

Université de Montréal

11634252

**La machinerie transcriptionnelle basale : Étude de son organisation  
moléculaire au promoteur génique et de ses interactions avec les  
régulateurs XAB1 et DDX5**

par :  
Cynthia Thérien

Département de biochimie  
Faculté de médecine

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de maître ès sciences (M. Sc.)  
en biochimie

Avril, 2005

© Cynthia Thérien, 2005



w  
4  
U58  
2005  
V. 151

**Direction des bibliothèques**

**AVIS**

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**

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  
Faculté des études supérieures

Ce mémoire intitulé :  
La machinerie transcriptionnelle basale : Étude de son organisation moléculaire au  
promoteur génique et de ses interactions avec les régulateurs XAB1 et DDX5

Présenté par :  
Cynthia Thérien

a été évalué par un jury composé des personnes suivantes :

\_\_\_\_\_  
James G. Omichinski  
*Président-rapporteur*

\_\_\_\_\_  
Benoit Coulombe  
*Directeur de recherche*

\_\_\_\_\_  
François Robert  
*Membre du jury*

## RÉSUMÉ

L'ARN polymérase (ARNP) II est une enzyme centrale du processus d'expression des gènes chez les eucaryotes. Mon projet de maîtrise se divisait en deux parties. La première consistait à déterminer l'organisation topologique du complexe de pré-initiation de la transcription (PIC) de l'ARNP II et la seconde visait à déterminer les partenaires protéiques de régulateurs interagissant avec la machinerie transcriptionnelle.

La réalisation de la première partie de mon projet a permis de déterminer le positionnement des protéines du complexe de pré-initiation de la transcription (PIC) relativement à l'ADN du promoteur. Pour ce faire, une méthode de photopontage en gel, permettant de ponter les protéines à des sites spécifiques sur l'ADN, a été utilisée. Ces expériences ont permis d'obtenir un modèle topologique de la trajectoire de l'ADN dans la structure de l'ARNP II et des protéines composant le PIC.

La deuxième partie de mon projet visait l'étude des interactions protéine-protéine de complexes transcriptionnels à l'échelle du génome dans des cellules de mammifères. Récemment, le laboratoire de Benoit Coulombe a utilisé une méthode de purification des protéines par marquage d'affinité avec une étiquette TAP (*Tandem Affinity Peptide*) pour purifier tous les facteurs généraux de transcription (FGTs) humains dans des conditions natives. Cette méthode permet de purifier les FGTs associés à des partenaires d'interactions additionnels, dont XAB1 et DDX5. Afin de confirmer la présence de ces polypeptides dans les complexes transcriptionnels et de tenter d'identifier de nouveaux partenaires protéiques, ces peptides ont été à leur tour étiquetés et purifiés par la méthode TAP. Ces études ont permis de révéler plusieurs d'informations sur les mécanismes transcriptionnels.

Mots Clés : Transcription, ARNP II, facteurs généraux de transcription, topologie, complexe de pré-initiation, photopontage en gel, EMSA, TAP, protéomique

## ABSTRACT

RNA polymerase (RNAP) II is the enzyme that synthesizes all eukaryotic mRNA. My master's project is divided in two parts. The aim of the first part was to determine the topology of the pre-initiation complex (PIC) and the second part focused on the determination of interacting partners of regulatory proteins of the transcription apparatus.

We first resolved the position of the proteins that compose the PIC relatively to the promoter DNA. To analyze the molecular organization of the PIC, a photo-cross-linking method was developed for the topological study of protein-DNA complexes on a native gel. The cross-linking results were used to determine the trajectory of the promoter DNA against the structure of the enzyme.

The second part of my project involved studying proteome-wide protein-protein interactions of regulatory proteins of the transcription apparatus. Recently, Benoit Coulombe and co-workers have used the Tandem Affinity Peptides (TAP) method for the purification of all human general transcription factors (GTFs) in native conditions. Compared to the GTFs purified with classical chromatography methods, the GTFs affinity purified using TAP tagging contained additional peptides. Among these peptides, we have found XAB1 and DDX5. To confirm the presence of these polypeptides in transcription complexes and to find new interacting partners, these polypeptides have also been tagged and purified using the TAP method. These studies have enlightened our understanding of the transcriptional mechanism.

Keys Words : Transcription, RNA polymerase II, general transcription factors, topology, pre-initiation complex, crosslinking in gel, EMSA, TAP-tag, Proteomics

## TABLE DES MATIÈRES

<b>RÉSUMÉ</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>iv</b>
<b>TABLE DES MATIÈRES</b>	<b>v</b>
<b>LISTE DES TABLEAUX</b>	<b>vii</b>
<b>LISTE DES FIGURES</b>	<b>viii</b>
<b>LISTE DES ABRÉVIATIONS</b>	<b>xi</b>
<b>REMERCIEMENTS</b>	<b>xiv</b>
<b>1- INTRODUCTION</b>	<b>1</b>
<b>1.1- La chromatine : un obstacle transcriptionnel</b>	<b>1</b>
<b>1.2- Les séquences régulatrices des gènes de classe II</b>	<b>2</b>
1.2.1- Le promoteur basal	3
1.2.1.1- La boîte TATA	4
1.2.1.2- L'initiateur (Inr)	5
1.2.1.3- Le DPE ( <i>downstream promotor element</i> )	5
1.2.1.4- Le BRE ( <i>TFIIB recognition element</i> )	6
1.2.2- Les séquences régulatrices	7
<b>1.3- Les facteurs transcriptionnels</b>	<b>7</b>
1.3.1- Les facteurs généraux de transcription	8
1.3.1.1- TBP	8
1.3.1.2- TFIIA	11
1.3.1.3- TFIIB	13
1.3.1.4- TFIIE	14
1.3.1.5- TFIIF	15
1.3.1.6- TFIIH	16
1.3.2- Les régulateurs transcriptionnels liant l'ADN	18
1.3.3- Les co-régulateurs	20
1.3.3.1- Le médiateur	20
1.3.3.2- TFIID	22
1.3.3.3- Les remodeleurs de la chromatine	23
1.3.3.4- Les histones acétyltransférases et les histones déacétylases	24
1.3.3.6- Autres co-régulateurs transcriptionnels	26
<b>1.4- L'ARNP II : structure Vs fonction</b>	<b>26</b>
1.4.1- Le domaine carboxy-terminal de l'ARNP II (CTD)	34
1.4.2- Le phénomène d'enroulement de l'ADN du promoteur autour de l'ARNP ( <i>wrapping</i> )	36
<b>1.5- Le cycle de transcription par l'ARNP II</b>	<b>41</b>
1.5.1- Pré-initiation de la transcription : l'assemblage séquentiel Vs l'holoenzyme	42
1.5.2- Initiation	43
1.5.3- Dégagement du promoteur	44
1.5.4- Élongation et terminaison	44
<b>1.6- Études des interactions protéine-ADN</b>	<b>45</b>
1.6.1- Les méthodes utilisées pour étudier les interactions protéine-ADN	46
1.6.1.1- Le gel de retardement de bandes (EMSA)	46
1.6.1.2- Les analyses d'empreinte à la DNase	46
1.6.1.3- Le photopontage protéine-ADN site spécifique	47

1.6.1.4- L'immunoprécipitation de la chromatine (CHIP)	47
1.6.2-Objectif de la première partie	48
<b>1.7- Études des réseaux d'interactions protéine-protéine</b>	<b>48</b>
1.7.1- Les principales méthodes utilisées pour étudier les interactions protéine-protéine à large échelle chez les eucaryotes	50
1.7.1.1- La méthode du double-hybride	50
1.7.1.2- La purification d'affinité utilisant une double étiquette TAP ( <i>Tandem Affinity Peptide</i> )	51
1.7.2- XAB1 et DDX5, deux nouveaux partenaires d'interaction de l'ARNP II et de TFIIS	54
1.7.2.1- Utilisation de la méthode TAP pour la purification de la machinerie transcriptionnelle basale	54
1.7.2.2- XAB1 ( <i>XPA binding protein 1</i> ) ou MBDin	55
1.7.2.3- DDX5 : <i>DEAD (Asp-Glu-Ala-Asp) Box polypeptide 5</i> ou p68	56
1.7.3- Objectifs de la deuxième partie	57
<b>2- MATÉRIELS ET MÉTHODES</b>	<b>58</b>
<b>2.1- Le photopontage en gel</b>	<b>58</b>
2.1.1- But de la méthode	58
2.1.2- Synthèse de la photosonde	58
2.1.3- Gel de retardement de bandes (EMSA)	60
2.1.4- Photopontage	61
2.1.5- Identification des protéines pontées	61
<b>2.2- La méthode TAP (<i>Tandem Affinity Peptide</i>)</b>	<b>62</b>
2.2.1- But de la méthode	62
2.2.2- Construction d'un vecteur d'expression encodant la protéine d'intérêt (appât) fusionnée à une étiquette TAP	63
2.2.2.1- PCR	64
2.2.2.2- Digestion du produit de PCR	64
2.2.2.3- Ligation du vecteur et de l'ADNc d'intérêt	65
2.2.2.4- Transformation des bactéries compétentes	65
2.2.3- Transfection des cellules humaines EcR-293 : Méthode de phosphate de calcium	66
2.2.4- Purification des complexes par double chromatographie d'affinité	66
2.2.4.1- Extraction des protéines cellulaires totales	68
2.2.4.2- Passage sur la première colonne d'affinité contenant les billes d'IgG	68
2.2.4.3- Passage sur la deuxième colonne d'affinité contenant les billes de calmoduline	68
2.2.4- Identification des partenaires protéiques de la protéine appât par spectrométrie de masse	69
<b>3- RÉSULTATS</b>	<b>70</b>
<b>3.1- Analyse de la topologie du complexe TBP-TFIIB-TFIIIE-TFIIF-ARNP II en utilisant le photopontage en gel</b>	<b>70</b>
<b>3.2- Identification des partenaires d'interaction des protéines XAB1 et DDX5</b>	<b>76</b>
<b>4- DISCUSSION</b>	<b>82</b>
<b>4.1- Analyse topologique du PIC</b>	<b>82</b>
<b>4.2- Étude des protéines XAB1 et DDX5 dans le réseau d'interactions protéiques de la machinerie transcriptionnelle basale</b>	<b>87</b>
<b>5- CONCLUSION ET PERSPECTIVES</b>	<b>93</b>
<b>6- LISTE DE RÉFÉRENCES</b>	<b>98</b>



## LISTE DES TABLEAUX

Tableau I. Résumé des principales kinases du CTD et leurs caractéristiques

Tableau II. Les facteurs d'élongation de l'ARNP II

Tableau III. Protéines purifiées par la méthode TAP et indentifiées par spectrométrie de masse MS/MS en utilisant les appâts Rpb2, Rpb11 et RPAP1.

Tableau IV. Identification des partenaires protéiques de XAB1 par spectrométrie de masse MS/MS

## LISTE DES FIGURES

Figure 1. Le promoteur basal.

Figure 2. Co-cristal TBP-ADN (boîte TATA).

Figure 3. Fonctions des sous-unités de TFIIF.

Figure 4. Structure du complexe médiateur et de l'ARNP II.

Figure 5. Structure cristallographique de l'ARNP II à une résolution de 2.8 Å présentant le positionnement de chacune des sous-unités.

Figure 6. Structure de l'ARNP II présentant les différents modules de l'enzyme

Figure 7. Voie empruntée par l'ADN lors de la transcription par l'ARNP II.

Figure 8. Interactions entre l'ARNP II et l'hybride ADN-ARN dans la structure de l'enzyme en élongation.

Figure 9. Schéma des interactions de l'ARNP II avec TFIIS.

Figure 10. Exposition du paradoxe "l'empreinte est plus grande que le pied" et modèle d'enroulement de l'ADN autour de l'ARNP.

Figure 11. Microscopie électronique des PIC assemblés sur le promoteur AdMLP en présence de TBP, TFIIB, TFIIF, TFIIE et l'ARNP II.

Figure 12. Modèle d'enroulement de l'ADN autour de la machinerie transcriptionnelle induit par RAP74.

Figure 13. Résumé du cycle transcriptionnel de l'ARNP II.

Figure 14. Vision actuelle de l'expression génique impliquant le couplage des processus.

Figure 15. Représentation schématique du système d'expression inductible à l'ecdysone chez les mammifères.

Figure 16. Principe pour la formation de la photosonde radioactive intégrant le photonucléotide à des positions spécifiques sur le promoteur.

Figure 17. Assemblage du PIC sur la sonde radioactive et purification en gel natif.

Figure 18. Carte du plasmide PMZI.

Figure 19. Résumé de la purification par la méthode TAP.

Figure 20. Sondes utilisées pour la méthode de photopontage en gel.

Figure 21. Représentation schématique de la méthode de photopontage en gel des complexes de pré-initiation de la transcription par l'ARNP II.

Figure 22. Gel de retardement de bandes démontrant les complexes intermédiaires assemblés sur la sonde en utilisant différentes combinaisons des FGTs et l'ARNP II.

Figure 23. Les conformères du photonucléotide (AB-dUMP) au promoteur AdMLP.

Figure 24. Résumé des résultats de photopontage en gel des complexes de pré-initiation comprenant les facteurs TBP, TFIIB, TFIIF (RAP30 et RAP74), TFIIE (TFIIE $\alpha$  et TFIIE $\beta$ ) et l'ARNP II.

Figure 25. Exemple de gel SDS-PAGE identifiant la sous-unité Rpb5.

Figure 26. Stratégie utilisée pour l'identification des partenaires protéiques *in vivo* de XAB1 et DDX5.

Figure 27. Gel coloré par la méthode SYPRO Ruby présentant les partenaires protéiques (directs et indirects) de XAB1 identifiés par spectrométrie de masse.

Figure 28. Gel présentant les partenaires protéiques de DDX5 identifiés par spectrométrie de masse.

Figure 29. Immunobuvardage réalisé sur l'éluat de DDX5 purifié par la méthode TAP.

Figure 30. Modèle représentant la trajectoire empruntée par l'ADN dans la structure de l'enzyme ARNP II et des facteurs TBP et TFIIB dans le PIC.

Figure 31. Localisation topologique des facteurs TFIIF et TFIIE établie en fonction des résultats de photopontage en gel.

## LISTE DES ABRÉVIATIONS

ABF1	<i>Autonomously replicating sequence binding factor 1</i>
AdMLP	<i>Adenovirus type 2 major late promoter</i>
ACF	<i>ATP-dependent chromatin assembly and remodeling factor</i>
ADN	Acide desoxyribonucléique
ARN	Acide ribonucléique
ARNP	ARNP
ARNm	ARN messenger
ATP	Adénosine triphosphate
BRE	<i>TFIIB recognition element</i>
CAK	<i>Cyclin-activating kinase complex</i>
CBP/p300	<i>C/EBP-binding protein</i>
CBP	<i>Calmodulin binding peptide</i>
CCD	<i>Charged cluster domain</i>
ADNc	ADN complémentaire
Cdk7	<i>Cyclin-dependent kinase 7</i>
CHIP	<i>Chromatin immunoprecipitation</i>
CHRAC	<i>Chromatin accessibility complex</i>
CPSF	<i>Cleavage polyadenylation specificity factor</i>
CR1/2	<i>Complement Receptor 1/2</i>
CTD	<i>Carboxyl terminal domain</i>
DDX5	<i>DEAD (Asp-Glu-Ala-Asp) Box polypeptide 5</i>
DMAP1	<i>DNA methyltransféranse 1-associated protein</i>
DNMT1	<i>DNA methyltransferase 1</i>
DPE	<i>Downstream promoter element</i>
Dr	<i>D repressor</i>
DSIF	<i>DRB sensitivity-inducing factor</i>
DUF	<i>DNA unwinding factor</i>
EcR-293	<i>Ecdysone receptor - 293</i>
ELL	<i>Eleven-nineteen lysine-rich leukemia</i>

EMSA	<i>Electrophoretic mobility shift assay</i>
FACT	<i>Facilitates chromatin transcription</i>
FBS	<i>Fetal bovine serum</i>
FCP1	<i>Transcription factor IIF (TFIIF)-associated carboxyl-terminal domain (CTD) phosphatase 1</i>
PCR	<i>Polymerase chain reaction</i>
FTG	Facteurs généraux de transcription
FRET	<i>Fluorescence resonance energy transfer</i>
GRE	<i>Glucocorticoid-response element</i>
HBx	<i>Hepatitis B virus X protein</i>
HDAC	<i>Histone deacetylases</i>
HIR	<i>Homomeric interaction region</i>
HSP	<i>Heat shock protein</i>
Inr	Initiateur
IPO11	Importine 11
KDa	Kilodalton
LCR	<i>Locus control region</i>
M	Mitose
MAT1	Ménage à trois 1
MBD	<i>Methyl-binding domain</i>
MeCP1	<i>Methyl-CpG-binding proteins</i>
NELF	<i>Negative elongation factor</i>
NLS	<i>Nuclear localisation signal</i>
NMR	<i>Nuclear magnetic resonance</i>
NURF	<i>Nucleosome remodeling factor</i>
N3R-dUMP	5-[N-(p-azidobenzoyl)-3-aminoallyl]-déoxyuridine monophosphate
NTP	Nucléotides
pb	Paire de bases
PFD	<i>Prefoldin domain</i>
PIC	<i>Pre-initiation complex</i>

P-TEFb	<i>Positive transcription elongation factor b</i>
Pu	Purine
Py	Pyrimidine
RAP	<i>RNA polymerase II-associated proteins</i>
RMP	<i>Rpb5-mediated protein</i>
RPAP1	<i>RNA polymerase-associated protein 1</i>
RPL12	<i>Ribosomal protein L12</i>
RSC	<i>Remodels the structure of chromatin</i>
SAGA	Spt-Ada-Gcn5 acetyltransferase
SAP30	<i>Sin3 associated polypeptide p30</i>
SNF	<i>Sucrose non-fermenting</i>
SnRNA	<i>Small nuclear RNA</i>
SRB	<i>Suppressor of RNA polymerase B</i>
STAP1	<i>SKP2-associated alpha PFD1</i>
SWI	<i>Switching defective</i>
TAF	<i>TBP associated factors</i>
TBP	<i>TATA box binding protein</i>
TLP	<i>TBP-like protein</i>
TRF	<i>TBP-related factor</i>
UAS	<i>Upstream activating sequences</i>
UBLCP1	<i>Ubiquitin-like domain containing CTD phosphatase 1</i>
UbcM2	<i>E2-type ubiquitin-conjugating enzyme</i>
URS	<i>Upstream repressing sequences</i>
UV	Ultraviolet
VDRIP	<i>Vitamin D receptor interacting protein</i>
VP16	<i>Viral protein 16</i>
XAB1	<i>XPA binding protein 1</i>
XPA	<i>Xeroderma pigmentosum A</i>
XPB	<i>Xeroderma pigmentosum B</i>
XPD	<i>Xeroderma pigmentosum D</i>
YY1	<i>Yin and Yang 1</i>

## REMERCIEMENTS

Je tiens sincèrement à remercier mon directeur de recherche, Benoit Coulombe, pour son encadrement, son soutien et ses conseils tout au long de ma maîtrise. Il m'a permis d'améliorer mes connaissances et de progresser dans un environnement de travail agréable. Je souhaite également remercier Diane Forget et Célia Jérónimo pour leur aide au cours de mon apprentissage pratique, leurs conseils et leur support moral. Je remercie Dominique Bergeron pour ses conseils et son aide précieuse en bioinformatique et Diane Bourque pour l'infographie. J'aimerais aussi remercier tous les gens du laboratoire avec qui j'ai aimé travailler et échanger. Finalement, merci du fond du cœur à mon conjoint, Olivier Bodard, qui m'a soutenu et appuyé durant mes années d'étude.



# 1- INTRODUCTION

## ***1.1- La chromatine : un obstacle transcriptionnel***

L'association entre l'ADN eucaryote et certaines protéines spécifiques constitue la chromatine. Les principales protéines de la chromatine sont les histones, petites protéines très riches en acides aminés basiques favorisant la liaison aux charges négatives de la molécule d'ADN. Il existe cinq principales histones, nommées H1, H2A, H2B, H3 et H4, variant à peine d'une espèce à l'autre (Cooper, 1999).

Le module de base de la chromatine est le nucléosome, décrit par Roger Kornberg en 1974. Le modèle proposé par Kornberg s'appuie sur deux types d'expériences. La première implique une digestion partielle de la chromatine par la nucléase de *micrococcus*. Cette digestion a permis d'obtenir des fragments d'ADN d'une longueur de 200 pb. Par contre, une digestion semblable avec de l'ADN nu a donné un dégradé de fragments de toutes les tailles. Ces résultats suggéraient que la fixation d'une ou de plusieurs protéines sur la molécule d'ADN protège cette dernière de l'attaque des nucléases (Kornberg, 1974). De plus, des images par microscopie électronique ont permis de montrer que la chromatine est constituée de plusieurs modules d'environ 200 pb appelés nucléosomes.

L'enroulement de l'ADN autour des histones produit une fibre de chromatine comprenant des nucléosomes de 10 nm de diamètre, séparés par des tronçons d'ADN d'environ 80 pb. La chromatine se condensera ensuite en se pelotonnant en fibres de 30 nm. Il semble que l'histone H1 joue un rôle important à cette étape de la condensation de la chromatine (Cooper, 1999). Le degré de condensation varie au cours du cycle cellulaire. Pendant l'interphase, la chromatine, appelée euchromatine, est assez bien décondensée et dispersée dans le noyau. À ce moment, les gènes sont activement transcrits et l'ADN se réplique pour la prochaine mitose. Contrairement à l'euchromatine, environ 10% de la chromatine

interphasique conserve un état très condensé appelé l'hétérochromatine (Cooper, 1999).

L'empaquetage des promoteurs dans la chromatine est un facteur influençant la transcription. Effectivement, les premières expériences de transcription *in vitro* utilisant des matrices de chromatine ont démontré que les nucléosomes pouvaient inhiber l'initiation de la transcription (Lorch et al., 1987). Par contre, une pré-incubation de l'ADN avec la machinerie transcriptionnelle avant la formation du nucléosome permet la transcription. À l'appui, des analyses *in vivo* supportent aussi ce modèle car la déplétion de certaines sous-unités d'histones chez la levure augmente la transcription de certains gènes (revu par Lee et Young, 2000). En somme, l'empaquetage de l'ADN procure un contexte physique différent pour chacun des promoteurs, ce qui permet une régulation spécifique des différents gènes.

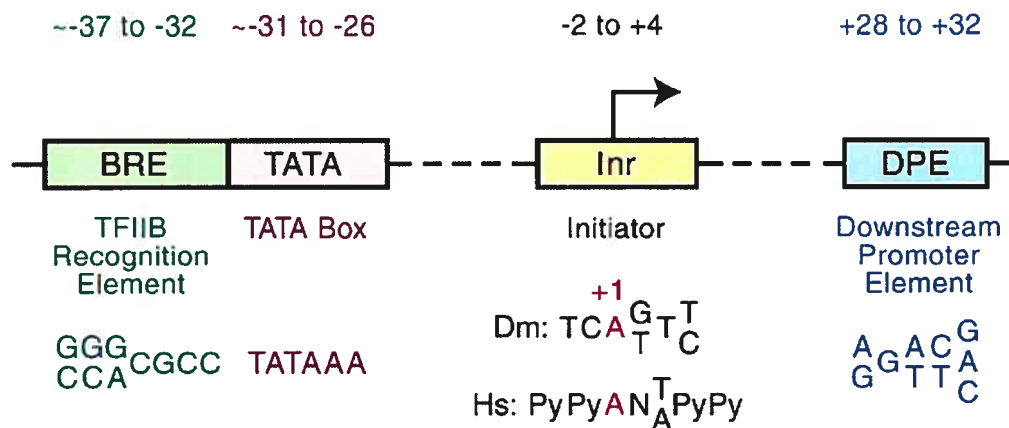
### ***1.2- Les séquences régulatrices des gènes de classe II***

Chez les eucaryotes, trois classes de gènes sont définies par le type d'enzyme ARNP qui les transcrit. L'ARNPI catalyse la transcription des gènes ribosomiaux, l'ARNP II permet la transcription des gènes codant pour les protéines et certains petits ARNs nucléaires (*snRNA*) et l'ARNP III transcrit essentiellement les gènes codant pour les ARNs de transfert et les ARNs 5S (revu par Zawel et Reinberg, 1995; Archambault et Friesen, 1993).

Les gènes de classe II requièrent certaines séquences particulières sur l'ADN pour permettre la transcription. Ces séquences peuvent être regroupées en deux classes, soit le promoteur basal et les séquences régulatrices.

### 1.2.1- Le promoteur basal

Selon une définition communément acceptée, le promoteur basal représente la région minimale continue d'ADN qui est suffisante pour permettre l'initiation de la transcription par la machinerie de l'ARNP II (revu par Butler et Kadonaga, 2002). Il existe plusieurs motifs retrouvés dans les séquences du promoteur basal incluant la boîte TATA, l'initiateur (Inr), l'élément DPE et l'élément BRE. Ces éléments sont montrés à la figure 1.



