride durant l'initiation [Majovski *et al.*, 2005].

4-3.2 Chez *S. cerevisiae*

Chez la levure, nos résultats ont montré que des mutations dans la sous-unité Rpb2 peuvent affecter la fonction de l'ARNPII. Ainsi, il apparaît que la mutation de l'acide aminé 512, de la boucle d'embranchement 2, en alanine est létale. La létalité de cette mutation peut s'expliquer de plusieurs façons : une perte de fonctionnalité de l'enzyme mais aussi un défaut dans le repliement de la sous-unité Rpb2 ou un défaut d'assemblage de Rpb2 avec les autres sous-unités de l'ARNPII. La boucle d'embranchement 2 est située en aval de la bulle de transcription. Il a été proposé que cette boucle soit impliquée dans le maintien de la bulle de transcription. Une étude précédente a montré que la mutation R512C n'était pas létale chez *S. cerevisiae* [Pappas et Hampsey, 2000]. D'autre part, des expériences préliminaires *in vitro* de cette mutation ont montré qu'elle présentait des difficultés d'élongation [Burton *et al.*, résultats non publiés]. Parmi les autres mutants créés dans la boucle d'embranchement 2, aucun n'est létal. Ils ne présentent pas de phénotype différent du phénotype sauvage en élongation. La présence de la drogue 6-azauracile ainsi que l'absence du facteur général de la transcription TFIIS n'affecte pas leur croissance. Les résultats que nous avons obtenus confirment l'importance du résidu 512 de la boucle d'embranchement 2 pour l'activité de l'ARNPII.

Parmi les trois mutations créées au niveau des résidus situés dans le site actif de l'ARNPII, l'acide aminé 1020 est essentiel à la survie cellulaire. De la même manière que pour le résidu 512, la létalité de la mutation peut s'expliquer par un défaut structurel de l'ARNPII ou par une perte de l'activité de l'enzyme. Les deux autres mutants créés dans ce domaine de Rpb2 présentent des défauts d'élongation. En effet, le mutant R983Q a un phénotype de croissance lent en l'absence de TFIIS mais n'a pas de phénotype particulier en présence de 6-AU. Alors que le mutant E1028Q n'est pas affecté par la délétion de TFIIS mais présente un phénotype de croissance lent en présence de 6-AU. De plus, des essais biochimiques préliminaires ont montré que ces deux mutants avaient un phénotype très lent en élongation [Burton *et al.*, résultats non publiés]. D'après les structures cristallographiques de l'ARNPII, ces deux résidus

interagissent entre eux par une liaison ionique. Il s'avère donc intéressant que ces mutants présentent tous les deux un phénotype similaire en élongation mais pas identique. La paire d'acides aminés R983 et E1028 semble donc avoir un rôle important dans le mécanisme de catalyse du site actif de l'ARNPII.

2 CONCLUSION ET PERSPECTIVES

L'ARNPII est responsable de la synthèse des ARNm chez les eucaryotes. Avec l'assistance des facteurs généraux de transcription (FGT) TFIIA, TFIIB, TFIID, TFIIIE, TFIIF et TFIIH, l'ARNPII reconnaît et lie les promoteurs géniques, ouvre la double hélice d'ADN au niveau du site d'initiation de la transcription, catalyse la polymérisation des nucléotides en utilisant l'un des brins d'ADN comme matrice, et se déplace le long de l'unité génique en allongeant l'ARN en synthèse. L'ARNPII s'associe aussi avec des coactivateurs qui transmettent les signaux régulateurs à la polymérase, des facteurs d'élongation qui permettent la production efficace de longs transcrits d'ARN et des facteurs multiprotéiques pour la maturation des ARNm et la terminaison. Suite à la résolution de la structure cristallographique de l'ARNPII, cela a permis de proposer un rôle pour certains domaines localisés près du site actif de l'enzyme.

L'objectif de ma maîtrise était d'étudier le rôle de différents éléments structuraux des sous-unités Rpb1 et Rpb2, entourant le site catalytique de l'ARNPII, qui avaient été révélés par les structures cristallographiques de l'enzyme.

Pour cela, dans un premier temps, nous avons créé une série de mutants portant des modifications dans Rpb1 au niveau de la fermeture à glissière et des boucles « switch » 1 et 2 en utilisant les cellules humaines EcR293. Nous avons purifié par la méthode TAP l'ARNPII de type sauvage et ses formes mutantes à partir de cellules humaines grâce à l'étiquette TAP placée à l'extrémité carboxyterminale de Rpb1. Nous avons pu observer que ces trois ARNPII mutantes sont capables de s'assembler et de former le complexe de polymérase à 12 sous-unités ; mais elles présentent toutes les trois des défauts de transcription. Il semble que ces trois mutations affectent des étapes précoces du mécanisme de transcription. Cependant, ces mutations ne paraissent pas toucher les mêmes fonctions de l'ARNPII. Ainsi, la délétion complète de la fermeture à glissière empêche la liaison de l'enzyme à l'ADN alors que les mutants des boucles « switch » 1 et 2 affectent vraisemblablement une étape suivant la formation du complexe de préinitiation.