Tiré de Butler et Kadonaga, 2002

**Figure 1. Le promoteur basal.** Celui-ci contient les éléments contribuant à la transcription basale. Il est important de noter que chacun des motifs identifiés se retrouve seulement dans certains promoteurs, selon des combinaisons variables.

Il est à noter que les facteurs de transcription seront abordés en détail dans la section 1.3, mais des renseignements sur quelques facteurs pouvant être reliés directement à l'élément de séquence discuté, seront brièvement présentés dans cette section.

### 1.2.1.1- La boîte TATA

La boîte TATA est le premier élément du promoteur basal qui a été découvert dans les gènes eucaryotes. En 1979, une étude comparative des séquences flanquantes en 5' de plusieurs gènes de drosophiles et de mammifères a permis de découvrir la boîte TATA située de 25 à 30 pb en amont du site d'initiation de la transcription (revu par Smale et Kadonaga, 2003). Le développement des systèmes de transfection et des essais de transcription *in vitro* a permis de démontrer que des mutations dans la séquence TATAAA réduisent ou anéantissent souvent l'activité des promoteurs (Hu et Manley, 1981). Chez la levure *Saccharomyces cerevisiae*, l'élément TATA semble aussi critique pour la transcription. Par contre, chez cet organisme, l'élément TATA est situé de 40 à 120 pb en amont du site d'initiation de la transcription (revu par Smale et Kadonaga, 2003). Dans les premières études sur la boîte TATA, il était spéculé que cet élément était essentiel pour l'initiation de la transcription de tous les gènes chez tous les organismes eucaryotes. Par contre, l'avancement des études et des recherches sur plusieurs gènes a permis d'évaluer une prévalence beaucoup plus faible des éléments TATA. En effet, en 2001, une étude d'analyse de banques de données des gènes humains a révélé que l'élément TATA n'était présent qu'à 32% dans les 1031 régions promotrices potentielles étudiées (Suzuki et al., 2001).

La séquence 5'- TATATAAG - 3' semble optimale pour la liaison de TBP (voir section 1.3.1.1) (Wong et Bateman, 1994). Par contre, des études révèlent qu'une variété de séquences riches en A/T peuvent fonctionner de la même manière que la boîte TATA et interagir avec TBP (revu par Smale et Kadonaga, 2003). Des co-cristaux du complexe TBP-ADN ont été produits avec 10 différentes séquences TATA et les résultats révèlent que TBP induit un changement de conformation similaire pour chacune des séquences TATA examinées (Patikoglou et al., 1999).

### 1.2.1.2- L'initiateur (Inr)

L'Inr est retrouvé dans les promoteurs contenant ou non une boîte TATA. La séquence consensus de l'Inr chez les mammifères est Py-Py A+1 -N-T/A-Py-Py. La position A+1 est le site d'initiation de la transcription +1 car la transcription commence souvent à ce nucléotide (revu par Butler et Kadonaga, 2002). Des études de l'espacement entre la boîte TATA et l'Inr ont démontré que les deux éléments agissent de manière synergique lorsqu'ils sont séparés de 25 à 30 pb et, de manière distincte lorsqu'ils sont séparés de plus de 30 pb (O'Shea-Greenfield et Smale, 1992).

Plusieurs facteurs interagissent avec l'Inr. Des évidences permettent de croire que le facteur TFIID (voir section 1.3.2.4) lie l'Inr de manière séquence spécifique. Plus précisément, les sous-unités TAF1 et TAF2 sont capables d'interagir avec l'Inr (Chalkley et Verrijzer, 1999). La synergie entre la boîte TATA et l'Inr corrèle avec la liaison combinée de ces deux éléments par TFIID. Il a aussi été observé que l'ARNP II purifiée et accompagnée de certains facteurs (TBP, TFIIB et TFIIF) est capable de reconnaître l'Inr et, ainsi permettre la transcription de manière Inr dépendante en absence des TAFs (Weis et Reinberg, 1997).

Deux facteurs liant l'ADN, TFII-I et YY1, interagissent avec l'Inr. Le facteur TFII-I est une protéine avec un motif bHLH (*basic-helix-loop-helix*) qui lie l'élément Inr et stimule la transcription. TFII-I participe à la communication entre les facteurs liant l'ADN et la machinerie transcriptionnelle. YY1 est une protéine à doigt de zinc liant aussi le promoteur basal au niveau de la séquence Inr. YY-1 interagit directement avec la machinerie transcriptionnelle (revu par Butler et Kadonaga, 2002).

### 1.2.1.3- Le DPE (*downstream promotor element*)

Le DPE est un élément fortement conservé de la drosophile à l'humain et est typiquement présent (mais pas toujours) dans les promoteurs sans boîte TATA. Le

DPE agit en conjonction avec l'Inr et sa séquence est localisée aux positions +28 à +32 relativement au nucléotide A+1 de l'Inr (revu par Smale et Kadonaga, 2003). Une mutation dans l'élément Inr ou l'élément DPE résulte en la perte de liaison de TFIID au promoteur basal (Burke et Kadonaga, 1996). L'espacement entre le DPE et l'Inr est identique pour tous les promoteurs et l'altération de l'espacement de seulement un nucléotide entre ces deux éléments résulte en une diminution sévère de la liaison par TFIID et, conséquemment, une diminution de la transcription basale (Kutach et Kadonaga, 2000). De plus, des études chez la drosophile effectuée sur des promoteurs sans boîte TATA montrent qu'une mutation dans le motif DPE produit une diminution de 10 à 50 fois de la transcription basale (Burke et Kadonaga, 1996). La séquence consensus pour le DPE est A/G+28 -G-A/T-C/T-G/A/C (revu par Butler et Kadonaga, 2002). Même si l'élément DPE a surtout été étudié chez la drosophile, on retrouve aussi des éléments DPE chez l'humain (Zhou et Chiang, 2001).

#### **1.2.1.4- Le BRE (*TFIIB recognition element*)**

Le BRE est un site de liaison de TFIIB qui est souvent localisé en amont de la boîte TATA (voir figure 1). TFIIB lie directement le BRE de manière séquence spécifique. La séquence consensus de BRE est G/C-G/C-G/A-C-G-C-C, où le C en 3' est suivi du T en 5' de la boîte TATA. Des essais de transcription *in vitro* avec les facteurs purifiés révèlent que le BRE facilite l'incorporation de TFIIB dans des complexes d'initiation de la transcription (Lagrange et al., 1998). Par contre, d'autres études utilisant des essais de transcription constitués à partir d'extraits totaux montrent que le BRE a un effet négatif sur la transcription basale (revu par Butler et Kadonaga, 2002).

### **1.2.2- Les séquences régulatrices**

En plus du promoteur basal, d'autres séquences sont essentielles pour réguler la transcription par l'ARNP II. Les séquences régulatrices servent de sites de liaison pour les activateurs ou les répresseurs. Parmi celles-ci on retrouve; les UASs (*upstream activating sequences*), les *enhancers*, les URSs (*upstream repressing sequences*), les *silencers* et les LCRs (*Locus control region*) (revu par Lee et Young, 2000). Les UASs sont des séquences liées par des activateurs qui influencent la transcription positivement. Cependant, elles ne peuvent agir que lorsqu'elles sont situées en amont de la boîte TATA (Guarente et Hoar, 1984). Les *enhancers* sont aussi liés par des activateurs et influencent la transcription. Par contre, contrairement au UASs, ils peuvent agir sur de longues distances, et ce, peu importe leur orientation (Blackwood et Kadonaga, 1998). Les URSs lient des facteurs inhibant la transcription. Ces facteurs peuvent interférer avec la liaison des activateurs, prévenir le recrutement de la machinerie transcriptionnelle par l'activateur et modifier la structure de la chromatine (Struhl, 1995). Les *silencers*, pour leur part, sont des séquences pouvant réprimer l'activité d'un promoteur indépendamment de leur orientation ou de leur position. L'ADN du domaine silencieux semble être inhibé au niveau transcriptionnel à la suite d'interactions entre des histones et certaines protéines qui agissent de manière coopérative pour couvrir la région et possiblement cibler la désacétylation (revu par Lee et Young, 2000). Finalement les LCRs sont semblables aux *enhancers*, mais sont souvent arrangés en complexes comprenant plusieurs éléments régulateurs. Les LCRs stimulent la transcription indépendamment de leur site d'intégration dans la chromatine et leur effet est limité à leur orientation et leur distance (Bulger et Groudine, 1999). Les *enhancers*, les LCRs et les *silencers* peuvent augmenter ou diminuer l'expression de différents gènes dans le génome.

### ***1.3- Les facteurs transcriptionnels***

Les facteurs transcriptionnels peuvent être divisés en 2 catégories; les facteurs généraux et les régulateurs transcriptionnels.

### 1.3.1- Les facteurs généraux de transcription

À la fin des années '70 et au début des années '80, les premiers essais de transcription ont permis de démontrer l'incapacité de l'ARNP II purifiée à reconnaître les séquences promotrices et à initier adéquatement la synthèse des ARNm. Ces résultats suggéraient la présence de facteurs additionnels nécessaires au phénomène transcriptionnel. Par la suite, plusieurs facteurs essentiels à l'initiation de la transcription ont été découverts et décrits par plusieurs laboratoires. Ces facteurs sont TFIIA, TFIIB, TFIID (TBP), TFIIE, TFIIF et TFIIH. Nommées les facteurs généraux de transcription, ces protéines s'assemblent au niveau de l'ADN du promoteur et recrutent l'ARNP II au site d'initiation. Chacun de ces facteurs sera traité en détails ci-dessous.

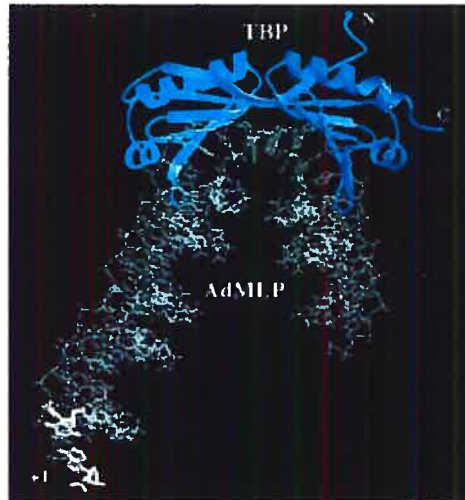
#### 1.3.1.1- TBP

Le facteur TBP (*TATA box binding protein*) a été nommé en fonction de sa propriété à lier la boîte TATA. Chez la levure, TBP est un monomère de 27 kDa qui est fonctionnellement interchangeable avec le facteur TBP des mammifères dans des systèmes de transcription *in vitro* (Cavallini et al., 1989; Buratowski et al., 1988). Ces résultats démontrent que le facteur de liaison à la boîte TATA est fortement conservé chez les organismes eucaryotes. Chez l'humain, TBP est une protéine de 38 kDa (Jupp et al., 1993). TBP a aussi été identifié comme étant un composant d'autres complexes contribuant à l'initiation de la transcription par les ARNPI et III (revu par Smale et Kadonaga, 2003).

*In vivo*, TBP est associé à d'autres protéines, nommées les TAFs, formant le facteur TFIID. Les TAFs, ont été clonés et certaines de leurs activités biologiques ont été définies. Parmi celles-ci : La reconnaissance du promoteur basal, une activité acétyltransférase, une activité kinase, des activités d'ubiquitination et de conjugaison et une fonction de co-activateurs par des interactions protéine-protéine avec des activateurs transcriptionnels (revu par Smale et Kadonaga, 2003). Les TAFs de TFIID seront décrits plus en détails dans la section 1.3.3.2.



Le clonage de TBP a été facilité par la purification du polypeptide chez la levure *Saccharomyces cerevisiae*. La séquence de TBP chez la levure a permis l'identification de la séquence codante chez plusieurs eucaryotes (revu par Smale et Kadonaga, 2003). Le co-cristal de TBP et de la boîte TATA a permis de déterminer de manière détaillée le mécanisme de liaison à l'ADN. La figure 2 illustre le co-cristal TBP-ADN. Il est à noter que la structure tridimensionnelle de la portion conservée de TBP est souvent comparée à une selle d'équitation (Nikolov et Burley, 1994; Smale et Kadonaga, 2003). Cette selle moléculaire est formée de deux domaines fortement symétriques contenant chacun 89 à 90 acides aminés. La portion N-terminale de TBP fait contact avec la partie 3' de la boîte TATA alors que sa portion en C-terminale se localise près de de la partie en 5'. Chaque moitié de la surface concave de la selle est formée de cinq feuillets- $\beta$  anti-parallèles. Huit de ces feuillets font contact avec le sillon mineur de l'ADN. La liaison au sillon mineur implique plusieurs interactions hydrophobes (Nikolov et al., 1996). TBP induit aussi une torsion de l'ADN aux extrémités 3' et 5' de la boîte TATA et désapparie partiellement l'ADN (Nogales, 2000).



Tiré de Nikolov et Burley, 1997

**Figure 2. Co-cristal TBP-ADN (boîte TATA).** TBP est illustré en bleu et adopte une forme de selle. Il induit une courbure de 80° de la boîte TATA sur le promoteur AdMLP.

Dans le co-cristal TBP-ADN, TBP est lié de manière orientée à la boîte TATA. La liaison polaire de TBP pourrait conduire à l'assemblage du complexe de pré-initiation (PIC) de façon orientée pour influencer la direction de la transcription (Kim et al., 1993). Par contre, même si TBP lie la boîte TATA dans une même orientation dans le co-cristal TBP-ADN, TBP peut lier l'élément TATA dans les deux orientations lorsqu'il est en solution, avec une faible préférence dans l'orientation retrouvée en co-cristal (Cox et al., 1997). Lorsque des promoteurs plus complexes sont utilisés, le déterminant majeur de l'orientation semble être la portion des sites de liaison d'activateurs, la boîte TATA et d'autres éléments du promoteur basal (Xu et al., 1991). Lorsque l'orientation de la boîte TATA est inversée sur un promoteur contenant des sites de liaisons d'activateurs distaux, la direction de la transcription n'est pas inversée (revu par Smale et Kadonaga, 2003). Des études récentes ont démontré que des activateurs distaux peuvent influencer la polarité de TBP, ce qui peut constituer un mécanisme dominant pour déterminer la direction de la transcription (Kays et Schepartz, 2002)

Le recrutement de TBP au promoteur génique est une étape limitante pour l'activation de la transcription. En fait, une part importante de la régulation du processus transcriptionnel est le retrait d'inhibiteurs associés au promoteur ou à la machinerie transcriptionnelle. L'accès de TBP au promoteur génique peut-être prévenue par la formation du nucléosome sur la boîte TATA, par certaines protéines inhibitrices incluant TAF1, NC2, Mot1, Spt3/Spt8, le complexe Ccr4-Not et par une deuxième molécule de TBP amenant la formation d'un homodimère (Kou et al., 2003).

En 1993, Crowley *et al.* ont démontré l'existence d'une protéine homologue à TBP chez la drosophile. Cette protéine, TRF (*TBP-related factor*), est structurellement similaire à TBP et est principalement impliquée dans l'expression des gènes du système nerveux (Crowley et al., 1993). À la suite de cette découverte, des chercheurs ont tenté d'identifier des orthologues de TRF chez les eucaryotes supérieurs. Chez la souris, TLP (*TBP-like factor*) a été identifié comme l'équivalent de TRF (Ohbayashi et al., 1999). TLP possède un rôle de régulateur positif sur quelques gènes (Nakadai et al., 2004).

### 1.3.1.2- TFIIA

Le facteur de transcription TFIIA est encodé par deux gènes. La petite sous-unité s'appelle TFIIA $\gamma$  et la plus grande TFIIA $\alpha\beta$  (Ranish et al., 1992). Cette dernière est clivée en deux polypeptides nommés TFIIA $\alpha$  et TFIIA $\beta$  (DeJong et Roeder, 1993). Des études récentes suggèrent que le clivage de TFIIA $\alpha\beta$  est un processus régulé et que les formes clivées et non-clivées de TFIIA ont des fonctions cellulaires distinctes (Hoiby et al., 2004). La localisation du site de clivage a été identifiée et une mutation dans ce dernier prolonge la demi-vie du précurseur TFIIA $\alpha\beta$ . En fait, le facteur TFIIA clivé serait un substrat pour la voie d'ubiquitination et de la dégradation par le protéasome. Ainsi, le clivage et

subséquemment la dégradation du facteur TFIIA permettrait une régulation fine de la quantité de ce facteur dans la cellule (Hoiby et al., 2004).

Il a été démontré que TFIIA favorise la transcription en interagissant avec TBP et en stabilisant sa liaison à l'ADN (Ranish et al., 1999). Aussi, TFIIA $\gamma$  possède des propriétés d'agent activateur en contrecarrant les facteurs régulant négativement TBP comme NC2/Dr1 et Dr2/PC3 (Ozer et al., 1994; DeJong et al., 1995), Mot1 (Auble et al., 1994) et TAF1 (Ozer et al., 1998b). TFIIA n'est pas requis pour la transcription basale dans des systèmes hautement purifiés utilisant TBP recombinant (sans les TAFs), mais il est requis dans des systèmes utilisant TFIID (Ma et al., 1993). L'association de TFIIA avec TFIID induit un changement de conformation dans les TAFs pour promouvoir des contacts additionnels avec le promoteur en amont de la boîte TATA (Oelgeschlager et al., 1996). La nécessité du facteur TFIIA en transcription est dépendante de la structure du promoteur basal et des activateurs transcriptionnels. Ces résultats sont consistants avec les rôles de TFIIA au niveau de la sélectivité du promoteur et la médiation des contacts avec les activateurs (Ozer et al., 1998a). La forme phosphorylée de TFIIA est trente fois plus efficace dans la formation d'un complexe stable avec TBP et la boîte TATA du promoteur d'ADN (Solow et al., 1999). Cette phosphorylation représenterait un important mécanisme pour moduler la transcription chez la levure.

Récemment, TFIIA a été retrouvé en complexe avec TLP (*TBP-like protein*). À la suite d'études biochimiques, il a été observé que TFIIA est requis pour la localisation cytoplasmique ou nucléaire de TLP et potentiellement pour la modulation de ses fonctions (Nakadai et al., 2004). TFIIA possède aussi un rôle important dans la dissociation des dimères de TBP, une étape limitante pour la liaison à l'ADN. La dissociation rapide des dimères de TBP et de TFIID permet de rendre ces facteurs disponibles pour la liaison à l'ADN (Coleman et al., 1999). Il a été démontré que TFIIA peut activer la transcription basale en stimulant les

fonctions de TFIIE34 et RAP74 durant l'étape d'initiation de la transcription (Langelier et al., 2001).

### 1.3.1.3- TFIIB

TFIIB est un polypeptide de 35 kDa chez l'humain (Ha et al., 1991). À la suite de la liaison de TBP au promoteur, TFIIB est recruté et semble jouer plusieurs rôles cruciaux dans la formation du PIC. Effectivement, TFIIB reconnaît le complexe TBP-TATA et forme un pont entre TBP et l'ARNP II. Des études ont démontré que TFIIB peut faire contact avec l'ADN du promoteur de manière séquence spécifique via un motif hélice-boucle-hélice. Il stabilise ainsi le complexe TFIIB-TBP-promoteur (Lagrange et al., 1998). Des expériences de FRET ont démontré qu'un changement de conformation du facteur TFIIB se produit lors de son entrée dans le complexe transcriptionnel. Ce changement de conformation et la vitesse de formation du PIC seraient augmentés en présence d'un activateur comme VP16. De plus, la formation du complexe TATA-TBP-TFIIB est influencée par la séquence promotrice (Zheng et al., 2004).

Au niveau structural, TFIIB est composé de deux domaines : un domaine C-terminal protéolytiquement stable et un domaine N-terminal contenant une structure en doigt de zinc (Bagby et al., 1995; Elsby et Roberts, 2004). Le côté N-terminal contient plusieurs résidus clés chargés et hautement conservés qui modulent les fonctions de TFIIB et que l'on nomme aussi le CCD (*charged cluster domain*). Les domaines N-terminal et C-terminal peuvent interagir ensemble et l'affinité entre ces deux domaines est modulée par le CCD. Des analyses de FRET indiquent que cette interaction permet la formation d'un équilibre entre les formes ouvertes et fermées de la protéine TFIIB (Zheng et al., 2004). Des mutations dans le CCD ont généré de plus fortes interactions intramoléculaires entre ces deux domaines (Elsby et Roberts, 2004). Aussi, des mutations dans la région du CCD de TFIIB modifient le positionnement du site d'initiation de la transcription de quelques nucléotides à des promoteurs de levure

et de mammifère (revu dans (Hampsey, 1998). Récemment, il a été démontré que les mutants de TFIIB responsables d'aberration dans la sélection du site d'initiation s'assemblent au promoteur dans une conformation différente que le TFIIB de type sauvage (Fairley et al., 2002).

#### 1.3.1.4- TFIIE

TFIIE humain est un hétérotétramère formé de deux sous-unités  $\beta$  (E34) et de deux sous-unités  $\alpha$  (E56) (Ohkuma et al., 1991). TFIIE recrute TFIIH au PIC et stimule les activités kinase et hélicase de ce dernier (Ohkuma, 1997). Le groupe de Robert Roeder a démontré qu'une forte phosphorylation de l'ARNP II est dépendante de TFIIE (Ohkuma et Roeder, 1994). Ainsi, TFIIE serait potentiellement impliqué dans le phénomène de dégagement du promoteur. De plus, TFIIE lie l'ARNP II et TFIIH, suggérant un rôle de pont entre ces deux protéines (Maxon et al., 1994).

Plusieurs évidences suggèrent un rôle de TFIIE au niveau de l'ouverture du promoteur. Premièrement, TFIIE lie l'ADN simple brin. Deuxièmement, le besoin en TFIIE peut être court-circuité en employant une séquence promotrice pré-ouverte. Finalement, des analyses de photopontage localisent TFIIE près de l'ADN et du centre actif de l'ARNP II (Meinhart et al., 2003).

TFIIE $\alpha$  comporte 439 acides aminés et la région N-terminale est essentielle pour la transcription basale (Ohkuma et al., 1995). Cette partie N-terminale présente de fortes homologies avec le facteur TFIIE $\alpha$  des archéobactéries (Bell et al., 2001). Des études récentes utilisant la mutagenèse ont permis de localiser du côté C-terminal de TFIIE $\beta$ , les résidus d'acides aminés responsables de la liaison à l'ARNP II et à l'ADN simple brin. De plus, ces mutants affectent la phosphorylation du domaine carboxy-terminal (CTD) de l'ARNP II par TFIIH (Watanabe et al., 2003).

### 1.3.1.5- TFIIF

TFIIF est un hétérotétramère composé de deux sous-unités RAP (*RNA polymerase II-associated proteins*) 30 et RAP 74. La mise en évidence de ce complexe hétérotétramérique est survenue à la suite d'une purification par filtration sur gel dans laquelle le facteur TFIIF présentait un poids apparent de 220 kDa (Flores et al., 1990). Toutefois, la structure identifiée par rayons-X montre TFIIF sous la forme d'un dimer seulement (Gaiser et al., 2000). Les deux sous-unités de TFIIF sont impliquées dans le recrutement stable de l'ARNP II (Flores et al., 1991; Killeen et Greenblatt, 1992) et sont nécessaires pour faciliter l'étape d'initiation de la transcription à partir de matrices d'ADN linéaires (Lei et al., 1998). Des études biochimiques ont démontré que TFIIF escorte l'ARNP II au promoteur et stabilise la formation du PIC (revu par Conaway et Conaway, 1993). Ce facteur de transcription est requis pour l'association subséquente de TFIIE et TFIID dans le PIC. Une fois le complexe assemblé, TFIIF facilite le dégagement de l'ARNP II du promoteur à la suite de la formation du premier lien phosphodiester du transcrit d'ARNm (Yan et al., 1999; Ghazy et al., 2004). TFIIF peut se lier directement à TFIIB (Ha et al., 1993; Fang et Burton, 1996), TFIID (Ruppert et Tjian, 1995; Dubrovskaya et al., 1996), TFIIE (Maxon et Tjian, 1994) et l'ARNP II (Flores et al., 1989). Ainsi, TFIIF orchestre l'organisation du PIC. Son rôle principal en élongation est de supprimer les pauses transitoires de l'ARNP II et stimuler l'élongation (Bengal et al., 1991). Jusqu'à ce jour, TFIIF a été retrouvé dans toutes les formes purifiées de l'holoenzyme ARNP II (revu par Greenblatt, 1997).

L'isomérisation du PIC est un processus progressif dans lequel des changements de conformation de l'ADN et des protéines se produisent avant l'initiation de la transcription. Des études de photopontage ont démontré que TFIIF joue un rôle fondamental dans ce processus. Ces études réalisées dans le laboratoire de Benoit Coulombe, en absence de RAP74 et en présence de mutants de RAP74, ont permis d'observer l'implication de TFIIF dans le phénomène d'enroulement de

l'ADN autour de l'ARNP II (Robert et al., 1998; Forget et al., 1997). Le phénomène d'enroulement sera expliqué plus en détail dans la section 1.4.2

Au niveau structural, la protéine RAP74 peut-être divisée en trois régions distinctes : un domaine N-terminal globulaire important pour la liaison de RAP30, l'assemblage du PIC et la stimulation de l'élongation; une région centrale fortement chargée dépourvue d'acides aminés hydrophobes; et finalement, un domaine C-terminal très basique avec une structure globulaire. Ce domaine fait contact avec TFIIB et l'ARNP II et stimule la protéine Fcp1, une phosphatase du CTD de l'ARNP II (Archambault et al., 1998). Pour sa part, RAP30, est composé d'un domaine N-terminal liant RAP74 et TFIIB (Fang et Burton, 1996), un domaine central agissant avec la sous-unité Rpb5 de l'ARNP II (Wei et al., 2001) et un domaine de liaison à l'ADN en C-terminal ayant des similarités avec le facteur sigma de l'ARNP procaryote (Wei et al., 2001). Aussi, des études de photopontage ont permis de suggérer que RAP30, pourrait agir comme une chaperonne moléculaire en catalysant l'assemblage l'ARNP II dans le PIC (Coulombe et al., 1994).

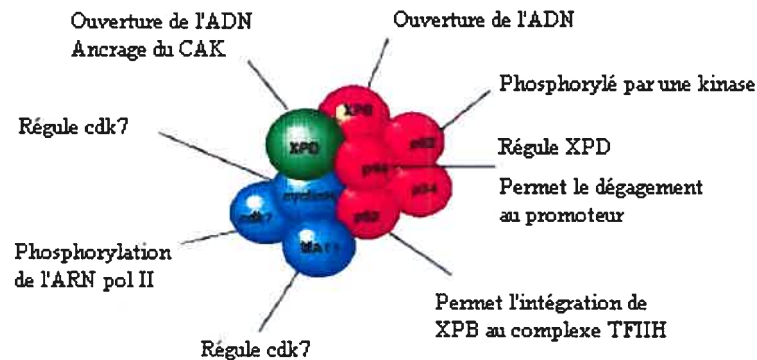
#### 1.3.1.6- TFIIH

Le facteur TFIIH est composé de 10 sous-unités divisées en deux sous-complexes : le complexe central et le complexe CAK (*cyclin-activating kinase complex*). Le domaine central est composé de cinq protéines : XPB (*xeroderma pigmentosum B*), p34, p52, p62 et p44. Le domaine CAK est formé de Cdk7 (*cyclin-dependent kinase 7*), Cyclin H et la protéine MAT1. XPD (*xeroderma pigmentosum D*) forme un lien entre le domaine central et le complexe CAK (revu par Zurita et Merino, 2003). Une dixième sous-unité, TFB5, a récemment été mise en évidence. Cette dernière est impliquée dans le trichothiodystrophie A, une maladie liée à un dysfonctionnement dans le système de réparation NER. La structure en trois dimensions de TFIIH dévoile une forme en anneau avec une cavité centrale interagissant possiblement avec l'ADN (Schultz et al., 2000).



Le facteur TFIIH est impliqué dans trois phénomènes cellulaires de première importance, soit la transcription, la réparation et le contrôle du cycle cellulaire. Au niveau de la réparation de l'ADN, TFIIH participe à la réparation par excision de nucléotide et la réparation de dommages oxydatifs (de Laat et al., 1999). Le sous-complexe CAK de TFIIH est responsable de la régulation du cycle cellulaire, notamment la transition G2 à M (Larochelle et al., 1998). Au niveau de la transcription, les rôles de TFIIH sont multiples. Au cours de l'initiation, TFIIH est responsable de la formation d'un complexe ouvert grâce à ses hélicases. L'activité hélicase retrouvée chez la sous-unité XPB est essentielle pour l'ouverture de l'ADN du promoteur. Par contre, cette activité peut être court-circuitée en utilisant des matrices d'ADN pré-ouvertes. L'hélicase de XPB est aussi requise pour la transition entre la production de transcrits avortés et l'élongation formant un ARNm complet (Bradsher et al., 2000). XPB prévient aussi les arrêts prématurés de l'ARNP II en élongation à des positions rapprochées du promoteur (revu par Lee et Young, 2000). À la suite de la formation du complexe ouvert, Cdk7 phosphoryle les 5 sérines du CTD de la plus grande sous-unité de l'ARNP II. Cette phosphorylation permettra à l'ARNP II de relâcher la plupart des facteurs généraux de transcription pour initier le processus d'élongation (revu par Zurita et Merino, 2003). La figure 3 présente les rôles de chacune des sous-unités de TFIIH.

### Fonctions des sous-unités de TFIIH



Tiré de Egly, 2001

**Figure 3. Fonctions des sous-unités de TFIIH.** La dixième sous-unité n'est pas représentée sur cette figure.

Des mutations dans les hélicase XPB et XPD sont responsables de maladies génétiques sévères chez les humains incluant le *Xeroderma pigmentosum*, le syndrome de Cockayne et la Trichothiodystrophie (revu (Egly, 2001)).

#### 1.3.2- Les régulateurs transcriptionnels liant l'ADN

La régulation de l'expression des gènes implique généralement une balance entre l'action des activateurs et des répresseurs. Typiquement les activateurs sont formés de deux domaines : un domaine de liaison à l'ADN et un domaine qui recrute ou stimule les activités de la machinerie transcriptionnelle. Les activateurs transcriptionnels peuvent recruter des complexes de modification de la chromatine comme Swi/Snf ou SAGA au promoteur (revu par Lee et Young, 2000). Les activateurs peuvent aussi lier et recruter des composants de la machinerie transcriptionnelle (revu par Lee et Young, 2000).

Une méthode permettant de photoponter les protéines à des sites spécifiques sur l'ADN a été utilisée dans le laboratoire de Benoit Coulombe afin de montrer qu'un activateur lié à des séquences spécifiques de reconnaissance en amont du promoteur basal peut interagir directement avec le complexe TBP-TFIIB-TFIIA situé au promoteur basal. Cette interaction implique la formation d'une boucle dans l'ADN (Dion et Coulombe, 2003). Dans cette même étude, il a été montré que l'interaction entre l'activateur et la machinerie transcriptionnelle modifie la topologie du complexe TBP-TFIIA-TFIIB, mais ne change pas l'efficacité de formation du complexe. Un autre rôle des activateurs est d'influencer l'activité de la machinerie de transcription. Quelques activateurs peuvent aussi stimuler plusieurs cycles de transcription pour des gènes hautement exprimés en facilitant la réinitiation. Un intermédiaire de réinitiation formé de TFIID, TFIIA, TFIIF, TFIIE et du médiateur a été identifié. Cet intermédiaire est dépendant de l'ATP et est stabilisé par l'activateur Gal4-VP16 (Yudkovsky et al., 2000).

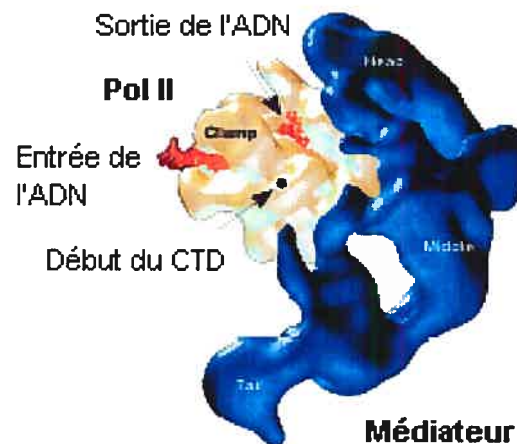
Les répresseurs peuvent être regroupés en deux classes : les répresseurs généraux et les répresseurs gènes spécifiques (Lee et Young, 2000). Les répresseurs transcriptionnels généraux sont génétiquement définis par des mutations causant une augmentation de la transcription de plusieurs gènes fonctionnellement non-reliés. Les répresseurs généraux lient l'ADN peuvent interagir avec la machinerie transcriptionnelle basale (souvent TBP) ou avec les nucléosomes (Hampsey, 1998). Par exemple, Mot1, lie les complexes TBP-ADN et cause la dissociation de TBP de l'ADN de manière ATP-dépendante. Pour leur part, les répresseurs gènes spécifiques fonctionnent en se liant à des activateurs ou en compétitionnant pour les sites de liaison d'activateurs. Par exemple, Hsp90 lie l'activateur transcriptionnel Hsf1 et prévient la formation d'un trimère Hsf1 nécessaire à la liaison à l'élément de réponse "heat shock" (Lee et Young, 2000). Les répresseurs lient l'ADN peuvent aussi recruter les histones désacétylases pour réprimer la transcription via leur action sur la chromatine (voir section 1.3.3.4)

### **1.3.3- Les co-régulateurs**

Certaines protéines ou complexes protéiques peuvent agir de manière indirecte sur la transcription en permettant aux activateurs de réaliser leurs fonctions (le médiateur et TFIID) ou en changeant l'accessibilité de la chromatine (les remodeleurs de la chromatine, les histones acétyltransférase et les histones desacétylases).

#### **1.3.3.1- Le médiateur**

Le médiateur est un complexe protéique qui a été isolé chez la levure sur la base de sa nécessité pour l'activation transcriptionnelle dans des systèmes reconstitués (Hampsey, 1998). Le médiateur peut jouer divers rôles dans la régulation de la transcription chez la levure. En plus de stimuler la transcription basale de 10 fois, le complexe médiateur peut stimuler de 30 fois la réponse aux activateurs. De surcroît, le médiateur provoque une augmentation de 30 à 50 fois de la phosphorylation du CTD de Rpb1 par le facteur TFIIF (Davis et al., 2002). Contrairement à l'ARNP II et les FGTs, le complexe médiateur n'est pas capable de lier des séquences spécifiques d'ADN. Par contre, ce dernier peut interagir physiquement avec l'ARNP II en liant son CTD (Woychik et Hampsey, 2002). Cette interaction forme l'holoenzyme, dont la structure est montrée à la figure 4. La phosphorylation du CTD joue un rôle important au niveau de l'interaction entre le médiateur et l'ARNP II. Effectivement, une étude réalisée en 1999 a permis de démontrer que l'ablation du CTD de Rpb1, prévient l'association entre le complexe médiateur et ARNP II *in vitro* (Asturias et al., 1999).



Tirée de Davis et al, 2002

**Figure 4. Structure du complexe médiateur et de l'ARNP II.**

Le médiateur, en bleu, est formé des domaines *head*, *middle* et *tail*. L'ARNP II, en beige, lie l'ADN (orange). Le point noir représente la localisation du CTD.

Les états hypophosphorylé et hyperphosphorylé du CTD de la sous-unité Rpb1 de l'ARNP II ont permis de suggérer un cycle de recyclage du complexe médiateur. En effet, il a été observé que l'ARNP II est hypophosphorylée lorsqu'elle est en complexe avec le médiateur alors que lors de l'élongation, elle possède un CTD fortement phosphorylé et n'est plus associée au complexe médiateur. Ces résultats ont permis de suggérer un modèle dans lequel la phosphorylation du CTD par TFIIF permet la dissociation du médiateur de l'ARNP II. Le médiateur libéré du complexe peut ensuite lier les ARNP II libres et le cycle se répète (Davis et al., 2002).

### 1.3.3.2- TFIID

En 1989, plusieurs groupes ont purifié et cloné le facteur de liaison à la boîte TATA (TBP) chez la levure. Le clonage de TBP chez la drosophile et l'humain a été décrit l'année suivante (revu par Albright et Tjian, 2000). À l'origine, on croyait que TBP et TFIID étaient identiques, mais des expériences de transcription *in vitro* ont rapidement contrecarré cette hypothèse. Effectivement, il a été observé que TFIID natif, mais pas TBP recombinant, peut supporter la stimulation de la transcription par un activateur dans des systèmes purifiés. Ceci a conduit à la proposition, et subséquemment à la démonstration, que les polypeptides associés à TBP sont essentiels pour médier l'action des activateurs dans des systèmes reconstitués (revu par Albright et Tjian, 2000). Ces facteurs sont connus sous le nom de TAFs et sont fortement associés à TBP chez les mammifères, alors qu'ils semblent moins fortement associés chez la levure (Martinez, 2002). TFIID est donc composé de TBP et des TAFs.

À une résolution de 35 Å, TFIID présente une forme de fer à cheval comprenant trois lobes dominants entourant une cavité centrale. Les lobes sont connectés entre eux par des régions de faible densité (Andel, III et al., 1999). En plus de former un pont entre les activateurs et la machinerie basale, les TAFs de TFIID possèdent certains rôles au niveau de la régulation transcriptionnelle. TAF1 permet de créer une zone d'échafaudage aux autres TAFs et à TBP dans le complexe TFIID; il peut lier certains activateurs pour recruter TFIID au promoteur; il régule la liaison de TBP à l'ADN; il peut aussi lier l'Inr du promoteur et les lysines acétylés des histones; il possède des activités kinase, ubiquitine ligase et acétylase modifiant les histones et les FGTs. Toutes ces activités de TAF1 sont dirigées vers deux buts principaux : aider au positionnement et à la stabilisation de TFIID sur l'ADN et altérer la structure de la chromatine au promoteur pour permettre l'assemblage des FGT (revu par Wassarman et Sauer, 2001).

Chez la levure, les promoteurs ont été groupés en deux classes en se basant sur leur besoin en TAFs. Les promoteurs TAF-dépendent ont besoin des TAFs pour la transcription et les promoteurs TAF-indépendant ne nécessitent pas les TAFs pour leur activité. Des analyses isolant des mutants sensibles à la température pour chacun des 13 TAFs essentiels, suivi d'une étude d'expression sur tout le génome (*genome-wide*), ont permis d'évaluer le pourcentage des gènes dépendant de chacun des TAFs. Chez la levure, environ 84% des gènes nécessitent au moins un TAF et 16% des gènes sont TAFs-indépendants. Ainsi, chacun des TAF aurait des rôles transcriptionnels distincts (Shen et al., 2003).

#### **1.3.3.3- Les remodeleurs de la chromatine**

Les remodeleurs de la chromatine peuvent agir de différentes manières; ils peuvent changer le positionnement du nucléosome sur l'ADN, créer un état remodelé dans lequel l'ADN est accessible et les histones restent liés, permettre la dissociation complète des contacts histones-ADN ou échanger une histone pour une histone "variant" (Mohrmann et al., 2004). Plusieurs complexes protéiques responsables de ce remodelage ont été identifiés. Les complexes les mieux étudiés sont SWI/SNF et RSC chez la levure et NURF, CHRAC et ACF chez la drosophile. Chacun de ces complexes contient une sous-unité possédant une activité ATPase qui est essentielle pour l'activité de remodelage (revu par Lee et Young, 2000).

Le complexe SWI/SNF de la levure est le mieux caractérisé. Ce complexe a été purifié et plusieurs propriétés biochimiques ont été rapportées. Les premiers modèles de remodelage ATP-dépendant de la chromatine pointaient sur les changements dans la composition en histones du nucléosome. En effet, il était suggéré que les enzymes du type SWI/SNF utilisaient l'énergie de l'hydrolyse de l'ATP pour enlever un ou les deux dimères d'histone H2A-H2B. Un peu plus tard, Hayes et Kingston proposaient un modèle alternatif dans lequel les dimères n'étaient pas perdus, mais simplement réarrangés, générant un nucléosome

remodelé (Aoyagi et al., 2002; Peterson, 2002). Le mécanisme précis de ce phénomène n'est pas très bien connu.

Basé sur l'homologie des composantes du complexe SWI-SNF, un autre complexe de remodelage de la chromatine a été isolé chez la levure et nommé le RSC (*remodels the structure of chromatine*) (Cairns et al., 1996). Comme pour SWI/SNF, RSC possède des fonctions ATPase ADN-dépendant, dont l'activité est stimulée par l'ADN libre et nucléosomal. RSC est un complexe de 15 sous-unités, incluant plusieurs peptides reliés aux SWI/SNF. RSC est approximativement 10 fois plus abondant que SWI/SNF et, contrairement au SWI/SNF, certains gènes encodant des sous-unités du RSC ne sont pas essentiels à la viabilité cellulaire (revu par Hampsey, 1998).

D'autres facteurs de remodelage de la chromatine ont été décrits. NURF et CHRAC ont été isolés chez la drosophile. NURF inclut des homologues de SWI2/SNF2. Un homologue de NURF n'a pas été décrit chez la levure, mais des homologues de certaines sous-unités ont été identifiés. CHRAC, pour sa part, inclut certaines sous-unités de NURF et facilite l'assemblage de la chromatine (revu par Hampsey, 1998). Ainsi les eucaryotes contiennent plusieurs complexes de remodelage de la chromatine jouant un rôle de co-activation.

#### **1.3.3.4- Les histones acétyltransférases et les histones déacétylases**

Les histones acétyltransférases (HAT) et les histones déacétylases régulent l'acétylation des histones et des facteurs de transcription. La structure de la chromatine peut être affectée par la modification de la partie N-terminale des queues des histones du nucléosome. L'acétylation des lysines est la modification la mieux caractérisée (revu par Legube et Trouche, 2003). L'acétylation permet d'augmenter l'accessibilité de l'ADN en neutralisant les charges positives des résidus lysines. Cette modification corrèle largement avec l'activation



transcriptionnelle, la réplication et la réparation de l'ADN (revu par Legube et Trouche, 2003).

L'acétylation de la chromatine a été évaluée comme étant un composant clé de la régulation de la transcription et est catalysée par les HATs (Chan et La Thangue, 2001). Les HATs peuvent être regroupés en deux classes en fonction de leur localisation cellulaire. Les HATs de type A sont situées dans le noyau et ont pour fonction d'acétyler les facteurs nucléaires. Pour leur part, les HATs de type B sont localisées dans le cytoplasme et semblent acétyler les histones nouvellement synthétisées pour faciliter leur assemblage (revu par Lee et Young, 2000). Des études biochimiques ont révélé que Gcn5 peut être purifié dans plusieurs complexes capables d'acétyler les histones. Deux complexes de levure et trois complexes humains ont été purifiés et contiennent un HATs semblable à Gcn5. Ceci suggère que la spécificité de substrat peut-être influencée par l'association avec d'autres protéines (revu par Lee et Young, 2000).

La première histone deacétylase, HDAC1, a été purifiée et clonée sur la base de son habileté à lier l'inhibiteur de deacétylase trapoxin. Une mutation de HDAC1 résulte en une hyperacétylation des queues des histones H3 et H4. Huit HDAC humaines ont été identifiées. Chez les mammifères, HDAC1 et HDAC2 existent aussi dans le complexe NuRD. Ce complexe est capable de remodeler les nucléosomes de manière ATP-dépendante (revu par Lee et Young, 2000).

En plus de l'acétylation, plusieurs autres modifications des histones sont impliquées dans le phénomène transcriptionnel. Effectivement, la méthylation, la phosphorylation et l'ubiquitination de résidus spécifiques des histones peuvent aussi affecter la liaison des facteurs de transcription sur l'ADN et changer la structure de la chromatine. Pour une revue récente, voir He et Lehming, 2003.

### 1.3.3.6- Autres co-régulateurs transcriptionnels

Un autre régulateur transcriptionnel purifié par le groupe de Robert Roeder a été nommé USA (*upsteam stimulatory activity*). Le USA implique des activités positives et négatives sur la transcription. Celui-ci interagit avec le PIC pour réprimer la transcription en absence d'activateur et stimuler la transcription en présence d'activateurs. Quelques composantes de USA ont été identifiées incluant PC4. Cette protéine interagit directement avec différents domaines d'activations et le complexe ADN-TBP-TFIIA, médiant l'interaction entre les activateurs et le PIC (revu par Hampsey, 1998).

La protéine chromosomique non-histone HMG2 agit aussi comme un co-activateur en stabilisant une conformation active du complexe TFIID-TFIIA-promoteur (Shykind et al., 1995; Hampsey, 1998).

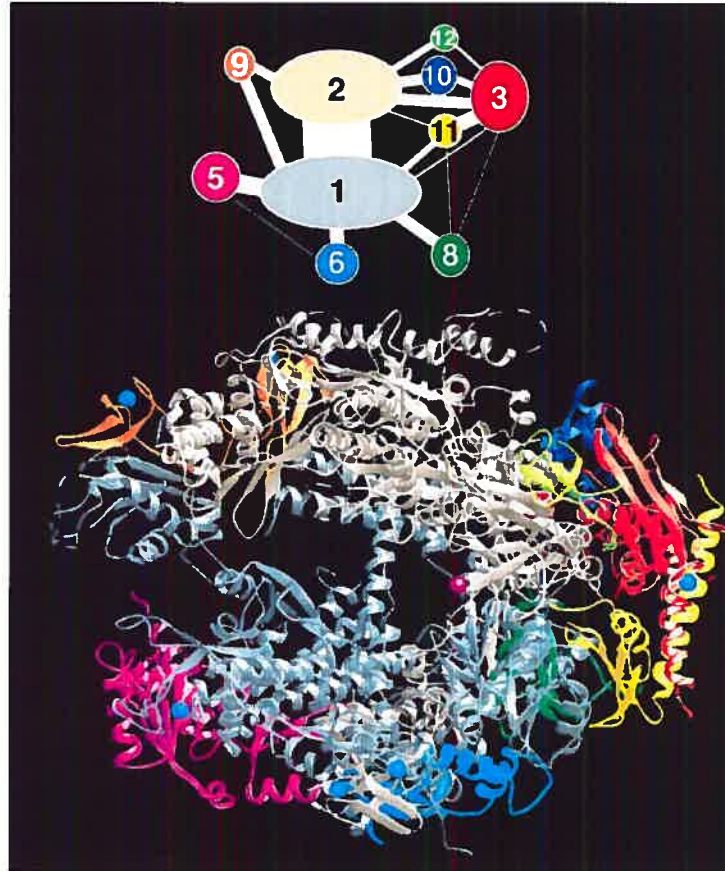
### 1.4- L'ARNP II : structure Vs fonction

La structure grossière de l'ARNP de la bactérie *T. aquaticus* définit une forme de pince de crabe comparable à celle de l'ARNP II. En comparant l'ARNP II et l'ARNP bactérienne, il a été démontré que la plupart des résidus conservés se situent au niveau du centre catalytique. Par contre, il y a très peu de conservation au niveau des résidus de surface. Les similarités dans la structure du centre catalytique suggèrent que le mécanisme d'action est analogue et les différences au niveau des résidus de surface sont cohérentes avec le besoin pour l'enzyme eucaryote de produire plusieurs contacts avec les FGTs et d'autres protéines régulatrices (revu par Woychik et Hampsey, 2002). Des études comparant l'ARNP de *E. coli* avec celle des eucaryotes ont permis de constater des homologies entre les deux enzymes. Effectivement, Rpb1 est structurellement et fonctionnellement homologue à la sous-unité  $\beta'$  de l'ARNP procaryotique et Rpb2 est l'homologue de la sous-unité  $\beta$  bactérienne (revu par Coulombe et Burton, 1999). Rpb1 et Rpb2 sont les deux plus grosses sous-unités de l'ARNP II et, ensemble, elles forment le centre catalytique de l'enzyme. Ces deux sous-

unités contribuent à elles seules pour plus de 60% du volume total de l'ARNP II avec des poids moléculaires respectifs de 220 et 140 kDa (Acker et al., 1997). Pour leur part, Rpb3 et Rpb11 semblent être les homologues distants de la sous-unité  $\alpha$  de *E. coli*. Donc, par analogie à la sous-unité  $\alpha$ , ces deux sous-unités seraient impliquées dans l'assemblage de l'ARNP II.

Les ARNP II de levure et de l'humain affichent un très haut niveau d'identité de séquence, dispersée de manière égale dans toutes les sous-unités. Il a été démontré que la plupart des sous-unités de l'ARNP II humaine sont interchangeable avec leur contrepartie chez la levure (revu par Woychik et Hampsey, 2002). La délétion des sous-unités Rpb4 et Rpb9 chez la levure a démontré que ces dernières ne sont pas essentielles pour la survie de la cellule (Woychik et Young, 1989). Chez les eucaryotes, l'ARNP II est formée de 12 sous-unités nommées Rpb1 à 12. Les sous-unités Rpb6, 8, 10 et 12 sont aussi des sous-unités des ARNPs I et III et celles-ci semblent être des composantes structurales centrales de l'enzyme (revu par Coulombe et Burton, 1999). L'ARNP II purifiée à partir du mutant de délétion de Rpb4 était aussi dépourvue de la sous-unité Rpb7, indiquant que Rpb7 est relié à l'ARNP II via Rpb4. Ce complexe Rpb4/Rpb7 se retrouve en quantité sous-stœchiométrique dans l'ARNP II. Les préparations d'ARNP II purifiées sont constituées d'un mélange protéique non-homogène, causant certains problèmes pour les études de structure cristallographique (Gnatt, 2002). C'est en utilisant un mutant de délétion pour Rpb4 qu'il a été possible de réaliser la première cristallographie de l'enzyme (Edwards et al., 1990).