Afin de terminer la caractérisation de ces trois mutants humains, il serait intéressant de confirmer nos résultats obtenus *in vitro* par des expériences *in vivo*. Il serait envisageable de vérifier les capacités de liaison aux promoteurs de ces mutants par des essais d'immunoprécipitation de la chromatine (ChIP).

La difficulté de création et de caractérisation de mutations en cellules humaines nous a amené à nous tourner vers un autre organisme d'étude de l'ARNPII. Dans un second temps, nous avons choisi d'utiliser la levure *S. cerevisiae* comme organisme modèle. Ainsi, nous avons pu créer et purifier avec la méthode TAP une série d'ARNPII mutantes modifiées au niveau de la boucle d'embranchement 2 et du site catalytique dans la sous-unité Rpb2. L'étude de ces mutants nous a permis de mettre en évidence l'importance du résidu R512 de la boucle d'embranchement 2 pour l'ARNPII et le rôle dans l'étape d'élongation des résidus R983 et E1028 situés dans le site catalytique de l'enzyme. Nos résultats ont montré l'intérêt et la pertinence du choix d'étudier l'ARNPII chez un eucaryote inférieur comme la levure *S. cerevisiae*.

Dans le but de poursuivre ce projet, il serait intéressant de continuer la mutagenèse, commencée chez *S. cerevisiae*, de la sous-unité Rpb2 notamment au niveau de la boucle d'embranchement 2 et des acides aminés impliqués dans la réaction de catalyse de l'ARNPII afin de mieux comprendre le mécanisme moléculaire précis de l'enzyme. Il serait également pertinent de réaliser une série de mutations au niveau de différents éléments de la sous-unité Rpb1. Par exemple, il serait intéressant d'analyser le rôle de l'hélice de pontage et de la boucle « trigger », toutes deux impliquées dans la translocation des acides nucléiques.

L'étude structure / fonction du site catalytique de l'ARNPII, qui a été réalisée en cellules humaines et en levures par mutagenèse, a permis de mieux comprendre le mécanisme moléculaire de l'enzyme. Cette étude a mis en évidence le rôle crucial des boucles conservées du site catalytique dans l'assemblage de l'ARNPII, l'assemblage au promoteur et la processivité de la transcription. Enfin, cette étude a également montré l'intérêt de mener des expériences biochimiques *in vitro* afin de confirmer les hypothèses émises après la résolution de la structure cristallographique de l'ARNPII.

6 LISTE DES RÉFÉRENCES

Andel,F., III, Ladurner,A.G., Inouye,C., Tjian,R. & Nogales,E. Three-dimensional structure of the human TFIID-IIA-IIB complex. *Science* 286, 2153-2156 (1999).

Archambault,J., Chambers,R.S., Kobor,M.S., Ho,Y., Cartier,M., Bolotin,D., Andrews,B., Kane,C.M. & Greenblatt,J. An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A* 94, 14300-14305 (1997).

Archambault,J., Pan,G., Dahmus,G.K., Cartier,M., Marshall,N., Zhang,S., Dahmus,M.E. & Greenblatt,J. FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* 273, 27593-27601 (1998).

Ares,M., Jr. & Proudfoot,N.J. The spanish connection: transcription and mRNA processing get even closer. *Cell* 120, 163-166 (2005).

Armache,K.J., Kettenberger,H. & Cramer,P. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A* 100, 6964-6968 (2003).

Armache,K.J., Mitterweger,S., Meinhart,A. & Cramer,P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol Chem.* 280, 7131-7134 (2005).

Awrey,D.E., Weilbaecher,R.G., Hemming,S.A., Orlicky,S.M., Kane,C.M. & Edwards,A.M. Transcription elongation through DNA arrest sites. A multistep process

involving both RNA polymerase II subunit RPB9 and TFIIS. *J. Biol. Chem.* 272, 14747-14754 (1997).

Awrey,D.E., Shimasaki,N., Koth,C., Weilbaecher,R., Olmsted,V., Kazanis,S., Shan,X., Arellano,J., Arrowsmith,C.H., Kane,C.M. & Edwards,A.M. Yeast transcript elongation factor (TFIIS), structure and function. II: RNA polymerase binding, transcript cleavage, and read-through. *J. Biol Chem.* 273, 22595-22605 (1998).

Bar-Nahum,G., Epshtein,V., Ruckenstein,A.E., Rafikov,R., Mustaev,A. & Nudler,E. A ratchet mechanism of transcription elongation and its control. *Cell* 120, 183-193 (2005).

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

Bhoite,L.T., Yu,Y. & Stillman,D.J. The Swi5 activator recruits the Mediator complex to the HO promoter without RNA polymerase II. *Genes Dev.* 15, 2457-2469 (2001).

Booth,V., Koth,C.M., Edwards,A.M. & Arrowsmith,C.H. Structure of a conserved domain common to the transcription factors TFIIS, elongin A, and CRSP70. *J. Biol Chem.* 275, 31266-31268 (2000).

Borggreffe,T., Davis,R., Bareket-Samish,A. & Kornberg,R.D. Quantitation of the RNA polymerase II transcription machinery in yeast. *J. Biol. Chem.* 276, 47150-47153 (2001).

Brand,M., Leurent,C., Mallouh,V., Tora,L. & Schultz,P. Three-dimensional structures of the TAFII-containing complexes TFIID and TFIIIC. *Science* 286, 2151-2153 (1999).

Bucheli,M.E., He,X., Kaplan,C.D., Moore,C.L. & Buratowski,S. Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI. *RNA* 13, 1756-1764 (2007).

Burke,T.W. & Kadonaga,J.T. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* 10, 711-724 (1996).

Bushnell,D.A., Cramer,P. & Kornberg,R.D. Structural basis of transcription: alpha-amanitin-RNA polymerase II cocystal at 2.8 A resolution. *Proc. Natl. Acad. Sci. U. S. A* 99, 1218-1222 (2002).

Bushnell,D.A. & Kornberg,R.D. Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc. Natl. Acad. Sci. U S. A* 100, 6969-6973 (2003).

Bushnell,D.A., Westover,K.D., Davis,R.E. & Kornberg,R.D. Structural basis of transcription: an RNA polymerase II-TFIIB cocystal at 4.5 Angstroms. *Science* 303, 983-988 (2004).

Casamassimi,A. & Napoli,C. Mediator complexes and eukaryotic transcription regulation: An overview. *Biochimie* 89, 1439-1446 (2007).

Chambers,R.S., Wang,B.Q., Burton,Z.F. & Dahmus,M.E. The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II and by the general transcription factors IIF and IIB. *J. Biol. Chem.* 270, 14962-14969 (1995).

Chen,Z. & Manley,J.L. In vivo functional analysis of the histone 3-like TAF9 and a TAF9-related factor, TAF9L. *J. Biol Chem.* 278, 35172-35183 (2003).

Cheng,B. & Price,D.H. Properties of RNA Polymerase II Elongation Complexes Before and After the P-TEFb-mediated Transition into Productive Elongation. *J. Biol. Chem.* 282, 21901-21912 (2007).

Cheong,J.H., Yi,M., Lin,Y. & Murakami,S. Human RPB5, a subunit shared by eukaryotic nuclear RNA polymerases, binds human hepatitis B virus X protein and may play a role in X transactivation. *EMBO J.* 14, 143-150 (1995).

Cho,E.J., Kobor,M.S., Kim,M., Greenblatt,J. & Buratowski,S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev.* 15, 3319-3329 (2001).

Cojocar M, Jeronimo C, Forget D, Bouchard A, Bergeron D, Côte P, Poirier GG, Greenblatt J & Coulombe B. Genomic location of the human RNA polymerase II general machinery: evidence for a role of TFIIF and Rpb7 at both early and late stages of transcription. *Biochem J.* 409(1), 139-147. 1-1-2008.

Ref Type: Generic

Conaway,R.C., Sato,S., Tomomori-Sato,C., Yao,T. & Conaway,J.W. The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci.* 30, 250-255 (2005).

Concino,M.F., Lee,R.F., Merryweather,J.P. & Weinmann,R. The adenovirus major late promoter TATA box and initiation site are both necessary for transcription in vitro. *Nucl. Acids Res.* 12, 7423-7433 (1984).

Corden,J., Wasylyk,B., Buchwalder,A., Sassone-Corsi,P., Kedinger,C. & Chambon,P. Promoter sequences of eukaryotic protein-coding genes. *Science* 209(4463), 1406-1414. 9-19-1980.

Ref Type: Generic

Costa,P.J. & Arndt,K.M. Synthetic lethal interactions suggest a role for the *Saccharomyces cerevisiae* Rtf1 protein in transcription elongation. *Genetics* 156, 535-547 (2000).