L'arrivée de la structure cristallographique de l'ARNP II de levure a propulsé l'avancement des recherches dans le domaine de la transcription. La plus forte résolution de l'enzyme obtenue à ce jour est de 2.8 Å avec l'ARNP II de levure à 10 sous-unités (voir la figure 5). Les deux sous-unités manquantes, Rpb4 et Rpb7, sont importantes pour l'initiation de la transcription, mais pas pour l'élongation (revu par Hahn, 2004).

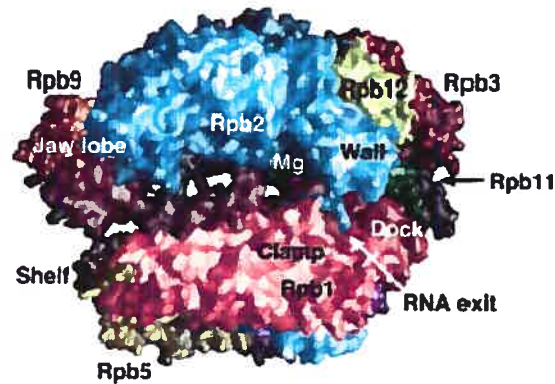


Tiré de Cramer *et al.*, 2000

**Figure 5. Structure cristallographique de l'ARNP II à une résolution de 2.8 Å présentant le positionnement de chacune des sous-unités.** Il est à noter que les sous-unités Rpb4 et Rpb7 sont manquantes sur cette figure.

En 2001, le groupe de Roger Kornberg a décrit la structure cristallographique de l'ARNP II à des résolutions de 3.1 et 2.8 Å (Cramer *et al.*, 2001). Ces structures ont permis de modéliser l'enzyme dans deux états : la forme ouverte et la forme partiellement fermée. En fait, ces deux formes diffèrent principalement par la position de la structure mobile que l'on nomme le *clamp* (revu par Klug, 2001). La structure cristallographique de l'ARNP II à haute résolution a permis d'identifier certaines structures (Cramer *et al.*, 2001). Au centre de l'enzyme, un

sillon est formé par le *core*, les mâchoires (*jaw*), le *lobe* et le *clamp*. Le sillon se termine par le *wall*. Ces structures sont montrées à la figure 6.

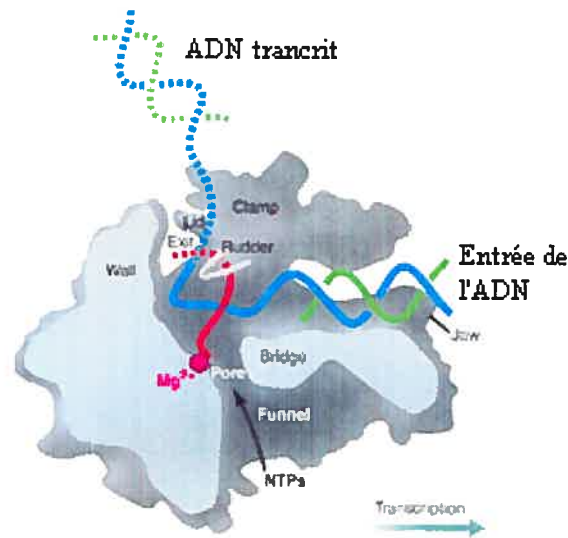


Tiré de Hahn et al, 2004

**Figure 6. Structure de l'ARNP II présentant les différents modules de l'enzyme.**

Le *core* constitue environ la moitié de la masse de l'ARNP II et est formé de Rpb3, 10, 11 12 et de certaines régions de Rpb1 et 2. Les mâchoires sont formées de régions de Rpb1, Rpb5 et Rpb9. Pour sa part, le lobe est formé par Rpb2. Finalement, la plus grande partie du *clamp* dérive de Rpb1, alors que la partie restante est formée par une portion de Rpb2 et de la queue N-terminale de Rpb6 (Woychik et Hampsey, 2002). Le *clamp* est relié au *core* par un ensemble d'éléments flexibles nommés les *switches* et permet au *clamp* de bouger jusqu'à 30 Å pour ouvrir et fermer le sillon (Cramer et al., 2001). La position ouverte permet l'entrée de l'ADN du promoteur dans la structure de l'enzyme. Lorsque Rpb6 est phosphorylé, il peut influencer la position du *clamp* (Kayukawa et al., 1999). Des études récentes ont dévoilé l'architecture de l'enzyme avec les sous-unités Rpb4 et Rpb7 (Armache et al., 2003). La structure de l'enzyme à 12 sous-unités a montré que le sous-complexe Rpb4-Rpb7 est lié à une poche formée de Rpb1, 2, et 6 à la base du *clamp*. Dans cette poche, Rpb7 agit comme un levier bloquant le *clamp* dans sa conformation fermée. Il a été proposé que durant la

formation du complexe ouvert, l'ADN simple brin matrice entre dans le sillon pour aller se positionner au site actif (revu par Hahn, 2004). De plus, un pore dans le complexe protéique peut permettre l'entrée des substrats NTPs pour la polymérisation, ainsi que la sortie des transcrits (Cramer et al., 2000). La réaction catalytique nécessite la présence de deux ions  $Mg^{2+}$  qui se localisent au site actif de l'ARNP II. Celui-ci participe directement dans la formation des liens phosphodiester nécessaires à l'assemblage des nucléotides composant l'ARNm (Coulombe et Burton, 1999). La trajectoire grossière de l'ADN dans la structure de l'ARNP II est décrite à la figure 7.



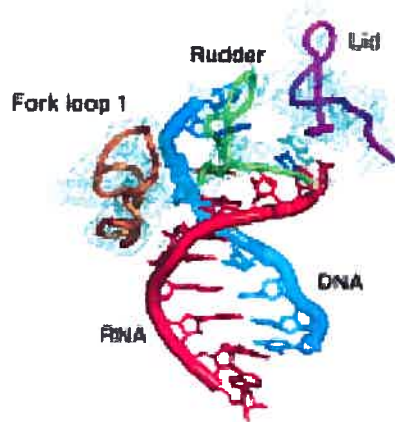
Tiré de Klug, 2001

**Figure 7. Voie empruntée par l'ADN lors de la transcription par l'ARNP II.**

Les surfaces plus rapprochées sont représentées en gris pâle. Le brin codant est en bleu, le brin non-codant en vert et l'ARN est en rouge. Lorsque l'ADN entre dans l'enzyme, il est maintenu par les structures protéiques nommées les *jaw*. Le bout 3' en croissance de l'ARN est localisé près de l'ion  $Mg^{2++}$  (rose) situé au site actif. Un mur protéique (le *wall*) bloque le passage direct de l'ADN à travers l'enzyme, ce qui oblige un changement d'axe de la trajectoire. Les nucléotides peuvent entrer dans une ouverture conique (le *funnel*) sur le côté de l'enzyme et peuvent ainsi avoir accès au centre actif de l'ARNP II en empruntant le pore. Les régions pointillées de l'ADN sont absentes du modèle cristallographique.

D'autres informations importantes ont été obtenues à la suite de la cristallisation de l'ARNP II en élongation à une résolution de 3.3 Å (Gnatt et al., 2001). La carte de densité des électrons permet de voir un hybride ADN-ARN de 9 pb dans la

bulle transcriptionnelle. Les auteurs ont proposé un modèle pour décrire l'étape de translocation. Cette étape suit l'addition du nucléotide à la chaîne d'ARN en élongation et impliquerait l'aide du *bridge helix* (revu par Klug, 2001). Aussi, la structure en élongation a été obtenue spécifiquement au niveau de la séparation de l'ADN et de l'ARN (Westover et al., 2004). La figure 7 montre les interactions entre l'ARNP II et l'hybride ADN-ARN. La structure en boucle de l'ARNP II nommée *lid* semble agir comme un levier pour conduire les brins d'ADN et d'ARN vers le bout en amont de la bulle transcriptionnelle et guider le brin d'ARN vers le sillon de sortie. La structure nommée *rudder* interagit avec l'ADN simple brin après la séparation du brin d'ARN, probablement pour prévenir la réassociation avec l'ARN sortant. Finalement le *fork loop 1* interagit et stabilise l'hybride ADN-ARN. Ces trois structures interagissent ensemble pour former un réseau d'interactions protéine-protéine et protéine-ADN stabilisant le complexe en élongation (voir la figure 8)

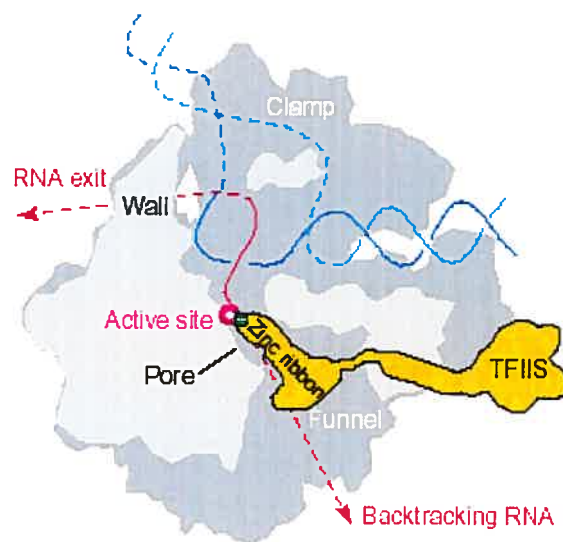


Tiré de Hahn, 2004

**Figure 8. Interactions entre l'ARNP II et l'hybride ADN-ARN dans la structure de l'enzyme en élongation.**



Durant l'élongation de l'ARNm, l'ARNP II peut rencontrer des séquences d'ADN capable de causer un mouvement de recul de l'enzyme. Ce mouvement est nommé *backtracking* et implique l'extrusion du bout 3' de l'ARNm par le pore de l'ARNP II, ce qui peut conduire à un arrêt de la transcription par l'ARNP II. Pour échapper à cet arrêt, le clivage de l'ARNm est requis et est stimulé par le facteur TFIIS. Le facteur d'élongation TFIIS augmente fortement l'activité nucléase de l'ARNP II. TFIIS est constitué du domaine I qui est faiblement conservé, du domaine central II qui est requis pour la liaison à l'ARNP II et du domaine III en C-terminal. Ce dernier est un domaine à ruban de zinc (*zinc ribbon*) essentiel pour le clivage de l'ARNm. Un modèle du squelette du complexe TFIIS-ARNP II a été produit récemment à 3.8 Å de résolution (Kettenberger et al, 2003). Selon ce modèle, le domaine II de TFIIS lie une des mâchoires (*jaw*) de l'ARNP II et son domaine III fait sailli du pore de l'ARNP II menant à son site actif (voir figure 9).



Tiré de Kettenberger et al., 2003

**Figure 9. Schéma des interactions de l'ARNP II avec TFIIS.**

Des études biochimiques et structurales ont permis de proposer un modèle du mécanisme de polymérisation des nucléotides et du clivage des ARNm par l'ARNP II. Ce dernier implique les deux ions métal du centre catalytique de l'enzyme. Dans ce modèle, la polymérisation et le clivage nécessitent le métal A ainsi qu'une implication différentielle du métal B. Pour la réaction de polymérisation, le métal B lie les phosphates du nucléotide (revu par Cramer, 2004). Par contre, pour stimuler le clivage, deux résidus acides essentiels et invariants de la structure en épingle du domaine III de TFIIS complètent le site actif de l'ARNP II et peuvent positionner le métal B et une molécule d'eau. Ce positionnement stimulerait le clivage hydrolytique de l'ARNm. TFIIS produit aussi un changement de conformation dans l'ARNP II pouvant réaligner les acides nucléiques au centre actif (Kettenberger et al., 2003).

#### **1.4.1- Le domaine carboxy-terminal de l'ARNP II (CTD)**

La plus grande sous-unité de l'ARNP II (Rpb1) possède un domaine carboxy-terminal (CTD) consistant en un heptapeptide répété avec la séquence consensus Tyr-Ser-Pro-Thr-Ser-Pro-Ser (Woychik et Hampsey, 2002). Le CTD est répété 28 fois chez la levure, 42 fois chez la drosophile et 52 fois chez l'humain et la souris (Corden, 1990). Même si le CTD n'est pas requis dans les essais d'activité promoteur-dépendent de l'ARNP II, il est essentiel *in vivo*. Effectivement, la délétion complète du CTD chez la drosophile et la levure entraîne la létalité (Nonet et al., 1987). Des études de photopontage réalisées dans le laboratoire de Benoit Coulombe ont permis de localiser le CTD relativement à l'ADN du promoteur. Le CTD a été localisé en aval du site d'initiation, soit aux positions +16 à +26, en juxtaposition avec le tunnel de sortie de l'ARN de l'enzyme ARNP II (Douziech et al., 1999). La phosphorylation du CTD de Rpb1 est importante pour la régulation de l'activité de l'ARNP II durant le cycle transcriptionnel; l'ARNP II hypophosphorylée est préférentiellement recrutée au PIC (Lu et al., 1991) et devient phosphorylée durant la transition de l'initiation à l'élongation (Laybourn et Dahmus, 1990). La phosphorylation se produit initialement sur la

sérine 5 puis sur la sérine 2 du CTD. L'hyperphosphorylation du CTD est aussi liée à d'autres processus essentiels durant la synthèse des ARNm, incluant le recrutement des enzymes de modification des ARNm et des facteurs d'épissage des pré-ARNm (revu par Palancade et Bensaude, 2003). Plusieurs kinases phosphorylent le CTD *in vitro*. Une liste de kinases impliquées dans la phosphorylation du CTD est présentée au tableau I.

**Tableau I. Résumé des principales kinases du CTD et leurs caractéristiques**

Kinases du CTD		Complexe multiprotéique	Positions phosphorylées	Implications transcriptionnelles
Mammi fère	Levure			
CDK1	CDC28	MPF	Ser-2, Ser-5	Répression de la transcription durant la mitose?
CDK7	KIN28	TFIIH	Ser-5 (Ser-2)	Activation de l'entrée en élongation, recrutement des enzymes de formation de la coiffe en 5'
CDK8	SRB10	Médiateur	Ser-2, Ser-5	Répression de l'initiation
CDK9	?	P-TEFb	?	Activation de l'élongation
?	CTK1	Complexe CTK1	?	?
ERK1/2	MPK1		Ser-5	?
DNA-PK		Complexe DNA-PK	Ser-7	?
c-abl			Tyr-1	?

Tiré de (Bensaude et al., 1999; Palancade et Bensaude, 2003)

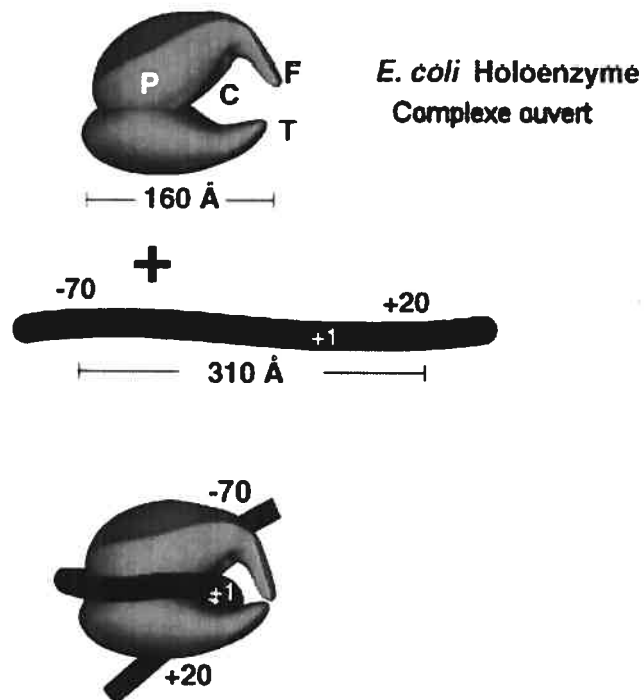
La phosphatase la plus largement connue est nommée Fcp1. Le recyclage de l'ARNP II après la transcription requiert la désphosphorylation du CTD par la phosphatase Fcp1 (Archambault et al., 1997). Fcp1 est conservée parmi les organismes eucaryotes et est essentielle à la viabilité cellulaire (Kamenski et al., 2004). L'activité de cette phosphatase est stimulée par la sous-unité RAP74 du

facteur de transcription TFIIF et cet effet stimulateur peut être bloqué par TFIIB (Chambers et al., 1995). Des études de NMR (*Nuclear magnetic resonance*) ont permis de localiser les domaines de liaison de la phosphatase FCP1 aux facteurs de transcription TFIIF et TFIIB (Nguyen et al., 2003). Récemment, une nouvelle phosphatase du CTD de l'ARNP II a été mise en évidence. Cette dernière, nommée SCP1 (*small CTD phosphatase*), déphosphoryle préférentiellement la sérine 5 (Yeo et al., 2003). De plus, la protéine UBLCP1 (*ubiquitin-like domain containing CTD phosphatase 1*) a été dernièrement identifiée comme étant aussi une phosphatase du CTD de l'ARNP II (Zheng et al., 2005).

#### **1.4.2- Le phénomène d'enroulement de l'ADN du promoteur autour de l'ARNP (*wrapping*)**

Il a été proposé que le phénomène d'enroulement de l'ADN autour de l'appareil transcriptionnel (*wrapping*), stimule la séparation de la double hélice d'ADN, causant une ouverture de l'ADN au site catalytique de l'enzyme avant l'initiation et pendant l'élongation du transcrit (Coulombe et Burton, 1999).

Le phénomène d'enroulement de l'ADN autour de l'ARNP a été mis en évidence chez *E. coli* par des études d'empreinte à la DNase I. Ces expériences, effectuées dans des complexes ouverts et fermés, ont permis d'observer que la liaison de l'ARNP procaryote sur l'ADN s'étendait à 310 Å et plus. Paradoxalement, selon des études cristallographiques, la dimension la plus large de l'ARNP procaryote est d'environ 160 Å (Coulombe et Burton, 1999). L'analogie "l'empreinte est plus grande que le pied" est parfois utilisée pour décrire ce paradoxe (voir la figure 10). Ainsi, il s'avère peu probable que l'ARNP interagisse uniquement par des contacts sur de l'ADN linéaire. Le modèle d'enroulement de l'ADN autour de l'ARNP a alors été proposé dans plusieurs laboratoires (Craig et al., 1998; Polyakov et al., 1995). L'établissement de ce modèle permet donc de résoudre le paradoxe posant une contradiction au niveau de la longueur de l'ADN protégé de la DNase I par l'holoenzyme versus la longueur physique de l'ARNP.



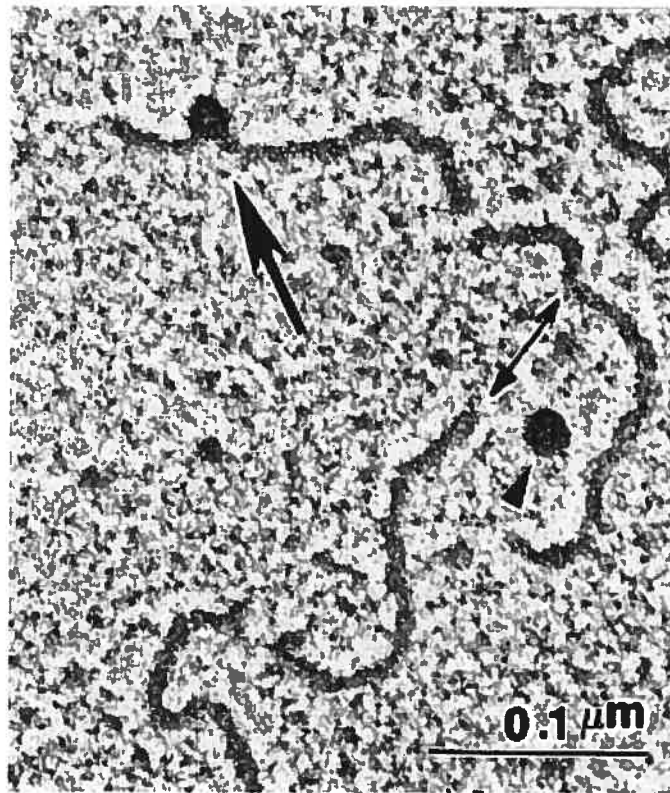
Tiré de Coulombe et Burton, 1999

**Figure 10. Exposition du paradoxe “l’empreinte est plus grande que le pied” et modèle d’enroulement de l’ADN autour de l’ARNP.** La dimension la plus longue de l’ARNP est de 160 Å, alors que les études d’empreinte à la DNase I montrent une protection s’étendant sur 300 Å. Le phénomène d’enroulement de l’ADN autour de l’enzyme permet de résoudre ce paradoxe.

La formation PIC en utilisant l’ARNP de *E. coli* a été étudiée à différentes températures pour permettre la caractérisation des structures intermédiaires. À une faible température (0-15 °C), l’ARNP procaryote a une structure ouverte et une empreinte à la DNase I s’étendant entre les nucléotides -70 ou -50 (dépendamment du promoteur utilisé) et -10. Par contre, à une température plus élevée (30-37 °C), l’ARNP adopte une conformation fermée et une empreinte à la

DNase I allant de -70 ou -50 à +20. Ces observations ont permis d'établir une perspective dynamique du phénomène et de distinguer un enroulement progressif de l'ADN autour de l'ARNP de *E. coli* (revu par Coulombe et Burton, 1999).

Chez les eucaryotes, les premières évidences de phénomène d'enroulement de l'ADN autour de l'ARNP II ont été obtenues par des analyses de photopontage des protéines du PIC sur le promoteur majeur tardif d'adénovirus (AdMLP) (Coulombe et al., 1994; Forget et al., 1997; Robert et al., 1998; Kim et al., 1997). En étudiant le PIC incluant TBP, TFIIB, TFIIF, TFIIE et l'ARNP II, le groupe de Benoit Coulombe a observé que l'ARNP II se localise près du promoteur à différents endroits entre les nucléotides -40 à +26. Cette région correspond à 66 pb, c'est-à-dire environ 225 Å, et est beaucoup plus longue que la plus longue dimension de l'ARNP II, soit 140 Å (Forget et al, 2003). De plus, des images par microscopie électronique de complexes de pré-initiation sur l'ADN du promoteur permettent la visualisation directe de ce phénomène d'enroulement de l'ADN autour de l'enzyme (voir figure 11).



Tiré de Forget et al, 1997

**Figure 11. Microscopie électronique des PIC assemblés sur le promoteur AdMLP en présence de TBP, TFIIB, TFIIF, TFIIE et l'ARNP II. Deux fragments d'ADN sans complexe (flèche à double sens), un complexe libre (tête de flèche) et un complexe lié à l'ADN (flèche à un sens) sont montrés.**

Comme il a été mentionné précédemment, le facteur de transcription TFIIF est impliqué dans ce phénomène d'enroulement. Des expériences de photopontage ont permis de localiser RAP74, la grande sous-unité de TFIIF, à proximité de l'ADN du promoteur à des distances similaires en amont et en aval de la courbure de l'ADN au niveau de la boîte TATA. La liaison de RAP74 induit un changement de conformation affectant la position de l'ARNP II relativement à l'ADN. La région N-terminale de RAP74 (1-205), est nécessaire et suffisante

pour la réorganisation du PIC et pour permettre des contacts de l'ARNP II en amont et en aval de la boîte TATA (Forget et al., 1997). Cette région contient le domaine de liaison à RAP30 et une autre région située entre les acides aminés 172 et 205 qui est aussi nécessaire pour permettre la transcription *in vitro*. Cette région, appelée HIR1, est responsable de l'homomérisation de RAP74. Celle-ci permet les interactions RAP74-RAP74 facilitant un enroulement étendu de l'ADN autour de l'ARNP II. Effectivement, plusieurs contacts du promoteur avec l'ARNP II requièrent la présence de HIR1 et du domaine de liaison de RAP30 (Robert et al., 1998). Les études de photopontage ont permis de montrer que l'utilisation d'un mutant RAP74 (1-172) provoque des contacts de l'ARNP II avec l'ADN entre les positions -19 et -5. Par contre, le mutant RAP74 (1-205) ou d'autres plus longs, étendent les contacts de l'ADN avec l'ARNP II aux positions -39/-40 à +13. (Forget et al., 1997; Robert et al., 1998). L'enroulement serré de l'ADN autour de la machinerie transcriptionnelle est représenté à la figure 12.



Tiré de Robert *et al*, 1998

**Figure 12. Modèle d'enroulement de l'ADN autour de la machinerie transcriptionnelle induit par RAP74.** TBP est en jaune, TFIIB en rouge, TFIIF en vert, TFIIE34 en bleu clair, l'ARNP II en bleu foncé, la région HIR1 est indiquée et représente le domaine de dimérisation de RAP74.



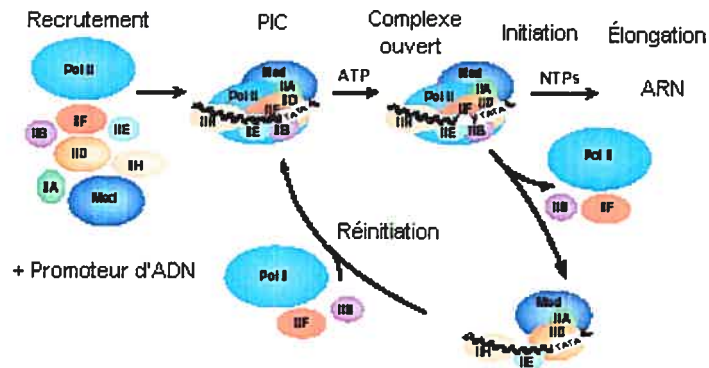
Des études subséquentes utilisant des mutants de délétion suggèrent que la région de RAP74 entre les acides aminés L155 et M177 stimule l'ouverture de l'hélice d'ADN pour l'initiation (Ren et al., 1999).

Les facteurs de transcription TBP et TFIIB facilitent le phénomène d'enroulement du promoteur d'ADN en induisant une courbure d'environ 80° dans l'ADN au niveau de la boîte TATA (Werner et Burley, 1997). Des études de photopontage ont révélé que la courbure au niveau de la boîte TATA induite par TBP et TFIIB permet le contact de l'ARNP II avec l'ADN du promoteur dans les régions -39/-40 (Forget et al., 1997; Robert et al., 1998; revu par Coulombe et Burton, 1999). Aussi, la présence de TFIIE favorise l'enroulement de l'ADN (Robert et al., 1998).

Cet enroulement pourrait faciliter l'initiation de la transcription en ouvrant partiellement le promoteur au niveau du site +1. L'ouverture partielle libérerait une matrice d'ADN simple brin, facilitant ainsi l'accès aux hélicases de TFIIF ATP dépendantes.

### ***1.5- Le cycle de transcription par l'ARNP II***

La transcription est un processus moléculaire qui se divise en étapes : la pré-initiation, l'initiation, le dégagement du promoteur, l'élongation et la terminaison. Les étapes sont présentées schématiquement à la figure 13.



Tiré de Hahn, 2004

Figure 13. Résumé du cycle transcriptionnel de l'ARNP II

### 1.5.1- Pré-initiation de la transcription : l'assemblage séquentiel Vs

#### l'holoenzyme

Le processus transcriptionnel débute généralement par la liaison d'un facteur de régulation spécifique au gène transcrit sur l'ADN. Ce facteur pourra agir indirectement sur la machinerie transcriptionnelle en recrutant les facteurs responsables de la modification de la chromatine ou directement en interagissant avec les composants de la machinerie basale. Le remodelage de la chromatine facilite l'accès de la machinerie transcriptionnelle à la région promoteur du gène (revu par Hahn, 2004). L'ARNP II pourra ensuite rejoindre le promoteur génique en compagnie des facteurs généraux de transcription TFIIA, TFIIB, TFIID, TFIIIE, TFIIF et TFIIH, afin de former le PIC. La théorie initialement admise sur la formation du PIC implique l'assemblage séquentiel de chacun des facteurs au niveau de l'ADN du promoteur. Dans ce modèle, TBP liant la boîte TATA, sert de plateforme pour le recrutement subséquent du facteur TFIIB. Par la suite, TFIIF escorte l'ARNP II pour se joindre le complexe ADN-TBP-TFIIB. Finalement, TFIIIE et TFIIH entrent dans le complexe et finalisent la formation du PIC. Le rôle de TFIIA dépendra de la composition du système. TFIIA stimule la transcription dans des systèmes utilisant des extraits cellulaires et dans des

systèmes reconstitués lorsque TFIID est employé (TBP avec les TAFs). Le facteur TFIIA peut s'associer au complexe à n'importe quelle étape après l'association de TBP à l'ADN (revu par Ranish et al., 1999).

D'autre part, ce modèle séquentiel a été remis en question par la découverte de complexes pré-assemblés en solution comprenant l'ARNP II, quelques facteurs généraux de transcription et d'autres protéines (revu par Myer et Young, 1998). Ce complexe comprenant l'ARNP II est communément appelé "holoenzyme". Une des formes de l'holoenzyme chez la levure comprend l'ARNP II, TFIIF et le complexe Srb/médiateur, comprenant Srb2, Srb4, Srb6, Srb7, Gal11, Rgr1, Sin4, Rox3 et 6 protéines Med (Myers et al., 1998). Une autre forme de l'holoenzyme de levure contient les facteurs cités ci-haut avec les facteurs TFIIB, TFIIH, des protéines Srb additionnelles et les protéines SWI/SNF (Wilson et al., 1996). Bref, l'holoenzyme a été identifiée chez la levure sous différentes formes. Récemment, la purification de l'ARNP II dans des cellules de mammifère par la méthode TAP (*tandem affinity peptide*) a permis l'identification d'un nouveau complexe soluble. Ce complexe est composé de l'ARNP II, TFIIB, TFIIF, Fcp1 et RPAP1 (Jeronimo et al., 2004). Chez la levure comme chez les mammifères, les facteurs TFIIA et TFIID ne semblent pas se retrouver dans les formes caractérisées de l'holoenzyme. Un recrutement indépendant de ces deux facteurs au PIC a été proposé (Ranish et al., 1999). À la suite de la formation du PIC (formé selon le modèle séquentiel ou de l'holoenzyme), TFIIH permettra l'ouverture de la bulle transcriptionnelle entre les positions -9 à +2, soit tout près du site d'initiation +1, pour libérer une matrice d'ADN simple brin. Il est à noter que l'activité hélicase du facteur TFIIH nécessite de l'énergie provenant de l'hydrolyse de l'ATP et est étroitement régulée par le facteur TFIIE. La matrice d'ADN libérée sera positionnée au niveau du sillon majeur de l'ARNP II.

### **1.5.2- Initiation**

La transcription peut ensuite être initiée par la formation du premier lien phosphodiester de l'ARN naissant. Le processus d'appariement des NTPs

nécessite deux ions  $Mg^{2+}$  situés au site actif de l'ARNP II. Après l'ajout de quatre NTPs, le PIC produit un changement de conformation menant à la stabilisation du complexe ouvert. Un cycle d'initiation avorté conduira à la formation de plusieurs courts transcrits d'ARN avant de procéder à l'élongation productive (Holstege et al., 1997).

### **1.5.3- Dégagement du promoteur**

Lorsque le transcrit atteint environ trente bases, l'ARNP II n'est plus en contact avec le promoteur. Avant de commencer l'étape d'élongation productive, l'ARNP II doit passer par l'étape de dégagement du promoteur. Durant cette étape, le PIC est partiellement désassemblé et certains facteurs restent au niveau du promoteur. Ceux-ci serviront d'échafaudage à un nouveau cycle transcriptionnel. La phosphorylation de la sérine 5 du CTD de l'ARNP II semble corrélérer avec l'étape de dégagement du promoteur (revu par Sims, III et al., 2004). De plus, le facteur TFIIF permet aussi de faciliter cette étape, en partie en prévenant l'arrêt prématuré de l'ARNP II. De tous les facteurs généraux de transcription, seul TFIIF se retrouve dans le complexe d'élongation. Une fois le promoteur dégagé, un nouveau cycle transcriptionnel peut avoir lieu (revu par Sims, III et al., 2004).

### **1.5.4- Élongation et terminaison**

L'ARNP II en élongation est en mesure de lier les facteurs d'élongation, les facteurs de modification des ARN et les facteurs de modification de la chromatine, formant un système intégré. Certains facteurs d'élongation ainsi que leurs fonctions sont décrits dans le tableau II (pour une liste plus complète et plus détaillée, consulter la revue de Sims et al, 2004).

**Tableau II : Les facteurs d'élongation de l'ARNP II**

<b>Facteurs d'élongation</b>	<b>Propriétés</b>
<b>TFIIF</b>	Essentiel pour la formation du PIC, stimule l'efficacité de l'ARNP II, module TFIIS, diminue les pauses
<b>Elongine</b>	Diminue les pauses, stimule l'efficacité de l'ARNP II
<b>ELL</b>	Diminue les pauses, stimule l'efficacité de l'ARNP II
<b>DSIF</b>	Stimule l'élongation, supprime la terminaison précoce, stimule la formation de la coiffe en 5'
<b>NELF</b>	Stimule une pause de l'ARNP II pour permettre le recrutement des facteurs de formation de la coiffe en 5'.
<b>FCP1</b>	Stimule l'élongation, recycle l'ARNP II par la déphosphorylation du CTD, implication dans la formation de la coiffe en 5'
<b>TFIIS</b>	Réactive l'ARNP II arrêtée en interagissant directement avec l'enzyme, stimule l'ARNP II à cliver le transcrit naissant en 3' (permet le réalignement de l'ARNP II au site actif et le 3'-OH pour passer outre le site d'arrêt)
<b>P-TEFb</b>	Phosphorylation de la Serine 2 sur le CTD, qui est critique pour la transition initiation à élongation
<b>FACT</b>	Module la structure de la chromatine, facilite l'élongation à travers la chromatine.

Tiré de Sims et al, 2004

L'étape finale du cycle transcriptionnel est la terminaison. À cette étape, l'ARNm est clivé, polyadénylé et transporté au cytoplasme où il sera traduit en protéine. Il est intéressant de noter que le facteur spécifique de clivage et polyadénylation (CPSF) peut être recruté par TFIID durant la formation du PIC (Dantonel et al., 1997). Ceci démontre que les processus d'initiation et de terminaison sont interconnectés et peuvent s'influencer l'un l'autre.

### ***1.6- Études des interactions protéine-ADN***

Le recrutement des machineries moléculaires à une localisation génique spécifique requiert des protéines régulatrices qui reconnaîtront de courts motifs

d'ADN. Les interactions entre les protéines régulatrices et la machinerie transcriptionnelle avec l'ADN génique sont essentielles au fonctionnement des processus fondamentaux.

### **1.6.1- Les méthodes utilisées pour étudier les interactions protéine-ADN**

Plusieurs méthodes ont été développées pour étudier les interactions protéine-ADN *in vitro* et *in vivo*. Parmi les méthodes les plus utilisées *in vitro*, on retrouve le gel de retardement de bandes, les analyses d'empreinte à la DNase (celle-ci peut parfois être utilisée *in vivo*) et le photopontage. *In vivo*, l'immunoprécipitation de la chromatine (CHIP) est fortement utilisée.

#### **1.6.1.1- Le gel de retardement de bandes (EMSA)**

La méthode de gel de retardement de bandes ou EMSA se base sur le principe que les protéines ont un poids beaucoup plus élevé que l'ADN et par conséquent, se déplacent beaucoup plus lentement dans des gels d'acrylamide natifs. Dans certaines conditions expérimentales favorables, l'ADN marqué radioactivement sera lié par une protéine et se distinguera de l'ADN libre (sonde) par une migration plus lente dans le gel (Garner et Revzin, 1981). Cette méthode permet de déterminer la présence d'une liaison d'une protéine ou d'un complexe protéique à un fragment d'ADN.

#### **1.6.1.2- Les analyses d'empreinte à la DNase**

Dans cette méthode, un segment d'ADN cible marqué à son extrémité 5' est partiellement dégradé par la DNase en présence et en absence de la protéine étudiée. Le patron de dégradation partiel est ensuite observé sur un gel dénaturant. Si la protéine lie l'ADN cible, le site de liaison de la protéine sur l'ADN sera protégé de la dégradation par la DNase et le patron de digestion sera modifié (Galas et Schmitz, 1978). Comparativement à la méthode de gel de retardement de bande, les analyses d'empreinte à la DNase permettent de localiser le site de liaison précis de la protéine.

### 1.6.1.3- Le photopontage protéine-ADN site spécifique

La méthode de photopontage permet de ponter les protéines à des sites spécifiques sur l'ADN par l'entremise de sondes radioactives contenant des nucléotides photoréactifs. Ainsi, des sondes d'ADN plaçant un photonucléotide à des positions déterminées sont initialement synthétisées. Par la suite un complexe est assemblé sur la sonde et le photonucléotide est activé en présence de lumière UV. Les protéines avoisinant le photonucléotide seront liées de manière covalentes par ce dernier et pourront être identifiées par migration sur gel SDS-PAGE (Forget et Coulombe, 2003). Cette méthode permet de localiser de manière précise la protéine sur l'ADN. De plus cette méthode peut être appliquée à des complexes multi-protéiques. Contrairement aux analyses d'empreinte à la DNase qui ne délimitent que l'endroit où débute et se termine la liaison sur l'ADN, le photopontage en gel permet de déterminer les interactions protéiques au long du fragment d'ADN.

### 1.6.1.4- L'immunoprécipitation de la chromatine (CHIP)

Contrairement aux trois méthodes présentées précédemment, le CHIP offre l'avantage de déterminer les interactions des protéines avec l'ADN *in vivo*. Dans cette méthode, la protéine est liée de manière covalente à l'ADN par un agent de pontage comme le formaldéhyde. Ensuite, la chromatine est cisailée et la protéine cible est précipitée en utilisant un anticorps spécifique, co-précipitant ainsi des fragments d'ADN. Les pontages sont ensuite renversés et l'ADN est relâché. Les fragments d'ADN peuvent être identifiés par PCR (Kuo et Allis, 1999). La méthode ChIP-chip combine les techniques d'immunoprécipitation de la chromatine avec la technologie des *microarray*. Cette méthode permettra des analyses à l'échelle du génome. Utilisant la méthode ChIP-chip, aussi appelée *location analysis*, Lee *et al.*, ont été défini un réseau régulateur chez la levure consistant en 106 facteurs de transcription et 2363 gènes cibles (Lee et al., 2002).

### **1.6.2-Objectif de la première partie**

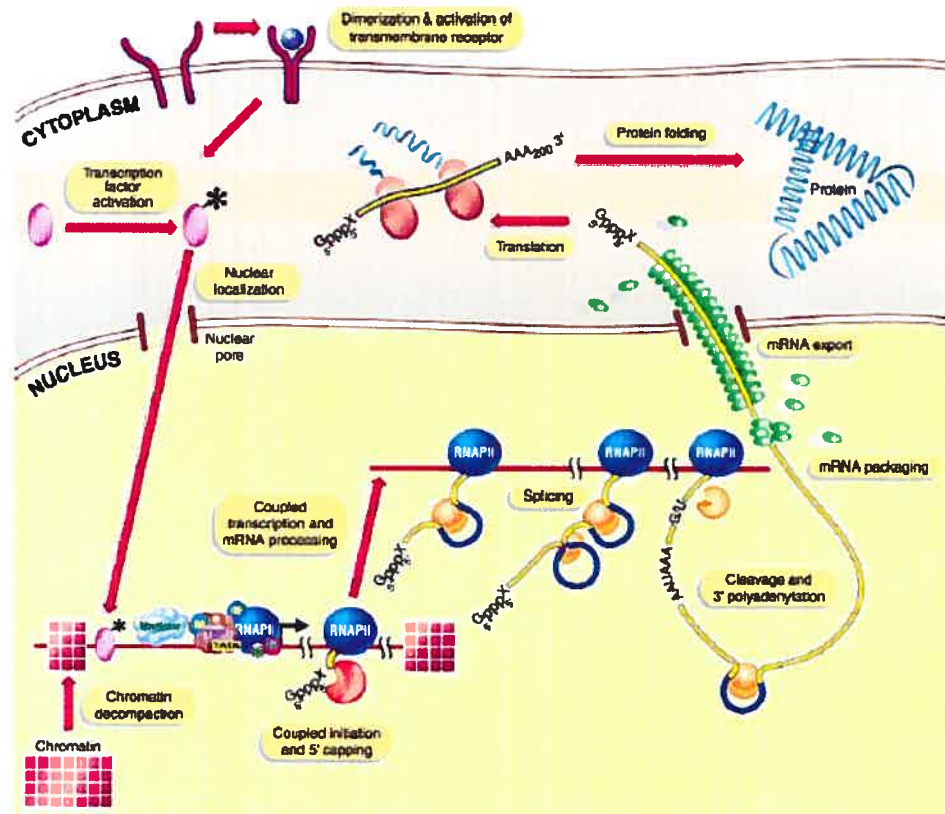
Le premier objectif de ce travail était de déterminer les interactions protéine-ADN du PIC comprenant les facteurs TBP, TFIIB, TFIIF (RAP30 et RAP74), TFIIE (E34 et E56) et l'ARNP II. Pour ce faire, une méthode de photopontage "en gel" fût utilisée. Les techniques servant à la visualisation directe de l'enroulement de l'ADN comme la microscopie électronique ou à force atomique, ne possèdent pas la résolution requise pour cartographier la trajectoire de l'ADN autour de l'enzyme. De plus, la cristallographie ne permet pas actuellement de former des cristaux de PIC entiers pour identifier les interactions protéine-ADN. Aussi, les méthodes d'EMSA, d'empreinte à la DNase et d'immunoprécipitation décrites précédemment ne permettent pas d'étudier la trajectoire précise de l'ADN à travers les nombreux polypeptides formant le PIC. Notre hypothèse est que la méthode de photopontage en gel permettra de déterminer la trajectoire de l'ADN dans la structure cristallographique de l'ARNP II établie par Cramer *et al* en 2000, ainsi que la position relative des FGTs. L'identification de la trajectoire de l'ADN dans le PIC permettra de mieux comprendre le mécanisme transcriptionnel par l'ARNP II.

### ***1.7- Études des réseaux d'interactions protéine-protéine***

La majorité des protéines ne fonctionnent pas seules et par elles-mêmes. Généralement, elles interagissent avec d'autres biomolécules pour réaliser leurs fonctions. Le réseau d'interaction des biomolécules constitue la base de la vie et les interactions entre les protéines, les effecteurs fonctionnels des processus cellulaires, détiennent un rôle extrêmement important. L'*interactome* des protéines constitue un réseau fortement complexe régissant de manière intégrée plusieurs processus cellulaires. Les mécanismes d'expression génique menant à la formation des protéines impliquent plusieurs étapes et de nombreux effecteurs. Par le passé, chacun des processus amenant à la génération des protéines avait été étudiée de manière distincte. Aujourd'hui, plusieurs évidences expérimentales ont mené à une vision beaucoup plus globale et intégrée du phénomène. La figure 14



illustre bien la vision actualisée de l'expression des gènes impliquant le couplage des processus. Selon cette vision, chaque étape de l'expression génique est une subdivision d'un processus continu (Orphanides et Reinberg, 2002).



Tiré de Orphanides et Reinberg, 2002

**Figure 14. Vision actuelle de l'expression génique impliquant le couplage des processus.**

En plus de faciliter la compréhension du phénomène de couplage des processus cellulaires, l'identification des partenaires d'une protéine inconnue peut aider à inférer sa fonction. Par contre, malgré le couplage des processus, les protéines interagissant ensemble sont souvent impliquées dans la même voie biochimique.

### **1.7.1- Les principales méthodes utilisées pour étudier les interactions protéine-protéine à large échelle chez les eucaryotes**

L'intérêt de plus en plus grandissant de définir l'*interactome* protéique a mené à la création de plusieurs banques de données informatiques emmagasinant des résultats d'interactions connues. Ces banques ont été conçues initialement à partir de données des études à petite échelle (Bader et al., 2001; Salwinski et al., 2004). Au niveau expérimental, les premières études d'interaction protéique à large échelle ont été publiées en 2000 (Uetz et al., 2000) et 2001 (Ito et al., 2001). Ces deux études utilisaient la technologie du double hybride. Deux années plus tard, les premiers complexes purifiés à large échelle ont été caractérisés (Ho et al., 2002; Gavin et al., 2002). Ces complexes ont été purifiés par la méthode TAP. Ces deux méthodes sont les plus utilisées actuellement pour déterminer le réseau d'interaction protéique à grande échelle dans les cellules eucaryotes.

#### **1.7.1.1- La méthode du double-hybride**

Cette méthode est largement utilisée pour déterminer les interactions physiques entre deux protéines. Le système double-hybride a été développé par Stanley Field (Fields et Song, 1989) et utilisait les domaines modulaires d'activation et de liaison de GAL4. Pour ce faire, la première protéine d'intérêt (l'appât) est fusionnée avec un domaine de liaison à l'ADN et la deuxième (la proie) est fusionnée avec le domaine d'activation du même facteur de transcription. Ces deux protéines hybrides sont introduites dans la levure et la transcription du gène rapporteur régulée par le facteur de transcription utilisé indique que les protéines interagissent ensemble. Ainsi, le domaine d'activation et le domaine de liaison à l'ADN du facteur de transcription utilisé sont reconstitués par la formation d'un pont créé par la liaison des protéines de fusion. Le système double-hybride a toutefois ses inconvénients. Premièrement, cette méthode ne peut pas (en théorie) détecter des interactions entre plus de deux protéines (Ito et al., 2001). Aussi, le système de deux hybrides ne convient pas à la détection des interactions impliquant des protéines membranaires (Ito et al., 2001) et ne convient pas très

bien aux protéines impliquées dans la transcription. Finalement, cette méthode peut générer beaucoup de faux positifs et négatifs (Coulombe et al., 2004)

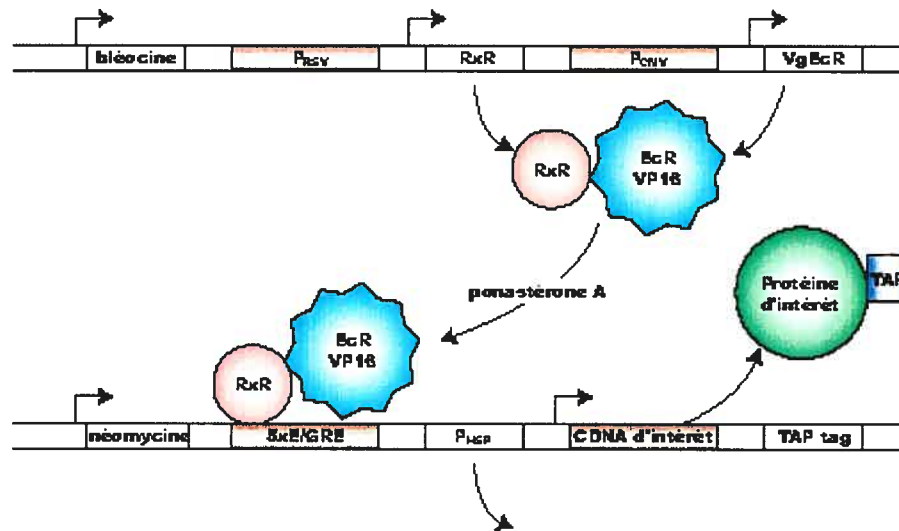
#### **1.7.1.2- La purification d'affinité utilisant une double étiquette TAP (*Tandem Affinity Peptide*)**

Dans cette méthode, une protéine de fusion (l'appât) portant une double étiquette est exprimée dans les cellules et utilisée pour purifier ses partenaires d'interaction. Il est à noter que la purification de complexes par la méthode TAP implique l'identification de partenaires liant directement ou indirectement l'appât. La purification du complexe en conditions natives utilise deux colonnes d'affinité. La purification TAP possède plusieurs avantages pour les études protéomiques à grande échelle. Un avantage important réside dans l'utilisation de conditions natives permettant de maintenir les interactions avec ses protéines associées et la purification de complexes entiers tels qu'ils existent *in vivo*. Comparativement à la méthode du double hybride, la purification TAP a pour avantage d'identifier les partenaires d'interaction directs ou indirects dans une seule expérience. De surcroît, la purification TAP peut fournir des indications sur la stoechiométrie des protéines des complexes et permet les analyses biochimiques du complexe identifié. La purification TAP permet aussi de purifier des protéines mutantes. Les mutants purifiés peuvent alors être caractérisés dans divers essais biochimiques. Certains inconvénients à cette méthode doivent aussi être mentionnés. Tout d'abord, il est possible que l'étiquette ajoutée à la protéine ne soit pas suffisamment exposée pour permettre la liaison de la protéine aux billes d'affinité. De plus, cette même étiquette peut interférer avec la fonction de la protéine ou l'assemblage d'un complexe. Finalement, la protéase TEV utilisée au cours de la purification (voir section 2.2.4.2) peut cliver dans certains cas rarissimes la protéine d'intérêt ou un des partenaires protéiques qui lui sont associés.