Coulombe,B. & Burton,Z.F. DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. *Microbiol. Mol. Biol. Rev.* 63, 457-478 (1999).

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

Cramer,P. RNA polymerase II structure: from core to functional complexes. *Curr. Opin. Genet. Dev.* 14, 218-226 (2004).

Davie,J.K. & Kane,C.M. Genetic Interactions between TFIIIS and the Swi-Snf Chromatin-Remodeling Complex. *Mol. Cell. Biol.* 20, 5960-5973 (2000).

Deng,W. & Roberts,S.G. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev.* 19, 2418-2423 (2005).

Dierks,P., van Ooyen,A., Cochran,M., Dobkin,C., Reiser,J. & Weissmann,C. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. *Cell* 32(3), 695-706. 1983.
Ref Type: Generic

Domecq,C., Trinh,V., Langelier,M.-F., Archambault,J. & Coulombe,B. Inhibitors of multisubunit RNA polymerases as tools to study transcriptional mechanisms in prokaryotes and eukaryotes. *Curr.Chem.Biol.* 2, 20-31. 2008.
Ref Type: Generic

Eberl,D.F., Duyf,B.J. & Hilliker,A.J. The Role of Heterochromatin in the Expression of a Heterochromatic Gene, the rolled Locus of *Drosophila melanogaster*. *Genetics* 134, 277-292 (1993).

Emami,K.H., Jain,A. & Smale,S.T. Mechanism of synergy between TATA and initiator: synergistic binding of TFIID following a putative TFIIA-induced isomerization. *Genes Dev.* 11, 3007-3019 (1997).

Epshtein,V., Cardinale,C.J., Ruckenstein,A.E., Borukhov,S. & Nudler,E. An Allosteric Path to Transcription Termination. *Molecular Cell* 28, 991-1001 (2007).

Evans,R., Fairley,J.A. & Roberts,S.G. Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes Dev.* 15, 2945-2949 (2001).

Exinger,F. & Lacroute,F. 6-Azauracil Inhibition of Gtp Biosynthesis in *Saccharomyces-Cerevisiae*. *Current Genetics* 22, 9-11 (1992).

Farago,M., Nahari,T., Hammel,C., Cole,C.N. & Choder,M. Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. *Mol Biol Cell* 14, 2744-2755 (2003).

Featherstone,M. Coactivators in transcription initiation: here are your orders. *Curr. Opin. Genet. Dev.* 12, 149-155 (2002).

Fish,R.N., Ammerman,M.L., Davie,J.K., Lu,B.F., Pham,C., Howe,L., Ponticelli,A.S. & Kane,C.M. Genetic Interactions Between TFIIF and TFIIS. *Genetics* 173, 1871-1884 (2006).

Forget,D., Robert,F., Grondin,G., Burton,Z.F., Greenblatt,J. & Coulombe,B. 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 (1997).

Fourel,G., Magdinier,F. & Gilson,E. Insulator dynamics and the setting of chromatin domains. *Bioessays* 26, 523-532 (2004).

Freire-Picos,M.A., Krishnamurthy,S., Sun,Z.W. & Hampsey,M. Evidence that the Tfg1/Tfg2 dimer interface of TFIIF lies near the active center of the RNA polymerase II initiation complex. *Nucl. Acids Res.* 33, 5045-5052 (2005).

Garrett,K.P., Serizawa,H., Hanley,J.P., Bradsher,J.N., Tsuboi,A., Arai,N., Yokota,T., Arai,K., Conaway,R.C. & Conaway,J.W. The carboxyl terminus of RAP30 is similar in sequence to region 4 of bacterial sigma factors and is required for function. *J. Biol. Chem.* 267, 23942-23949 (1992).

Gaston,K. & Jayaraman,P.S. Transcriptional repression in eukaryotes: repressors and repression mechanisms. *Cell Mol Life Sci.* 60, 721-741 (2003).

Gaszner,M. & Felsenfeld,G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* 7, 703-713 (2006).

Gnatt,A.L., Cramer,P., Fu,J., Bushnell,D.A. & Kornberg,R.D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292, 1876-1882 (2001).

Grosschedl,R. & Birnstiel,M.L. Delimitation of Far Upstream Sequences Required for Maximal in vitro Transcription of an H2A Histone Gene. *Proceedings of the National Academy of Sciences* 79, 297-301 (1982).

Guarente,L. & Hoar,E. Upstream activation sites of the CYC1 gene of *Saccharomyces cerevisiae* are active when inverted but not when placed downstream of the "TATA box". *Proc. Natl. Acad. Sci. U. S. A* 81, 7860-7864 (1984).

Guglielmi,B., Soutourina,J., Esnault,C. & Werner,M. TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proceedings of the National Academy of Sciences* 104, 16062-16067 (2007).

Hahn,S., Buratowski,S., Sharp,P.A. & Guarente,L. Yeast TATA-Binding Protein TFIID Binds to TATA Elements with Both Consensus and Nonconsensus DNA Sequences. *Proceedings of the National Academy of Sciences* 86, 5718-5722 (1989).

Hanna-Rose,W. & Hansen,U. Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet.* 12, 229-234 (1996).

Hartzog,G.A., Wada,T., Handa,H. & Winston,F. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes Dev.* 12, 357-369 (1998).

Hemming,S.A., Jansma,D.B., Macgregor,P.F., Goryachev,A., Friesen,J.D. & Edwards,A.M. RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. *J. Biol. Chem.* 275, 35506-35511 (2000).

Hoiby,T., Zhou,H., Mitsiou,D.J. & Stunnenberg,H.G. A facelift for the general transcription factor TFIIA. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1769, 429-436 (2007).

Hull,M.W., McKune,K. & Woychik,N.A. RNA polymerase II subunit RPB9 is required for accurate start site selection. *Genes Dev.* 9, 481-490 (1995).

Javahery,R., Khachi,A., Lo,K., Zenzie-Gregory,B. & Smale,S.T. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.* 14, 116-127 (1994).

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. & Coulombe,B. 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 (2004).

Kamenski,T., Heilmeyer,S., Meinhart,A. & Cramer,P. Structure and mechanism of RNA polymerase II CTD phosphatases. *Mol Cell* 15, 399-407 (2004).

Kang,X., Hu,Y., Li,Y., Guo,X., Jiang,X., Lai,L., Xia,B. & Jin,C. Structural, Biochemical, and Dynamic Characterizations of the hRPB8 Subunit of Human RNA Polymerases. *J. Biol. Chem.* 281, 18216-18226 (2006).

Kapanidis,A.N., Margeat,E., Ho,S.O., Kortkhonjia,E., Weiss,S. & Ebright,R.H. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314, 1144-1147 (2006).

Kettenberger,H., Armache,K.J. & Cramer,P. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* 114, 347-357 (2003).

Kim,B., Nesvizhskii,A.I., Rani,P.G., Hahn,S., Aebersold,R. & Ranish,J.A. The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proceedings of the National Academy of Sciences* 104, 16068-16073 (2007).

Kim,J.L., Nikolov,D.B. & Burley,S.K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365, 520-527 (1993).

Kim,Y., Geiger,J.H., Hahn,S. & Sigler,P.B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365, 512-520 (1993).

Kim,Y.J., Bjorklund,S., Li,Y., Sayre,M.H. & Kornberg,R.D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77, 599-608 (1994).

Krishnamurthy,S., He,X., Reyes-Reyes,M., Moore,C. & Hampsey,M. Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell* 14, 387-394 (2004).

Kuras,L., Borggreffe,T. & Kornberg,R.D. Association of the Mediator complex with enhancers of active genes. *Proc. Natl. Acad. Sci. U. S. A* 100, 13887-13891 (2003).

Kutach,A.K. & Kadonaga,J.T. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol Cell Biol* 20, 4754-4764 (2000).

Kuznedelov,K., Minakhin,L., Niedziela-Majka,A., Dove,S.L., Rogulja,D., Nickels,B.E., Hochschild,A., Heyduk,T. & Severinov,K. A role for interaction of the

RNA polymerase flap domain with the sigma subunit in promoter recognition. *Science* 295, 855-857 (2002).

Kuznedelov,K., Korzheva,N., Mustaev,A. & Severinov,K. Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length. *EMBO J.* 21, 1369-1378 (2002).

Lagrange,T., Kapanidis,A.N., Tang,H., Reinberg,D. & Ebright,R.H. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor α IIB. *Genes Dev.* 12, 34-44 (1998).

Langelier,M.F., Baali,D., Trinh,V., Greenblatt,J., Archambault,J. & Coulombe,B. The highly conserved glutaminic acid 791 of Rpb2 is involved in binding of NTP and Mg(B) in the active center of human RNA polymerase II. *Nucleic Acids Res.* 33, 2629-2639 (2005).

Lee,D.H., Gershenson,N., Gupta,M., Ioshikhes,I.P., Reinberg,D. & Lewis,B.A. Functional Characterization of Core Promoter Elements: the Downstream Core Element Is Recognized by TAF1. *Mol Cell Biol* 25, 9674-9686 (2005).