La purification TAP chez les mammifères, mise en place dans notre laboratoire, permet l'expression de la protéine d'intérêt à un niveau endogène en employant le

système d'expression-Ecdysone illustré à la figure 15. Un système inductible permet d'avoir un contrôle négatif servant à éliminer les faux-positifs, comme la présence de protéines contaminantes liant de manière non-spécifique les billes ou détenant des sites de liaison pour la calmoduline. De plus, l'introduction du plasmide d'expression dans la cellule et l'expression de la protéine exogène peuvent altérer la physiologie de la cellule. Conséquemment, la cellule peut répondre à des stimuli de manière différente de la lignée mère non transfectée. Un contrôle négatif utilisant des cellules non-induites permet de contourner ce biais de la méthode.

*Système d'expression-Ecdysone :*



**Figure 15. Représentation schématique du système d'expression inductible à l'ecdysone chez les mammifères.** Le plasmide du haut est présent dans les cellules ECR-293 et exprime constitutivement le récepteur de l'acide rétinoïque (RXR), le récepteur ecdysone (VgEcR) qui contient le domaine de transactivation VP16 et le gène de résistance à la bléocine. Les récepteurs RXR et VgEcR hétérodimérisent et lient l'élément de réponse spécifique à l'ecdysone (E/GRE) en présence de ponastérone A (No et al., 1996; Vickers et Sharrocks, 2002). La boîte P du récepteur VgEcR a été modifiée pour reconnaître un demi-site de l'élément de réponse aux glucocorticoïdes (GRE). Le plasmide du bas est formé du gène encodant la protéine d'intérêt s'exprimant à partir d'un promoteur contenant l'élément de réponse (E/GRE). Il contient aussi un gène de résistance à la néomycine.

## **1.7.2- XAB1 et DDX5, deux nouveaux partenaires d'interaction de l'ARNP II et de TFIIS**

### **1.7.2.1- Utilisation de la méthode TAP pour la purification de la machinerie transcriptionnelle basale**

La méthode TAP a été utilisée dans le laboratoire de Benoit Coulombe afin de purifier l'ARNP II et tous les FGTs humains dans des conditions natives. Comparés aux FGTs purifiés selon des méthodes chromatographiques classiques, les "holo-FGTs" comportent des polypeptides additionnels dont certains pourraient représenter des sous-unités régulatrices. En 2004, ces études ont permis de décrire la purification et la caractérisation d'un complexe comprenant toutes les sous-unités de l'ARNP II, les facteurs de transcription TFIIB et TFIIF (RAP30 et RAP74), la phosphatase Fcp1 et une protéine à fonction inconnue RPAP1 (Jeronimo et al, 2004). Plus récemment, un polypeptide additionnel, XAB1 a été identifié dans ce complexe en utilisant les appâts Rpb2, Rpb11 et RPAP1 (tableau III).

Les rôles décrits de XAB1 au niveau de la régulation transcriptionnelle (voir section 1.7.2.2) et ses interactions avec Rpb2, Rpb11 et RPAP1 nous ont poussé à l'utiliser comme appât pour identifier ses partenaires d'interactions. Cette expérience permettra de confirmer l'interaction de XAB1 avec l'ARNP II et RPAP1 et d'identifier de nouveaux partenaires d'interaction.

Les partenaires d'interaction de TFIIS, ont aussi été identifiés par l'équipe de Benoit Coulombe en utilisant la méthode TAP et comprend plusieurs polypeptides dont certains impliqués dans la machinerie d'épissage comme DDX5 (Jeronimo et al, en préparation). La purification par la méthode TAP a permis d'identifier DDX5 lorsque TFIIS était utilisé comme appât, indiquant que DDX5 interagit avec TFIIS. Des rôles de DDX5 en transcription ont aussi été décrits (voir section 1.7.2.3), ce qui nous a incité à l'utiliser comme appât pour identifier ses partenaires d'interactions.

**Tableau III. Protéines purifiées par la méthode TAP et indentifiées par spectrométrie de masse MS/MS en utilisant les appâts Rpb2, Rpb11 et RPAP1.**

Rpb2	Rpb11	RPAP1
Rpb1	Rpb1	
<b>Rpb2</b>	Rpb2	Rpb2
Rpb3	Rpb3	Rpb3
Rpb4	Rpb4	
Rpb5	Rpb5	
Rpb6	Rpb6	
Rpb7	Rpb7	
Rpb8	Rpb8	
Rpb9	Rpb9	
Rpb10	Rpb10	
Rpb11	<b>Rpb11</b>	
Rpb12	Rpb12	
<b>XAB1</b>	<b>XAB1</b>	<b>XAB1</b>
TFIIB	TFIIB	
RAP74	RAP74	
RAP30	RAP30	
Fcp1	Fcp1	
RPAP1	RPAP1	<b>RPAP1</b>

Aujourd'hui, les résultats obtenus par la méthode TAP, nous permettent d'établir un large réseau régulateur génique de la machinerie transcriptionnelle basale comprenant l'ARNP II, les FGTs, des polypeptides du médiateur, des facteurs impliqués dans l'épissage ainsi que des protéines de fonctions inconnus. (Jeronimo et al, en préparation).

#### **1.7.2.2- XAB1 (*XPA binding protein 1*) ou MBDin**

La liaison de XAB1 avec XPA a été identifiée par des analyses de doubles hybrides sur une librairie d'ADNc provenant de cellules HeLa. XAB1 est une GTPase cytoplasmique impliquée dans la localisation nucléaire de XPA et des analyses de sa séquence en acides aminés indiquent la présence d'un NES (*nuclear export signal*) potentiel. XAB1 se localise majoritairement dans le

cytoplasme, mais il s'accumule dans le noyau s'il y a perturbation du NES ou avec un traitement à la leptomycin, un inhibiteur du transport cytoplasmique médié par le NES (Lembo et al., 2003).

Une étude identifiant les partenaires d'interactions de MBD2 (*Methyl-CpG binding protein 2*) par la méthode du double hybride avec une librairie d'ADNc provenant de cellules HeLa a permis l'identification de MBDin (XAB1). MBD2 est un composant d'un large complexe de protéines, le MeCP1, qui réprime la transcription des gènes fortement méthylés. MeCP1 inclut HDAC1, HDAC2 et RbAp46/48. De plus, MBD2 peut s'associer à différentes protéines. La région C-terminale de MBD2 est nécessaire pour l'interaction avec MBDin (XAB1) et des analyses fonctionnelles démontrent que MBDin est capable d'inhiber les fonctions de MBD2. Selon cette étude, MBDin, en liant MBD2, réactive les promoteurs méthylés sur de l'ADN épisomal (Lembo et al., 2003).

#### 1.7.2.3- DDX5 : *DEAD (Asp-Glu-Ala-Asp) Box polypeptide 5* ou p68

Les protéines à boîte DEAD sont des hélicases caractérisées par le motif conservé Asp-Glu-Ala-Asp. DDX5 n'a pas une localisation cellulaire unique, celle-ci variant d'exclusivement nucléaire à prédominance cytoplasmique dans trois différentes lignées cellulaires étudiées (Rossow et Janknecht, 2003). La protéine DDX5 a différents rôles au niveau de l'épissage des ARNm. Entre autre, la délétion de DDX5 bloque l'assemblage du splicéosome à l'étape du pré-splicéosome (Liu, 2002). De plus, DDX5 est impliquée dans la régulation de l'épissage alternatif du pré-ARNm c-*H-ras* (Guil et al., 2003)

La microscopie confocale et l'immunocoloration de cellules *Myc-tagged* DDX5 et *HA-tagged* p300 montre une co-localisation de DDX5 et p300. Les résultats obtenus au cours de cette étude suggèrent que DDX5 peut co-activer la transcription (Rossow et Janknecht, 2003). Des analyses de *pull down*, utilisant la protéine de fusion GST-DDX5 et des extraits nucléaires dérivés de cellules 293T



ont permis d'identifier l'interaction de DDX5 avec l'ARNP II. Rossow et Janknecht suggèrent l'existence d'un complexe multiprotéique incluant DDX5, CBP/p300 et l'ARNP II promouvant conjointement l'expression génique (Rossow et Janknecht, 2003). DDX5 (p68) et p72 sont deux hélicases de la famille DEAD box qui interagissent ensemble et existent sous la forme d'un dimère (Ogilvie et al., 2003). Récemment, des études utilisant un gène rapporteur couplé à différents promoteurs ont démontré que p68 et p72 peuvent agir en tant que répresseurs transcriptionnels et que cette répression serait spécifique au promoteur utilisé (Wilson et al., 2004). Dans cette même étude, plusieurs résultats portent à croire que p68, p72 et HDAC1 existent dans le même complexe, suggérant certains mécanismes pour expliquer la répression transcriptionnelle (Wilson et al., 2004).

### **1.7.3- Objectifs de la deuxième partie**

La deuxième partie de mon projet consistait à utiliser la méthode TAP avec XAB1 et DDX5. L'identification de leurs partenaires d'interaction permettra notamment d'élucider certains rôles de ces protéines et possiblement d'identifier de nouveaux partenaires.

## 2- MATÉRIELS ET MÉTHODES

### ***2.1- Le photopontage en gel***

#### **2.1.1- But de la méthode**

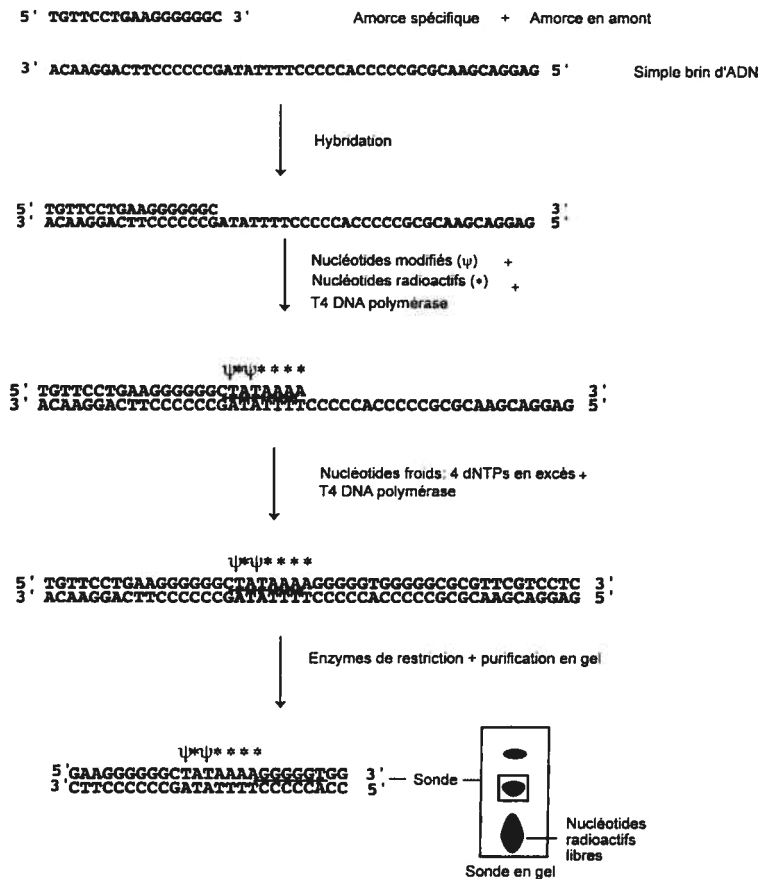
Identification des interactions entre les protéines du PIC et l'ADN du promoteur. L'expérience porte sur le complexe TBP-TFIIB-TFIIE-TFIIF-ARNP II-promoteur. La méthode est décrite en détail dans (Forget et Coulombe, 2003).

*La méthode de photopontage en gel peut se diviser en quatre étapes:*

1. Synthèse de la photosonde radioactive d'ADN du promoteur AdMLP dans laquelle un nucléotide photoréactif est incorporé à une (ou des) position(s) spécifique(s). Ce photonucléotide peut être activé en présence de radiation UV.
2. Le gel de retardement de bandes. Assemblage et isolement des complexes de pré-initiation sur la sonde *in vitro*.
3. Photopontage. Activation du photonucléotide avec des rayonnements UV dans le gel. Cette activation permettra la formation de liens covalents des protéines à proximité du photonucléotide.
4. Identification des protéines ayant ponté l'ADN à des positions spécifiques en utilisant un traitement dénaturant et une migration sur gel SDS-PAGE.

#### **2.1.2- Synthèse de la photosonde**

Le principe pour la formation de la sonde est illustré à la figure 16.



**Figure 16. Principe pour la formation de la photosonde radioactive intégrant le photonucléotide à des positions spécifiques sur le promoteur.**

La préparation des sondes s'effectue sur glace. La première réaction consiste à mélanger 500 ng de l'ADN simple brin du promoteur AdMLP (environ 0.5 pmol) avec deux amorces (40 ng chacune), une amorce spécifique et une amorce en amont. L'amorce spécifique s'intègre à l'endroit où l'on veut placer le photonucléotide et l'amorce en amont se place beaucoup plus loin sur l'ADN et servira à compléter la synthèse de l'ADN double-brin. La réaction s'effectue en présence de 1  $\mu$ l de tampon A 10X (300 mM Tris-HCl, pH8.0, 500 mM KCl et 70 mM MgCl<sub>2</sub>). Les tubes tests sont chauffés à 90°C pendant 5 min et incubés pendant 30 min à 25°C. À la suite de l'incubation, les nucléotides modifiés AB-

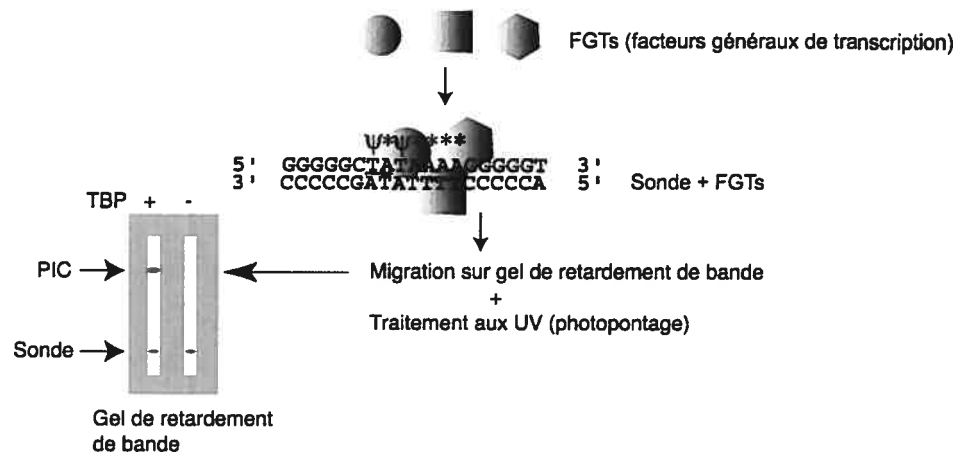
dUMP ou 5[N-(azidobenzoyl)-3-aminoallyl] (0.5 à 2  $\mu$ l) et les nucléotides radioactifs appropriés (20  $\mu$ Ci) sont ajoutés au mélange en présence de 5-10 U de T4 DNA polymérase (Pharmacia). La réaction s'effectue en présence de 0.5  $\mu$ l de BSA 10 mg/ml et de 1  $\mu$ l de tampon A (10X). La synthèse des nucléotides AB-dUMP a été décrite précédemment (Bartholomew et al., 1990). Le mélange est ensuite incubé à 25°C pendant 30 min pour permettre la polymérisation des NTPs radioactifs et photosensibles. Les 4 NTPs froids (concentration finale de 10 mM) sont ensuite ajoutés et incubés à 25°C pendant 5 min et à 37°C pendant 20 min pour compléter la formation de l'ADN bicaténaire. La sonde est digérée avec 10-20 U de SmaI. Les sondes sont purifiées par migration sur gel d'acrylamide 8% dans du tampon TBE (1X), pendant 60 min à 150 V. À la suite d'une autoradiographie du gel, la sonde est repérée et excisée du gel. L'élution de la sonde est faite dans une solution de Tris 10mM (160  $\mu$ l) pendant 16 h.

Puisque le nucléotide photosensible est dérivé de l'uracile, son incorporation sur l'ADN de la sonde par la T4 DNA polymérase s'effectuera à l'endroit d'insertion d'une thymine. L'intégration du photonucléotide aux positions -15, -10 et -5 a été possible en générant des mutations dans le promoteur, convertissant le nucléotide présent en thymine.

### **2.1.3- Gel de retardement de bandes (EMSA)**

Les sondes sont placées en présence des FGTs dont TBP, TFIIB, RAP30, RAP74, ARNP II, E34 et E56. La purification des facteurs a été décrite précédemment (Robert et al., 1996; Robert et al., 1998; Hodo, III et Blatti, 1977). Tous les facteurs sont ajoutés en quantité environ équimolaires dans la réaction : TBP (80 ng), TFIIB (80 ng), RAP30 (300 ng), RAP74 (600 ng), RNAP II (300 ng), TFIIE34 (160 ng), TFIIE56 (380 ng), dans 20  $\mu$ l de tampon G (12 mM HEPES, pH7.9, 60 mM KCl, 0.12 mM EDTA, 8 mM MgCl<sub>2</sub>, 50 ng/ml de BSA, 5 mM  $\beta$ -mercaptoethanol et 12% glycérol). Pour prévenir la formation de complexes non-spécifiques, le poly(dI-dC) est ajouté à la réaction (2.5 mg/ml) (Forget et al.,

1997). Finalement, la sonde (6000 cpm) est incorporée au mélange. La réaction est incubée pendant 30 min à 30°C. Les complexes sont ensuite migrés sur gel 4.5% de polyacrylamide :BAC (19 :1) à 400 V dans du TBE (0.5X) froid (Killeen et al., 1992). La synthèse du BAC (N,N'-Bisacryloylcysamine) est décrite dans Forget et Coulombe (2000). La figure 17 illustre le principe de la méthode.



**Figure 17. Assemblage du PIC sur la sonde radioactive et purification en gel natif.**

#### **2.1.4- Photopontage**

À la suite de cette migration, le gel est irradié aux UV (254 nm) pendant 10 min, induisant le photopontage entre les protéines et le photonucléotide. La séparation des complexes peut être observée sur PhosphoImager. Le complexe comprenant tous les facteurs de transcription se retrouvera dans le haut du gel (voir figure 17).

#### **2.1.5- Identification des protéines pontées**

La bande de gel contenant le complexe est alors excisée et incubée dans 10 µl de DTT (1 M) pendant 5 min à 37°C. Ensuite 40 µl du tampon ND (20 mM HEPES,

pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA, 20% glycérol et 0.4M DTT) est ajouté au mélange et les tubes sont incubés à 37°C pendant 25 min. Subséquemment, 13 µl de DNase (0.5 U/µl) diluée dans 30 mM CaCl<sub>2</sub> sont ajoutés pour digérer l'ADN non-ponté aux protéines. Une incubation de 20 min à 37°C est alors requise. 3.9 µl de SDS (10%) sont ajoutés au mélange et les tubes sont chauffés à 100°C pendant 3 min. Finalement, un mélange de 2.5% acide acétique contenant 15 mM ZnCl<sub>2</sub> (5.2 µl) et 80 000 U/ml nucléase S1 (2.6 µl) sont ajoutés à la réaction et incubés à 37°C pendant 20 min. La réaction est arrêtée en ajoutant 15 µl de tampon de charge (80 mM Tris, pH 6.8, 12.5% glycérol, 2.5% SDS, 0.9 M β-mercaptoéthanol, 0.2% bleu de bromophénol). L'échantillon doit bouillir 5 min à 100°C et migrer sur un gel SDS-PAGE, pendant 1 h à 30 mA et 3 h à 55 mA. Le gel est placé dans une cassette contenant une pellicule autoradiographique détectant la radioactivité des protéines pontées.

## ***2.2- La méthode TAP (Tandem Affinity Peptide)***

### **2.2.1- But de la méthode**

Cette méthode vise à purifier des complexes protéiques en conditions natives en créant une protéine de fusion comprenant une double étiquette d'affinité. La construction plasmidique contenant l'ADNc encodant la protéine de fusion est introduite dans des cellules humaines où elle sera exprimée et formera des complexes *in vivo* avec ses partenaires protéiques endogènes. Ces complexes pourront être purifiés et les composantes identifiés par spectrométrie de masse. La méthode TAP a été décrite dans (Vickers et Sharrocks, 2002; Puig et al., 2001; Rigaut et al., 1999).

*La méthode peut se diviser en quatre étapes :*

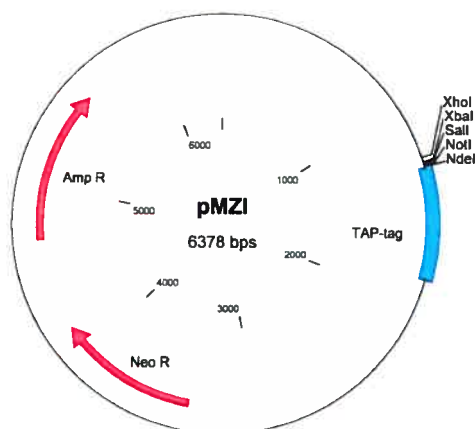
1. Construction d'un vecteur d'expression encodant la protéine d'intérêt (l'appât) fusionnée à une étiquette TAP : Le clonage permet d'insérer l'ADNc d'intérêt dans le plasmide pMZI (figure 18). Un PCR est réalisé dans le but d'introduire des sites de restrictions aux extrémités de l'ADNc d'intérêt. La digestion du

plasmide et du produit de PCR permet de former des extrémités cohésives pour la ligation des deux fragments (Sambrook et al., 1989). Les bactéries compétentes XL-1 sont transformées avec le produit de ligation.

2. Transfection des cellules humaines EcR-293 : Les cellules humaines EcR-293 sont transfectées avec le plasmide contenant l'ADNc d'intérêt. La méthode de coprécipitation au phosphate de calcium est utilisée pour la transfection. La détection des clones ayant intégré la construction plasmidique et exprimant la protéine appât suite à une induction avec la ponastérone A se fait par immunobuvardage, à l'aide d'un anticorps reconnaissant la protéine A de l'étiquette d'affinité.
3. Purification des complexes par double colonne d'affinité : Le clone sélectionné est produit en grande quantité et les protéines cellulaires totales sont extraites. L'appât et ses partenaires associés sont purifiés par deux passages successifs sur colonne d'affinité contenant des billes d'IgG et de calmoduline.
4. Identification des partenaires protéique associés à la protéine appât : Les éluats obtenus, contenant la protéine appât, sont migrés sur un gel SDS-PAGE et l'identification des polypeptides est effectuée par des analyses de spectrométrie de masse MS/MS.

### **2.2.2-Construction d'un vecteur d'expression encodant la protéine d'intérêt (appât) fusionnée à une étiquette TAP**

Les clones contenant l'ADNc d'intérêt sont obtenus de Invitrogène. L'ADNc d'intérêt est incorporé dans le plasmide pMZI. Ce plasmide code pour l'étiquette TAP en C-terminal (TAP-TAG). Le TAP-TAG comprend un double site de liaison pour les IgG de la protéine A de *Staphylococcus aureus* et un peptide de liaison à la calmoduline (CBP), séparés par un site de clivage de la protéase TEV.



**Figure 18. Carte du plasmide pMZI.** Ce plasmide comprend un gène de résistance à l'ampicilline, un gène de résistance à la néomycine et le TAP-TAG situé en aval du site de clonage multiple.

#### 2.2.2.1- PCR

Le PCR avec l'ADNc d'intérêt permet l'insertion de sites de restriction aux extrémités 3' et 5'. Les sites de restriction Not1 (en 5') et Nde1 (en 3') sont ajoutés à l'ADNc de XAB1 et les sites Xho1 (en 5') et Not1 (en 3') sont ajoutés à l'ADNc de DDX5. La réaction de PCR contient : 1 ng de matrice d'ADN, 17.5  $\mu$ M de chaque amorce, 7.5  $\mu$ M de dNTPs, 2.5 U de Pfu turbo (Invitrogene), le tampon Pfu (1X). Le programme comprend les étapes suivantes : 95°C / 2 min, 95°C / 30 sec, (Tm-5) / 30 sec, 72 °C / (1 min par kb), les trois dernières étapes recommencent en boucle 30 fois et finalement 72°C / 10 min.

#### 2.2.2.2- Digestion du produit de PCR

Le produit de PCR est purifié avec le kit de purification de Qiagen et digéré avec les enzymes Not1 et Nde1 (Pharmacia) ou les enzymes Xho1 et Not1 (*New England Biolabs*). Le vecteur pMZI et le produit de PCR sont digérés pendant 16 h dans le tampon recommandé par le fabricant. Les produits de digestion sont



migrés sur gel d'agarose 1% et l'ADN est récupéré en utilisant le kit d'extraction de gel de Qiagen. L'ADN est quantifié par spectrométrie en utilisant une longueur d'onde de 260 nm (1 unité  $A_{260}$  d'ADN double brin, correspond à 50  $\mu\text{g/ml H}_2\text{O}$ ).

#### **2.2.2.3- Ligation du vecteur et de l'ADNc d'intérêt**

La réaction de ligation est réalisée à 16°C pendant 3 h. Le mélange réactionnel pour la ligation comprend: 1 U de T4 DNA ligase, 20 fmol d'ADN de l'insert (ADNc d'intérêt) et 60 fmol d'ADN du vecteur. La réaction de ligation se produit dans 20  $\mu\text{l}$  de tampon de ligation 1X (Invitrogene).

#### **2.2.2.4- Transformation des bactéries compétentes**

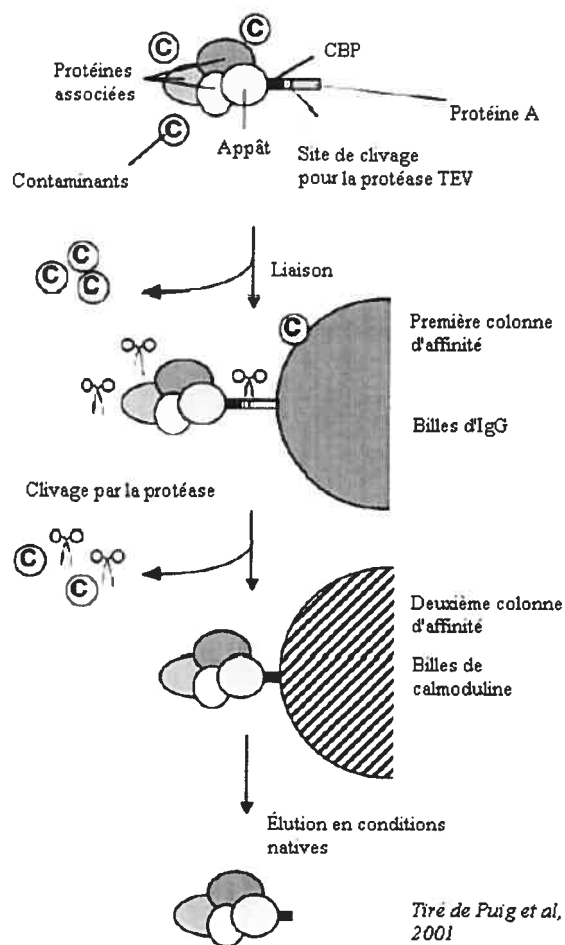
Les bactéries compétentes XL-1 sont transformées avec le produit de ligation. Tout le produit de ligation est ajouté à 100  $\mu\text{l}$  de bactéries compétentes et le mélange est incubé sur glace pendant 30 min. Le choc thermique subséquent est produit par une incubation de 90 sec à 42°C, suivi d'un repos sur glace pendant 2 min. 400  $\mu\text{l}$  de milieu LB sont ensuite ajoutés au mélange et la réaction est incubée à 37°C pendant 60 min. Les bactéries compétentes sont ensuite étalées sur un pétri avec 100  $\mu\text{g/ml}$  d'ampicilline. Les colonies bactériennes qui ont poussé après une incubation de 16 h sont testées par PCR. La réaction de PCR est la même que décrite précédemment, sauf que la matrice provient d'un lysat de bactéries (préalablement diluées dans 5  $\mu\text{l}$  d'eau) traité aux micro-ondes pendant 90 sec. La colonie positive est séquencée et le plasmide est produit en grande quantité (kit de Qiagen).

### **2.2.3- Transfection des cellules humaines EcR-293 : Méthode de phosphate de calcium**

Les cellules EcR-293 sont cultivées dans le milieu DMEM contenant 10% de FBS (*fetal bovine serum*), 2 mM de glutamine, 100 µg/ml de pénicilline, 100 µg/ml de streptomycine et 30 µg/ml de bléocine. Les cellules sont cultivées jusqu'à l'obtention de  $2 \times 10^6$  cellules par pétri de 100 mm. Après 24 h d'incubation dans un incubateur humidifié ayant une température de 37°C et 5% de CO<sub>2</sub>, les cellules sont transfectées avec un mélange de 500 µl comprenant : 8 µg d'ADN et 250 mM de CaCl<sub>2</sub>; ce mélange est complétement délicatement de 500 µl de HeBS (2x) (1.5 mM Na<sub>2</sub>HPO<sub>4</sub>.7H<sub>2</sub>O, 274 mM NaCl, 55 mM HEPES, pH 7.0). La réaction est incubée 10 min à 25°C avant d'être ajoutée aux cellules EcR-293 en croissance. Après 24 h, les cellules sont transférées dans le milieu de sélection DMEM contenant 300 µg /ml de généticine, 10% de FBS, 2 mM de glutamine et 30 µg/ml de bléocine. Les clones cellulaires sont isolés et produits en grande quantité. Un test d'expression de la protéine d'intérêt est effectué en induisant les clones avec 3 mM de ponasterone A (voir système d'expression-ecdysone à la figure 15) pendant environ 24 h. L'immunobuvardage est utilisé pour vérifier l'induction de la protéine. Trois clones positifs sont sélectionnés, congelés et testés en purification sur colonne par la méthode TAP.

### **2.2.4- Purification des complexes par double chromatographie d'affinité**

Le principe de la méthode est présenté à la figure 19.



**Figure 19. Résumé de la purification par la méthode TAP.**

L'extrait cellulaire comprenant la protéine appât et ses partenaires est placé en présence des billes d'IgG. Par affinité, la protéine A de *Staphylococcus aureus* lie les IgG. Les billes sont lavées et la protéine d'intérêt est récupérée en clivant une partie de l'étiquette avec la protéase TEV qui reconnaît le site de clivage situé entre le site de liaison des IgG de la protéine A et le CBP. Une deuxième colonne d'affinité faite de billes de calmoduline permet la liaison du CBP, et, suite au lavage, la protéine appât est récupérée par l'ajout d'un agent chélateur des ions calcium (EGTA). Toutes les procédures ont lieu dans des conditions natives à 4°C.

#### **2.2.4.1- Extraction des protéines cellulaires totales**

Les cellules doivent être homogénéisées sur glace afin d'obtenir un extrait total. Cette extraction s'effectue dans des conditions natives (faible quantité de sel). Un volume et 1/3 de tampon A (10 mM HEPES, pH 7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM de KCl, 0.5 mM de DTT, 0.5 mM AEBSF et 1 tablette de *Complete EDTA-free* #1873580 de Roche) sont ajoutés aux cellules récoltées. Les cellules sont homogénéisées 10 fois à l'aide d'un homogénéisateur manuel. Ensuite, un volume de solution B (50 mM HEPES, pH 7.9, 1.5 mM MgCl<sub>2</sub>, 0.5 mM de DTT, 0.5 mM AEBSF et 1.26 M KAc) est ajouté et les cellules sont homogénéisées 10 autres fois dans cette solution.

#### **2.2.4.2- Passage sur la première colonne d'affinité contenant les billes d'IgG**

L'extrait est placé en présence des billes d'IgG (50 µl par g de cellules traitées) préalablement équilibrées dans le tampon IPP (10 mM Tris, pH 8, 100 mM NaCl, 0.1% Triton X100, 10% glycérol) et incubé pendant 1 h à 4°C. Les billes sont ensuite placées dans une colonne pour procéder aux lavages avec le tampon IPP. Par la suite, dans le tampon TEV (10 mM Tris, pH 8, 100 mM NaCl, 0.1% Triton X100, 0.5 mM EDTA, 10% glycérol et 1 mM DTT), 30 U de protéase TEV (Invitrogen) sont ajoutés aux billes par g de cellules traitées pour libérer l'appât et les protéines liées à l'appât. Cette réaction a lieu pendant 16 h à 4°C.

#### **2.2.4.3- Passage sur la deuxième colonne d'affinité contenant les billes de calmoduline**

Le premier éluat est récupéré dans un tube contenant les billes de calmoduline (50 µl par g de cellules traitées) préalablement équilibrées avec le tampon CBB (10 mM Tris, pH 8, 100 mM NaCl, 1 mM imidazole, 1 mM MgAc, 2 mM CaCl<sub>2</sub>, 0.1% Triton X100, 10% glycérol, 10 mM β-mercaptoéthanol). Les billes et l'éluat sont ensuite incubées à 4°C pendant 2 h en présence de CaCl<sub>2</sub> (2 mM). Les billes sont placées dans une colonne pour procéder aux lavages avec le tampon CBB. La

colonne est éluée en présence d'un agent chélateur de calcium, l'EGTA, pour rompre la liaison du CBP avec la calmoduline. Ainsi, 2 volumes de colonne, suivi de 3 volumes de colonne du tampon CEB (10 mM Tris, pH 8, 100 mM NaCl, 1 mM imidazole, 1 mM MgAc, 2 mM EGTA, 10% glycérol, 10 mM  $\beta$ -mercaptoéthanol) seront utilisés pour récupérer la protéine appât et ses partenaires.

#### **2.2.4- Identification des partenaires protéiques de la protéine appât par spectrométrie de masse**

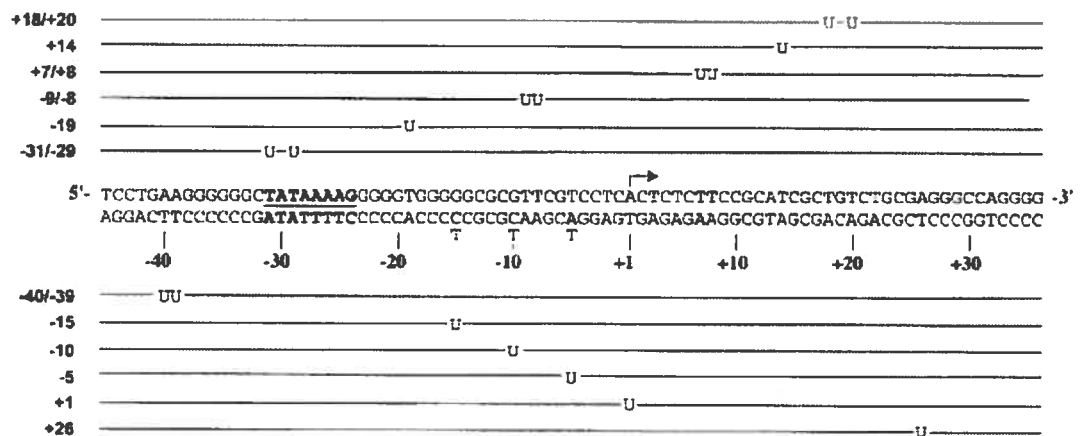
Les protéines sont ensuite concentrées à l'aide d'un concentrateur rotatif avant d'être chargées sur gel SDS-PAGE. Les protéines sont préalablement bouillies à 100°C pendant 5 min. Les bandes protéiques obtenues sont ensuite identifiées par spectrométrie de masse MS/MS. Pour procéder à la spectrométrie de masse, les bandes de gel sont découpées et les protéines sont digérées en peptides par une digestion avec une protéase du pancréas, la trypsine. Ces fragments sont ensuite ionisés. Lorsque le spectromètre de masse rencontre un ion, il le fragmente et la masse des fragments générés est mesurée. Les spectres MS/MS de l'échantillon sont comparés aux spectres théoriques pour retrouver les séquences correspondantes (Delahunty et Yates III, 2005).

## 3- RÉSULTATS

### ***3.1- Analyse de la topologie du complexe TBP-TFIIB-TFIIIE-TFIIF-ARNP II en utilisant le photopontage en gel***

[NOTE : Au niveau des résultats de photopontage présentés dans ce mémoire, j'ai réalisé les figures 21, 22, 25 et j'ai participé aux résultats de la figure 24, ceux-ci ayant été principalement effectués par Diane Forget. La figure 23 a été conçue par Vincent Trinh. Les études de photopontage ont mené à la publication d'un article (Forget D, Langelier MF, Therien C, Trinh V, Coulombe B (2004) Mol Cell Biol. 24:1122-3). Dans cet article, j'ai démontré que nos conditions d'assemblages étaient efficaces pour former un complexe TBP-TFIIB-TFIIIE-TFIIF-ARNP II-promoteur (figure 2A et B). J'ai localisé Rpb5 relativement à l'ADN du promoteur (figure 4B) et reproduit pour validation certains résultats de Diane Forget concernant la localisation de Rpb1, Rpb2 et des FTGs (figure 3).]

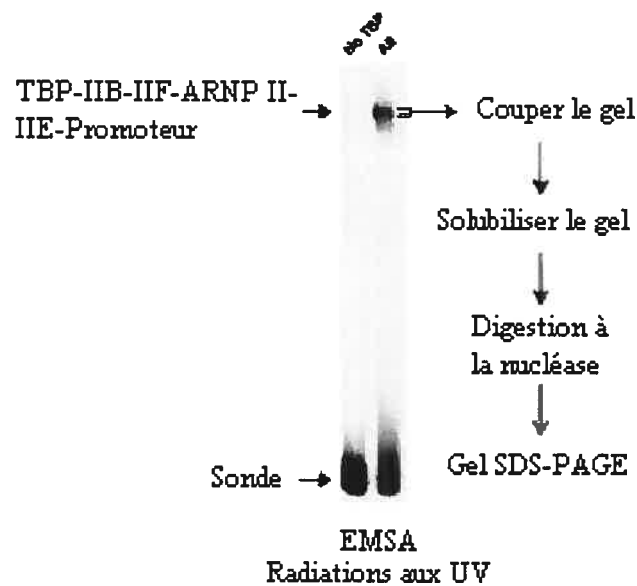
Dans le but d'élucider l'organisation moléculaire du PIC de l'ARNP II, une méthode qui permet de ponter les protéines à des sites spécifiques sur l'ADN a été utilisée. Douze sondes placant un photonucléotide à des positions déterminées du promoteur AdMLP juxtaposé d'un nucléotide radioactif ont été utilisées afin d'évaluer la position des sous-unités de l'ARNP II et des FGTs relativement à l'ADN (figure 20).



**Figure 20. Sondes utilisées pour l'analyse topologique du PIC.**

La boîte TATA est montrée en caractères gras et le site +1 est représenté par une flèche. Les différentes sondes placent le photonucléotide (U) à des sites spécifiques le long de l'ADN du promoteur entre les positions -40 et +26 sur les brins codants et non-codants. Des mutations (N à T) sont indiquées pour quelques photosondes.

Chaque photosonde a été mise en présence des FGTs TBP, TFIIB, TFIIE et TFIIF et de l'ARNP II dans des conditions favorisant l'assemblage des complexes de pré-initiation (figure 21). Les complexes ont été séparés dans des conditions natives par EMSA et irradiés aux UV. Suite à l'excision des bandes et la solubilisation du gel, les complexes TBP-TFIIB-TFIIE-TFIIE-ARNP II-promoteur ont été traités à la nucléase S1 et la DNase I. Les protéines pontées ont ensuite été migrées sur SDS-PAGE pour permettre leur identification.

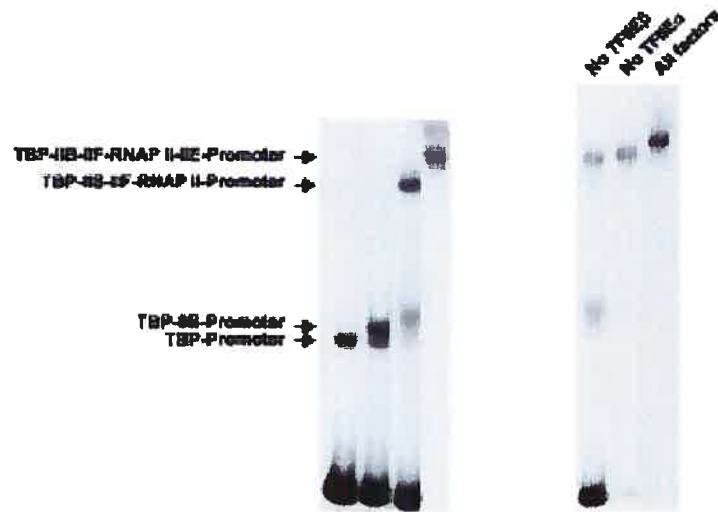


**Figure 21. Représentation schématique de la méthode de photopontage en gel des PIC de la transcription par l'ARNP II.**

Les analyses ont été effectuées en présence et en absence de TBP pour obtenir un contrôle de spécificité. L'absence de TBP prévient la formation de complexes spécifiques, mimant l'absence de boîte TATA au promoteur (Forget et Coulombe, 2003).

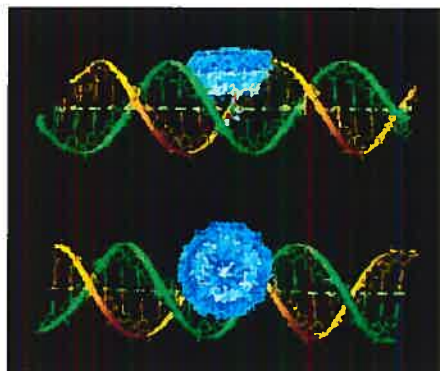
Afin de vérifier que le complexe formé à la figure 21 est bien un complexe TBP-TFIIB-TFIIE-TFIIF-ARNP II-promoteur, nous avons réalisé des expériences de EMSA utilisant diverses combinaisons de FGTs. La figure 22 montre que tous les facteurs sont requis pour obtenir le complexe de faible mobilité obtenue à la figure 21, indiquant que nos conditions d'assemblages sont efficaces et que le complexe étudié est bien un complexe TBP-TFIIB-TFIIE-TFIIF-ARNP II-promoteur.





**Figure 22. EMSA démontrant les complexes intermédiaires assemblés sur la sonde en utilisant différentes combinaisons des FGTs et l'ARNP II.**

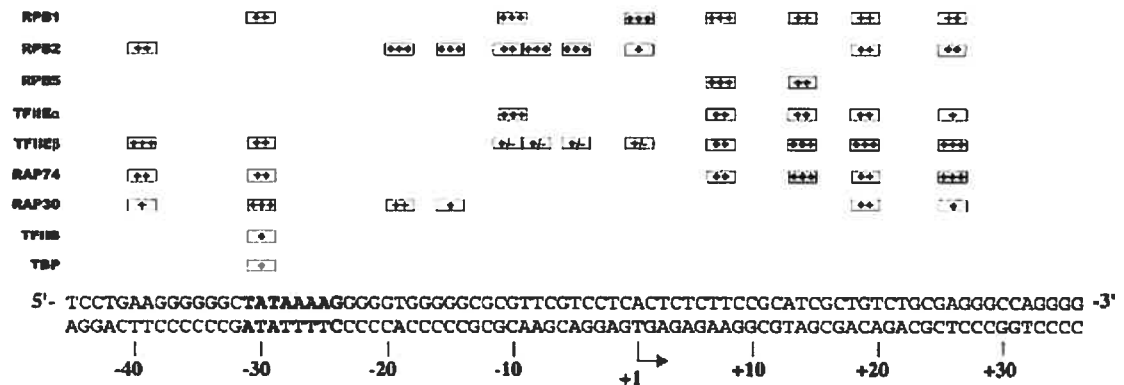
Le photonucléotide AB-dUTP peut adopter différentes conformations résultant de la rotation du groupement aryl autour des axes de certains liens chimiques. Les différentes configurations adoptées ont été modélisées (figure 23). Le modèle indique que le groupement nitrene réactif du photonucléotide AB-dUTP peut occuper un espace de 20 par 20 par 10 Å dans le sillon majeur de l'hélice d'ADN. Les protéines situées dans cet espace pourront être pontées de manière covalente par le photonucléotide.



Tiré de Forget et al, 2004

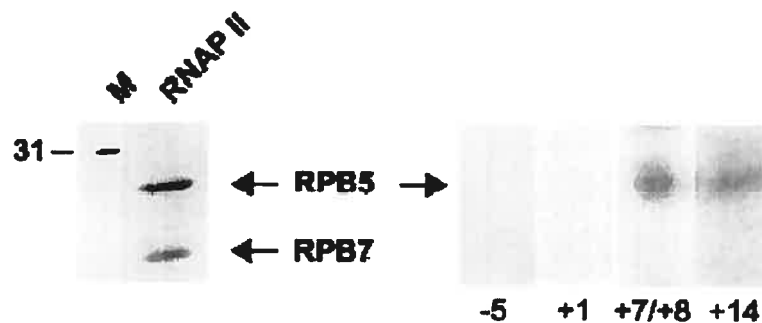
**Figure 23. Les conformères du photonucléotide (AB-dUMP) au promoteur AdMLP.**

Les résultats obtenus en utilisant les douze sondes du promoteur AdMLP sont montrées à la figure 24.



**Figure 24. Résumé des résultats de photopontage en gel des complexes de pré-initiation comprenant les facteurs TBP, TFIIB, TFIIF (RAP30 et RAP74), TFIIE (TFIIE $\alpha$  et TFIIE $\beta$ ) et l'ARNP II. Les boîtes indiquent les photopontages des facteurs et de l'ARNP II à différentes positions sur le promoteur AdMLP. Les symboles +++, ++ et + représentent l'intensité moyenne des signaux de photopontage obtenus pour chaque photosonde. Les symboles +/- de TFIIE $\beta$  avec les sondes -10, -9/-8, -5 et +1 indique une incertitude face la spécificité des complexes (Forget et al., 2004).**

La figure 25 montre les résultats de photopontages obtenus pour la sous-unité Rpb5 de l'ARNP II. Les protéines pontées ont été migrées sur un gel SDS-PAGE en parallèle avec l'ARNP II purifiée. Les bandes correspondant à l'ARNP II ont été colorées à l'argent et identifiées par spectrométrie de masse et la partie de gel contenant les protéines pontées a été soumise à l'autoradiographie.



**Figure 25. Exemple de gel SDS-PAGE identifiant la sous-unité Rpb5.**

***Les contacts de l'ARNP II avec l'ADN du promoteur (figure 24)***

Rpb1 fait contact avec le photonucléotide à plusieurs positions, dans une région s'étendant de -31/-39 à +26. Rpb2 est ponté entre les positions -40/-39 et +26, et des pontages importants sont observés autour de la position -10, suggérant que cette région du promoteur est très proche de Rpb2. Rpb1 fait contact avec l'ADN situé au niveau de la boîte TATA (-31/-29), alors que Rpb2 ponté l'ADN en amont de la boîte TATA aux positions -40/-39. Finalement, et tel que montré précédemment, Rpb5 ponté l'ADN aux positions +7/+8 et +14, soit seulement en aval du site +1. Aucune petite sous-unité de l'ARNP II (Rpb6 à Rpb12) n'a réagit avec les photosondes.

***Les contacts des FGTs avec l'ADN du promoteur (figure 24)***

Les résultats de la figure 24 montrent que RAP30 se situe près de l'ADN en amont (-40/-39 et -15) et en aval (+18/+20 et +26) du site +1. Pour sa part, RAP74 fait contact avec l'ADN entre les positions +7/+8 et +26 et aux positions -31/-29 et -40/-39. De manière générale, les polypeptides RAP30 et RAP74 se situent de part et d'autre des positions +7/+8 et -15, soit relativement éloignées du site +1.

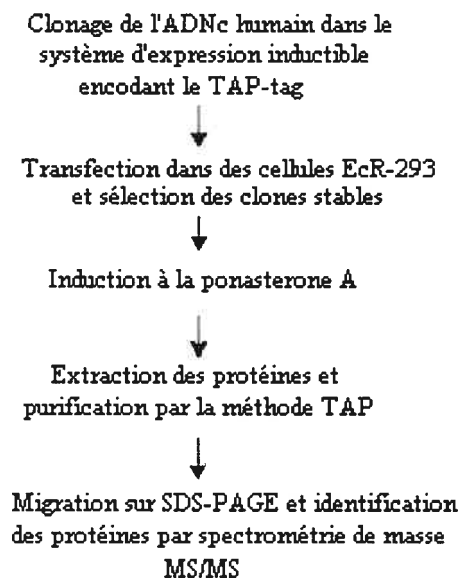
Les sous-unités du facteur TFIIE pontent fortement l'ADN entre les positions +7/+8 et +26. De plus, il est à noter que TFIIE $\alpha$  pontent l'ADN d'une manière marquée à la position -10. Le facteur TFIIE $\beta$ , pour sa part, présente deux autres pontages; l'un au niveau de la boîte TATA et un autre à la position -40/-39. Par contre, au niveau des positions -10, -9/-8, -5 et +1, nous ne pouvons être certains des pontages de TFIIE $\beta$ , car ces 4 sondes permettent l'assemblage de complexes en absence du facteur TBP. Ces complexes non-spécifiques, obtenus en absence de TBP suggèrent l'agrégation de molécules de TFIIE $\beta$ . Un fait à noter : TFIIE $\beta$  interagit avec le photonucléotide aux mêmes positions que RAP74. Finalement, les facteurs TFIIB et TBP pontent l'ADN seulement au niveau de la boîte TATA.

### ***3.2- Identification des partenaires d'interaction des protéines XAB1 et DDX5***

[NOTE : Les résultats de purification par la méthode TAP des partenaires d'interaction de XAB1 et DDX5 présentés dans ce mémoire ont tous été réalisés par moi-même. Ces résultats contribueront à la publication d'un article décrivant le réseau d'interaction de la machinerie transcriptionnelle basale (Jeronimo et al, 2005; en préparation).]