Lee,T.I. & Young,R.A. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* 34, 77-137 (2000).

Lewis,B.A., Kim,T.K. & Orkin,S.H. A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proc. Natl. Acad. Sci. U. S. A* 97, 7172-7177 (2000).

- Lewis,B.A. & Reinberg,D.** The mediator coactivator complex: functional and physical roles in transcriptional regulation. *J. Cell Sci.* 116, 3667-3675 (2003).
- Li,S. & Smerdon,M.J.** Rpb4 and Rpb9 mediate subpathways of transcription-coupled DNA repair in *Saccharomyces cerevisiae*. *EMBO J.* 21, 5921-5929 (2002).
- Lim,C.Y., Santoso,B., Boulay,T., Dong,E., Ohler,U. & Kadonaga,J.T.** The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev.* 18, 1606-1617 (2004).
- Lo,K. & Smale,S.T.** Generality of a functional initiator consensus sequence. *Gene* 182, 13-22 (1996).
- Lotan,R., Bar-On VG, Harel-Sharvit,L., Duek,L., Melamed,D. & Choder,M.** The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes Dev.* 19, 3004-3016 (2005).
- Lotan,R., Goler-Baron,V., Duek,L., Haimovich,G. & Choder,M.** The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *J. Cell Biol.* 178, 1133-1143 (2007).
- Lykke-Andersen,S. & Jensen,T.H.** Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* 89, 1177-1182 (2007).
- Majovski,R.C., Khapersky,D.A., Ghazy,M.A. & Ponticelli,A.S.** A functional role for the switch 2 region of yeast RNAPII in transcription start site utilization and abortive initiation. *J. Biol Chem.* (2005).

Malagon,F., Tong,A.H., Shafer,B.K. & Strathern,J.N. Genetic Interactions of DST1 in *Saccharomyces cerevisiae* Suggest a Role of TFIIS in the Initiation-Elongation Transition. *Genetics* 166, 1215-1227 (2004).

Martinez,E., Zhou,Q., L'Etoile,N.D., Oelgeschlager,T., Berk,A.J. & Roeder,R.G. Core promoter-specific function of a mutant transcription factor TFIID defective in TATA-box binding. *Proc. Natl. Acad. Sci. U. S. A* 92, 11864-11868 (1995).

Mason,P.B. & Struhl,K. Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell* 17, 831-840 (2005).

McCracken,S. & Greenblatt,J. Related RNA polymerase-binding regions in human RAP30/74 and *Escherichia coli* sigma 70. *Science* 253, 900-902 (1991).

McKune,K., Moore,P.A., Hull,M.W. & Woychik,N.A. Six human RNA polymerase subunits functionally substitute for their yeast counterparts. *Mol. Cell Biol.* 15, 6895-6900 (1995).

Minakhin,L., Bhagat,S., Brunning,A., Campbell,E.A., Darst,S.A., Ebright,R.H. & Severinov,K. Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc. Natl. Acad. Sci. U. S. A* 98, 892-897 (2001).

Mooney,R.A., Artsimovitch,I. & Landick,R. Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *J. Bacteriol.* 180, 3265-3275 (1998).

Morin,P.E., Awrey,D.E., Edwards,A.M. & Arrowsmith,C.H. Elongation factor TFIIS contains three structural domains: solution structure of domain II. *Proc. Natl. Acad. Sci. U. S. A* 93, 10604-10608 (1996).

Myer,V.E. & Young,R.A. RNA polymerase II holoenzymes and subcomplexes. *J. Biol. Chem.* 273, 27757-27760 (1998).

Myers,L.C. & Kornberg,R.D. Mediator of transcriptional regulation. *Annu. Rev. Biochem.* 69, 729-749 (2000).

Nesser,N.K., Peterson,D.O. & Hawley,D.K. RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. *Proceedings of the National Academy of Sciences* 103, 3268-3273 (2006).

Nonet,M., Sweetser,D. & Young,R.A. Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell* 50, 909-915 (1987).

Nonet,M.L. & Young,R.A. Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of *Saccharomyces cerevisiae* RNA polymerase II. *Genetics* 123, 715-724 (1989).

O'Shea-Greenfield,A. & Smale,S.T. Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J. Biol. Chem.* 267, 6450 (1992).

Ogbourne,S. & Antalis,T.M. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem. J.* 331 (Pt 1), 1-14 (1998).

Ohler,U., Liao,G.c., Niemann,H. & Rubin,G. Computational analysis of core promoters in the Drosophila genome. *Genome Biology* 3, research0087 (2002).

Olmsted,V.K., Awrey,D.E., Koth,C., Shan,X., Morin,P.E., Kazanis,S., Edwards,A.M. & Arrowsmith,C.H. Yeast transcript elongation factor (TFIIS), structure and function. I: NMR structural analysis of the minimal transcriptionally active region. *J. Biol Chem.* 273, 22589-22594 (1998).

Opalka,N., Chlenov,M., Chacon,P., Rice,W.J., Wriggers,W. & Darst,S.A. Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* 114, 335-345 (2003).

Orphanides,G., Wu,W.H., Lane,W.S., Hampsey,M. & Reinberg,D. The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* 400, 284-288 (1999).

Pan,G., Aso,T. & Greenblatt,J. Interaction of elongation factors TFIIS and elongin A with a human RNA polymerase II holoenzyme capable of promoter-specific initiation and responsive to transcriptional activators. *J. Biol. Chem.* 272, 24563-24571 (1997).

Pappas,D.L., Jr. & Hampsey,M. Functional interaction between Ssu72 and the Rpb2 subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* 20, 8343-8351 (2000).

Park,J.M., Werner,J., Kim,J.M., Lis,J.T. & Kim,Y.J. Mediator, not holoenzyme, is directly recruited to the heat shock promoter by HSF upon heat shock. *Mol. Cell* 8, 9-19 (2001).

Patikoglou,G.A., Kim,J.L., Sun,L., Yang,S.H., Kodadek,T. & Burley,S.K. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* 13, 3217-3230 (1999).

Pokholok,D.K., Hannett,N.M. & Young,R.A. Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol Cell* 9, 799-809 (2002).

Prather,D.M., Larschan,E. & Winston,F. Evidence that the elongation factor TFIIS plays a role in transcription initiation at GAL1 in *Saccharomyces cerevisiae*. *Mol Cell Biol* 25, 2650-2659 (2005).

Price,D.H. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell Biol.* 20, 2629-2634 (2000).

Qureshi,S.A. & Jackson,S.P. Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol. Cell* 1, 389-400 (1998).

Ranish,J.A., Yi,E.C., Leslie,D.M., Purvine,S.O., Goodlett,D.R., Eng,J. & Aebersold,R. The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* 33, 349-355 (2003).

Reed,R. Coupling transcription, splicing and mRNA export. *Curr. Opin. Cell Biol.* 15, 326-331 (2003).

Ren,D., Nedialkov,Y.A., Li,F., Xu,D., Reimers,S., Finkelstein,A. & Burton,Z.F. Spacing requirements for simultaneous recognition of the adenovirus major late

promoter TATAAAAG box and initiator element. *Archives of Biochemistry and Biophysics* 435, 347-362 (2005).

Revyakin,A., Liu,C., Ebright,R.H. & Strick,T.R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314, 1139-1143 (2006).

Robert,F., Douziech,M., Forget,D., Egly,J.M., Greenblatt,J., Burton,Z.F. & Coulombe,B. Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. *Mol. Cell* 2, 341-351 (1998).

Runner,V., Podolny,V. & Buratowski,S. The Rpb4 subunit of RNA polymerase II contributes to co-transcriptional recruitment of 3' processing factors. *Mol. Cell. Biol.* MCB (2008).

Sims,R.J., III, Mandal,S.S. & Reinberg,D. Recent highlights of RNA-polymerase-II-mediated transcription. *Curr. Opin. Cell Biol.* 16, 263-271 (2004).

Smale,S.T. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta* 1351, 73-88 (1997).

Smale,S.T. & Kadonaga,J.T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72, 449-479 (2003).

Sopta,M., Burton,Z.F. & Greenblatt,J. Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II. *Nature* 341, 410-414 (1989).

Sosunova,E., Sosunov,V., Kozlov,M., Nikiforov,V., Goldfarb,A. & Mustaev,A.

Donation of catalytic residues to RNA polymerase active center by transcription factor

Gre. Proc. Natl. Acad. Sci. U. S. A 100, 15469-15474 (2003).

Struhl,K. Molecular mechanisms of transcriptional regulation in yeast. *Annu. Rev.*

Biochem. 58, 1051-1077 (1989).

Sun,Z.W. & Hampsey,M. Identification of the gene (SSU71/TFG1) encoding the

largest subunit of transcription factor TFIIF as a suppressor of a TFIIB mutation in

Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A 92, 3127-3131 (1995).

Sun,Z.W., Tessmer,A. & Hampsey,M. Functional interaction between TFIIB and the

Rpb9 (Ssu73) subunit of RNA polymerase II in *Saccharomyces cerevisiae. Nucleic*

Acids Res. 24, 2560-2566 (1996).

Suzuki,Y., Tsunoda,T., Sese,J., Taira,H., Mizushima-Sugano,J., Hata,H., Ota,T.,

Isogai,T., Tanaka,T., Nakamura,Y., Suyama,A., Sakaki,Y., Morishita,S., Okubo,K.

& Sugano,S. Identification and Characterization of the Potential Promoter Regions of

1031 Kinds of Human Genes. *Genome Res.* 11, 677-684 (2001).

Takagi,Y., Masuda,C.A., Chang,W.H., Komori,H., Wang,D., Hunter,T.,

Joazeiro,C.A. & Kornberg,R.D. Ubiquitin ligase activity of TFIIF and the

transcriptional response to DNA damage. *Mol Cell* 18, 237-243 (2005).

Talkington,C.A. & Leder,P. Rescuing the in vitro function of a globin pseudogene

promoter. *Nature* 298, 192-195 (1982).

Tan,Q., Prysak,M.H. & Woychik,N.A. Loss of the Rpb4/Rpb7 Subcomplex in a Mutant Form of the Rpb6 Subunit Shared by RNA Polymerases I, II, and III. *Mol. Cell Biol.* 23, 3329-3338 (2003).

Tokunaga,K., Hirose,S. & Suzuki,Y. In monkey COS cells only the TATA box and the cap site region are required for faithful and efficient initiation of the fibroin gene transcription. *Nucl. Acids Res.* 12, 1543-1558 (1984).

Toulokhonov,I., Artsimovitch,I. & Landick,R. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* 292, 730-733 (2001).

Toulokhonov,I. & Landick,R. The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. *Mol Cell* 12, 1125-1136 (2003).

Toulokhonov,I., Zhang,J., Palangat,M. & Landick,R. A Central Role of the RNA Polymerase Trigger Loop in Active-Site Rearrangement during Transcriptional Pausing. *Molecular Cell* 27, 406-419 (2007).

Vassylyev,D.G., Vassylyeva,M.N., Zhang,J., Palangat,M., Artsimovitch,I. & Landick,R. Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448, 163-168 (2007).

Wang,D., Bushnell,D.A., Westover,K.D., Kaplan,C.D. & Kornberg,R.D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* 127, 941-954 (2006).

Wery,M., Shematorova,E., Van Driessche,B., Vandenhaute,J., Thuriaux,P. & Van,M., V. Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIS. *EMBO J.* 23, 4232-4242 (2004).

Westover,K.D., Bushnell,D.A. & Kornberg,R.D. Structural Basis of Transcription; Nucleotide Selection by Rotation in the RNA Polymerase II Active Center. *Cell* 119, 481-489 (2004).

Westover,K.D., Bushnell,D.A. & Kornberg,R.D. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* 303, 1014-1016 (2004).

Wilson,C.J., Chao,D.M., Imbalzano,A.N., Schnitzler,G.R., Kingston,R.E. & Young,R.A. RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* 84, 235-244 (1996).

Wong,J.M. & Bateman,E. TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs. *Nucl. Acids Res.* 22, 1890-1896 (1994).

Yang,C., Bolotin,E., Jiang,T., Sladek,F.M. & Martinez,E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 389, 52-65 (2007).

Yeo,M., Lin,P.S., Dahmus,M.E. & Gill,G.N. A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. *J. Biol Chem.* 278, 26078-26085 (2003).

Yonaha,M., Aso,T., Kobayashi,Y., Vasavada,H., Yasukochi,Y., Weissman,S.M. & Kitajima,S. Domain structure of a human general transcription initiation factor, TFIIF. *Nucleic Acids Res.* 21, 273-279 (1993).

Zaros,C., Briand,J.F., Boulard,Y., Labarre-Mariotte,S., Garcia-Lopez,M.C., Thuriaux,P. & Navarro,F. Functional organization of the Rpb5 subunit shared by the three yeast RNA polymerases. *Nucl. Acids Res.* 35, 634-647 (2007).

Zhang,G., Campbell,E.A., Minakhin,L., Richter,C., Severinov,K. & Darst,S.A. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98, 811-824 (1999).

Zheng,H., Ji,C., Gu,S., Shi,B., Wang,J., Xie,Y. & Mao,Y. Cloning and characterization of a novel RNA polymerase II C-terminal domain phosphatase. *Biochem. Biophys. Res. Commun.* 331, 1401-1407 (2005).

Ziegler,L.M., Khapersky,D.A., Ammerman,M.L. & Ponticelli,A.S. Yeast RNA Polymerase II Lacking the Rpb9 Subunit Is Impaired for Interaction with Transcription Factor IIF. *J. Biol. Chem.* 278, 48950-48956 (2003).