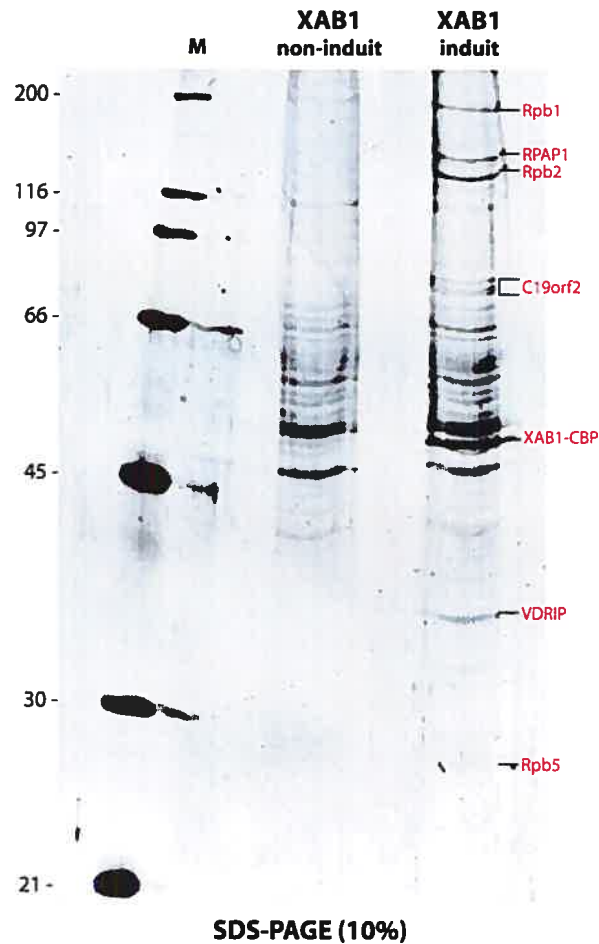
Les protéines XAB1 et DDX5 ont été identifiées suite à la purification d'affinité par la méthode TAP de la machinerie transcriptionnelle basale. XAB1 a été identifié en utilisant exclusivement Rpb2, Rpb11 et RPAP1 comme appât et DDX5 a été identifié exclusivement avec l'appât TFIIS (voir section 1.7.2). Dans le but d'identifier les partenaires protéiques *in vivo* de XAB1 et DDX5, la méthode TAP a été appliquée. L'étude vise essentiellement à confirmer l'interaction de ces protéines avec la machinerie transcriptionnelle basale et de découvrir de nouveaux partenaires protéiques, le cas échéant.

Pour ce faire, des lignées cellulaires stables capables d'exprimer XAB1 et DDX5 suite à une induction à la ponastérone A ont été générées. La stratégie globale de la méthode est présentée à la figure 26



**Figure 26. Stratégie utilisée pour l'identification des partenaires protéiques *in vivo* de XAB1 et DDX5.**

La figure 27 montre les résultats de purification TAP obtenus avec la protéine XAB1. Les cellules ont été cultivées en présence et en absence de ponastérone A pour obtenir un contrôle de spécificité. Les bandes ont été colorées par SYPRO ruby et identifiées par spectrométrie de masse MS/MS.



**Figure 27. Gel coloré au SYPRO Ruby présentant les partenaires protéiques (directs ou indirects) de XAB1 identifiés par spectrométrie de masse.**

Les résultats de spectrométrie de masse sont représentés par un pointage qui permet d'évaluer le niveau de confiance de l'identification. En spectrométrie de masse, plusieurs éléments doivent être considérés, notamment le nombre de peptides identifiés et la qualité du spectre de masse. Le tableau IV montre les partenaires protéiques de XAB1 obtenus par la purification d'affinité TAP. Le pointage pour chacune des identifications est indiqué.

**Tableau IV. Identification des partenaires protéiques de XAB1 par spectrométrie de masse MS/MS**

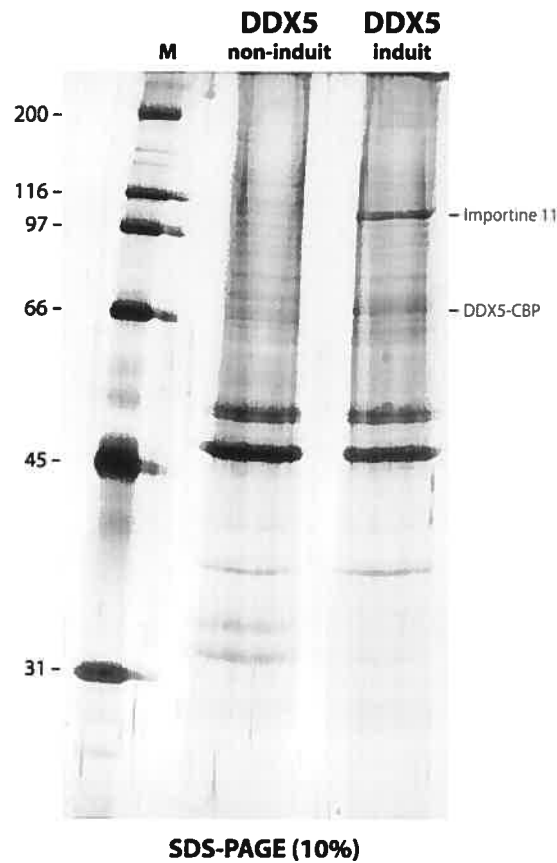
Protéines	Poids moléculaire estimé sur gel (kDa)	Pointage Mascot / nombre de peptides	
XAB1 (appât)	45	410 / 29	454 / 16
Rpb1	200	142 / 5	94 / 4
Rbp2	125	317 / 9	445 / 14
Rpb5	28	44 / 1	133 / 6
RPAP1	140	141 / 4	218 / 11
C19orf2	60	82 / 3	-
VDRIP	35	36 / 1	-

**Légende** : Rouge = Très bon pointage; Noir = bon pointage

Le tableau IV montre que l'appât XAB1, 3 sous-unités de l'ARNP II et la protéine RPAP1 ont un pointage élevé qui est validé par une double identification en spectrométrie de masse. Ces 5 polypeptides ont été identifiés antérieurement en utilisant différents appâts (voir section 1.7.2). D'autre part, la protéine C19orf2, impliquée dans la répression de la transcription par sa liaison à Rpb5 et la protéine VDRIP, une composante du médiateur de l'ARNP II (voir section 1.3.2.3) ont été nouvellement identifiées. Toutefois, les résultats de spectrométrie de masse avec ces deux protéines sont faibles et ont été obtenus qu'une seule fois. Nous considérons donc ces identifications comme "tentatives". Mes résultats confirment la présence d'un sous-complexe comprenant XAB1, l'ARNP II et

RPAP1 et suggèrent l'existence de partenaires (direct ou indirect) additionnels, C19orf2 et VDRIP.

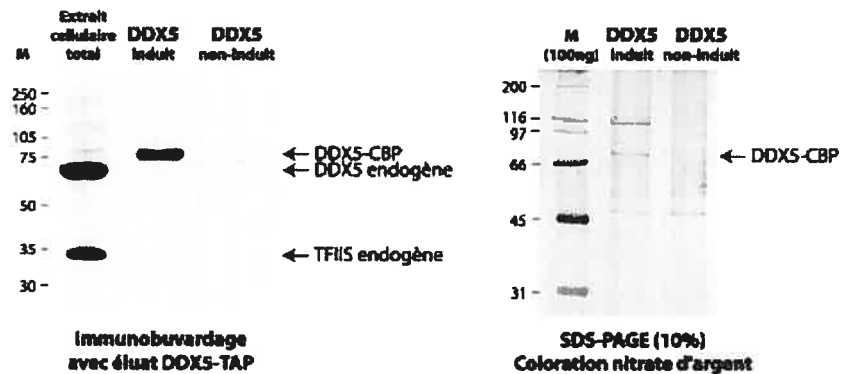
La protéine DDX5 a été identifiée antérieurement comme partenaire d'interaction du facteur d'élongation TFIIIS (voir section 1.7.2.1). La figure 28 montre les résultats de purification TAP obtenus avec la protéine DDX5. La coloration a été effectuée avec le nitrate d'argent.



**Figure 28.** Gel coloré au nitrate d'argent présentant les partenaires protéiques de DDX5 identifiés par spectrométrie de masse.



La protéine appât DDX5 a un poids de 68 kDa sur le gel. Un seul partenaire est retrouvé à environ 100 kDa et correspond à une importine (IPO11). L'intensité de la bande correspondant à l'importine 11 est de plus forte intensité que celle correspondant à l'appât. Le facteur TFIIS n'a pas été identifié sur le gel comme il était attendu. Dans le but de confirmer l'absence de TFIIS dans la purification par TAP de DDX5, un test d'immunobuvardage a été réalisé (figure 29).



**Figure 29. Immunobuvardage réalisé sur l'éluat de DDX5 purifié par la méthode TAP.**

Un extrait cellulaire total a été utilisé dans les essais pour obtenir un contrôle de spécificité des anticorps. En utilisant l'anticorps anti-DDX5, un signal est présent dans l'échantillon induit et absent dans le non-induit. Par contre, aucun signal n'est obtenu avec l'anticorps anti-TFIIS dans les éluats induit et non-induit de DDX5 (figure 29, à gauche). Ces mêmes éluats ont été migrés sur un gel SDS-PAGE et colorés au nitrate d'argent afin de vérifier le bon fonctionnement de la purification par affinité TAP (figure 29, à droite). Mes résultats ne permettent pas de confirmer l'interaction de DDX5 avec TFIIS, mais suggèrent l'existence d'un nouveau partenaire, IPO11.

## 4- DISCUSSION

### *4.1- Analyse topologique du PIC*

#### *Le photopontage protéine-ADN en gel : une nouvelle méthode d'étude des complexes nucléoprotéiques.*

Le photopontage des protéines est une méthode de choix pour analyser l'organisation moléculaire de grands complexes nucléo-protéiques. Par contre, les premières études de photopontage étaient vraisemblablement limitées, dû au fait que la spécificité d'assemblage des complexes ne pouvait pas être vérifiée. Puisque la plupart des FGTs et l'ARNP II ont une affinité pour l'ADN, il est souvent difficile d'ajuster les conditions qui permettront de former exclusivement des complexes spécifiques sur les photosondes. Pour surpasser cette limite, nous avons développé une méthode qui permet de purifier les complexes en gel natif et d'étudier les pontages protéine-ADN dans chaque complexe isolé. La méthode est décrite dans la section 3.1 (figure 21).

Des essais de retard sur gel, en omettant certains facteurs, ont été réalisés afin de vérifier si nos conditions expérimentales permettent l'assemblage des PIC. La figure 22 montre que tous les facteurs s'assemblent en un complexe unique, qui représente un complexe TBP-TFIIB-TFIIIE-TFIIF-ARNP II-promoteur.

L'utilisation du logiciel *Sybyl Molecular Modeling Package* a permis d'élaborer un modèle illustrant les conformères du photonucléotide (voir figure 23). Le groupement nitrène réactif du photonucléotide AB-dUTP a été positionné relativement à l'ADN du promoteur. Ce dernier peut s'étendre dans l'espace de 20 Å par 20 Å par 10 Å dans le sillon majeur de l'hélice d'ADN. Ces résultats ont permis de construire avec une plus grande précision le modèle topologique du PIC (voir plus bas et figure 30).

La figure 25 montre Rpb5 liée a un fragment radioactif des photosondes +7/+8 et +14, migrant sur une distance d'environ 30 kDa. L'identification des protéines pontées se base principalement sur le poids moléculaire des polypeptides pontés. Par contre, certaines protéines peuvent co-migrer ce qui peut rendre l'interprétation difficile. Ainsi, l'identité des polypeptides pontés est souvent confirmée par immunoprécipitation ou par l'emploi de facteurs étiquetés dans la réaction de photopontage.

### ***Comparaison des photopontages obtenus avec les études antérieures***

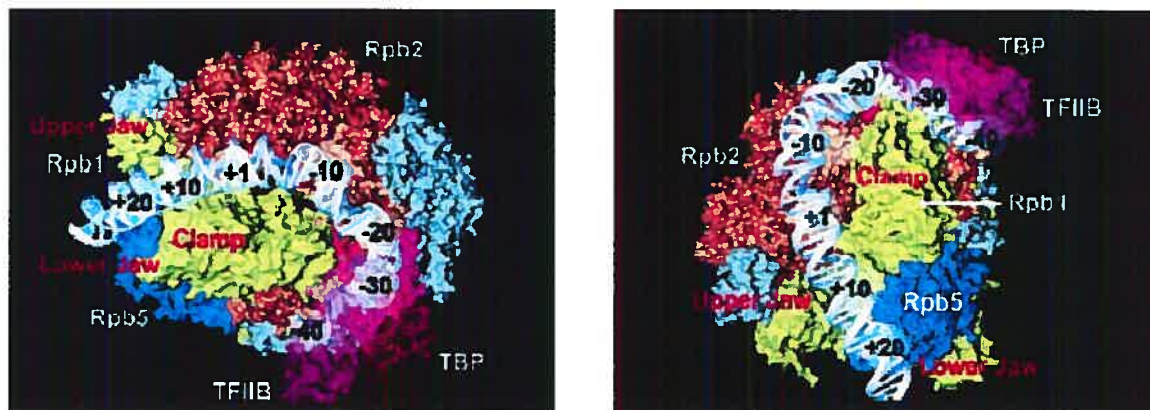
Comparativement aux résultats que nous avons obtenu en utilisant la méthode de photopontage classique, le photopontage en gel d'un complexe TBP-TFIIB-TFIIE-TFIIF-ARNP II-promoteur purifié réduit légèrement le nombre total de pontage de la machinerie transcriptionnelle avec l'ADN du promoteur. Cette réduction s'applique plus spécialement aux facteurs TFIIF et TFIIE qui font moins de contacts avec l'ADN, notamment autour du site +1 (Robert et al., 1998; Douziech et al., 2000). Rpb5 n'a pas été identifié dans les analyses de photopontage classique. Toutefois, le positionnement de Rpb5 dans la structure cristallographique de l'enzyme appuie nos résultats montrant des pontages en amont du site +1.

Dans plusieurs publications du laboratoire Coulombe, précédant Forget *et al* (2004), aucun pontage de TFIIE56 n'avait été obtenu en l'absence de TFIIH; en présence de ce facteur TFIIE56 ponte l'ADN en amont de la boîte TATA, au niveau de la bulle transcriptionnelle et en aval du site +1 (Douziech et al., 2000). Nos résultats de photopontage en gel utilisant des quantités équimolaires des FGTs montrent que TFIIE56 peut s'intégrer de manière stable dans le complexe, même en absence de TFIIH. Des contacts plus prononcés de TFIIE56 en amont du site +1 sont aussi observés.

Toujours comparé aux résultats précédents, la longueur du promoteur faisant contact avec la machinerie transcriptionnelle reste inchangée. L'ARNP II pontre l'ADN entre les régions -40 et +26, correspondant à une région de 66 pb (225Å). Cette dimension est nettement plus grande que l'ARNP II mesurant 140 Å sur sa dimension la plus longue. Ces résultats supportent le modèle d'enroulement de l'ADN autour de l'enzyme (voir section 1.4.2). L'enroulement serré du promoteur autour de l'ARNP a aussi été retrouvée chez les procaryotes (Rivetti et al., 1999), ce qui suggère qu'il s'agisse d'un mécanisme fondamental.

### *Pontages avec l'ARNP II*

Le modèle présenté à la figure 30 montre la trajectoire de l'ADN dans le PIC en se basant sur nos résultats obtenus par le photopontage en gel. Seule cette trajectoire du promoteur semble permettre d'expliquer les pontages spécifiques des sous-unités de l'ARNP II.



**Figure 30. Modèle représentant la trajectoire empruntée par l'ADN dans la structure de l'enzyme ARNP II.** Seuls les facteurs TBP (rose) et TFIIB (magenta) sont montrés par souci de clarté. Les sous-unités Rpb1 (jaune), Rpb2 (orange) et Rpb5 (bleu) de l'ARNP II sont indiquées.

Dans notre modèle, la partie en aval du promoteur semble suivre le tunnel localisé entre les deux mâchoires de l'enzyme (*jaws*), impliquant des contacts avec Rpb1, Rpb2 et Rpb5. Entre les positions +1 et -20, l'ADN se rapproche de Rpb2 et bifurque vers la droite pour guider la boîte TATA du promoteur vers TBP et TFIIB. Une légère courbure redirige le promoteur vers Rpb2 dans la région -40.

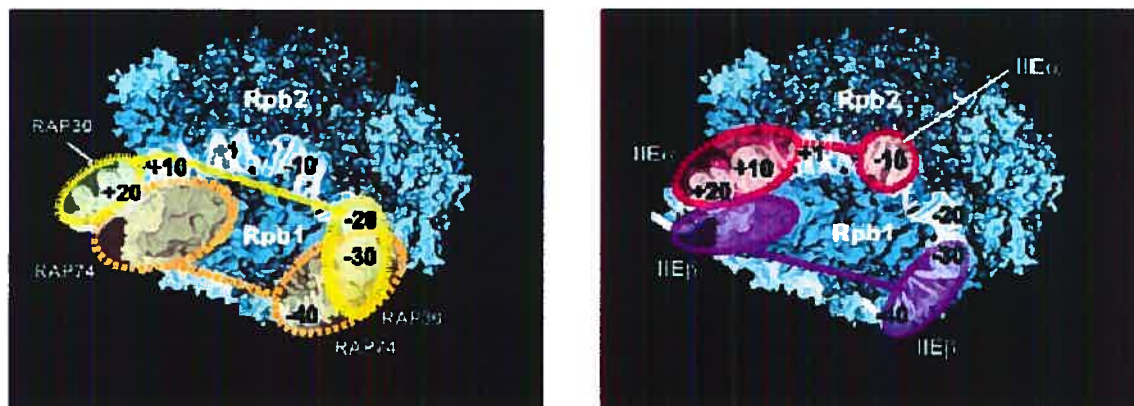
Le *clamp*, est formée de Rpb1, Rpb2 et Rpb6. Ce dernier peut se déplacer pour ouvrir et fermer le tunnel de l'ARNP II et stabiliser la matrice d'ADN au site actif. Ces données sont cohérentes avec notre modèle où l'ADN entre dans l'enzyme au niveau des mâchoires et conserve des interactions avec le *clamp* sur une très grande distance. À la figure 30, on observe que la courbure de l'ADN se produit au niveau du côté gauche de la surface du *clamp*, suggérant un rôle de ce dernier dans l'initiation de la transcription. Un enroulement vers la gauche de l'ADN induirait une relaxation de l'ADN, produisant un désappariement partiel de la double hélice. Cette relaxation de l'ADN du promoteur pourrait jouer un rôle important en fournissant une région d'ADN simple brin nécessaire à l'action des hélicases à ADN simple brin de TFIIF pour la formation du complexe ouvert (Douziech et al., 2000).

Des études de photopontage chez la levure réalisée en 2004 ont permis d'observer un patron de pontage très similaire entre le PIC de la levure et celui employant un système transcriptionnel humain (Chen et al., 2004). Ces résultats supportent la notion que la structure du PIC soit similaire chez tous les eucaryotes. Par contre, une différence notable entre les pontages de l'ARNP II de la levure et des eucaryotes supérieurs est le pontage de Rpb7 au lieu de Rpb5. Cette distinction reflète une différence dans les mécanismes transcriptionnels entre les deux systèmes.

### ***Pontages avec les facteurs de transcription***

L'interaction des facteurs TFIIB et de TBP avec la même sonde (position -29/-31) suggère que ces deux facteurs sont situés très près l'un de l'autre dans le PIC. En support à ces données, le co-cristal de TBP-TFIIB à 4.5Å a révélé que le domaine C-terminal de TFIIB (TFIIBc) interagit avec TBP et l'ARNP II. De plus, la structure du complexe avec un fragment d'ADN du promoteur a révélé l'interaction de TFIIBc avec l'élément BRE (Bushnell et al., 2004).

La figure 31 montre le positionnement des sous-unités de TFIIF et de TFIIE avec l'ADN et l'ARNP II.



**Figure 31. Localisation topologique des facteurs TFIIF et TFIIE établie en fonction des analyses de photopontage en gel.**

Les nombreux photopontages de TFIIE et TFIIF en amont de la boîte TATA et en aval du site +1 indiquent que ces facteurs peuvent s'associer au domaine *clamp* de l'ARNP II. De plus, ce positionnement pourrait faciliter le maintien des contacts prolongés avec le *clamp* et faciliter l'enroulement serré de l'ADN autour de l'enzyme. Les facteurs TFIIE et TFIIF sont pontés à des positions distantes de 60 pb (-40 et +26). La présence de deux molécules de chacune des sous-unités de

TFIIF (Flores et al., 1990) et TFIIE (Inostroza et al., 1991) peut expliquer les contacts distants de ces facteurs sur le promoteur d'ADN.

TFIIE exerce un rôle important dans la stimulation de l'ouverture de l'ADN du promoteur *in vitro* (Holstege et al., 1995). Ce facteur possède une activité de liaison à l'ADN qui pourrait être importante pour médier cette ouverture (Ohkuma et Roeder, 1994). Le photopontage de TFIIE56 au nucléotide -10, directement en amont de la bulle de transcription, pourrait donc jouer un rôle dans le mécanisme d'ouverture du promoteur et l'initiation de la transcription. Dans le but de vérifier l'implication de TFIIE56 au cours de l'initiation, des essais de transcription ont été effectués en utilisant un promoteur artificiellement ouvert dans la région de la bulle transcriptionnelle. L'ouverture partielle du promoteur dans différentes régions entre les positions -1 et -11 révèle que le maximum de stimulation se produit dans la région de l'ouverture de la bulle transcriptionnelle qui est la plus en amont et à l'endroit où TFIIE56 pontage l'ADN (Forget et al., 2004). Ces résultats suggèrent que le contact de l'ADN à la position -10 par TFIIE56 pourrait stabiliser l'ouverture du promoteur et stimuler l'initiation de la transcription.

#### ***4.2- Étude des protéines XAB1 et DDX5 dans le réseau d'interactions protéiques de la machinerie transcriptionnelle basale***

##### ***Purification des protéines XAB1 et DDX5 par la méthode TAP***

La méthode TAP permet la purification de complexes multiprotéiques dans des conditions natives (Rigaut et al., 1999). L'intérêt grandissant de cette méthode provient, entre autre, de son efficacité à purifier une protéine d'intérêt en association avec ses partenaires protéiques. Comme il a été mentionné dans la section 1.7.1.2, les partenaires protéiques identifiés avec la méthode TAP peuvent être reliés à la protéine appât par une interaction directe ou indirecte. La

purification des complexes en conditions natives fournit des données utiles pour mieux comprendre la dynamique et les fonctions des protéines cellulaires.

La méthode TAP permet de purifier un ensemble de complexes contenant la protéine appât. Pour identifier les protéines formant ces complexes spécifiques, plusieurs appâts doivent pouvoir être utilisés de concert pour purifier les complexes. Cette méthode a permis de purifier les protéines XAB1 et DDX5 dans des conditions natives, tel que montré à la figure 28.

***La protéine XAB1 est retrouvée en complexe avec l'ARNP II et RPAP1***

La figure 29 montre la migration sur gel des protéines purifiées par la méthode TAP en utilisant l'appât XAB1. Les bandes protéiques ont été excisées du gel et digérées par la trypsine avant d'être analysées par spectrométrie de masse MS/MS. Le bilan des protéines est présenté au tableau IV avec le pointage des résultats de spectrométrie de masse. Les protéines les plus abondantes sur ce gel sont XAB1 (l'appât), Rpb1, Rpb2 et RPAP1. Ces trois dernières protéines ainsi que Rpb5 font partie du complexe RNAPII identifié précédemment par Jeronimo et al (Jeronimo et al, 2004).

L'identification répétée des protéines XAB1, RPAP1, Rpb1, Rpb2 et Rpb5 en utilisant plusieurs appâts (voir tableau III), semble indiquer que ces protéines sont fortement reliées entres-elles physiquement. La protéine RPAP1 a une masse de 153 kDa et n'a pas de fonction connue. Par contre, celle-ci montre une homologie de séquence avec la protéine Ydr527wp de *Saccharomyces cerevisiae*, une protéine essentielle à la viabilité et dont la fonction est également inconnue. Une étude éteignant l'expression de la protéine Ydr527wp chez la levure a montré que son absence induit un changement global dans l'expression des gènes. Ce changement serait similaire à celui causé par l'extinction de de la sous-unité Rpb11 de l'ARNP II (Jeronimo et al, 2004). L'identification d'une fonction précise de ce sous-réseau comprenant au minimum les facteurs XAB1, l'ARNP



II et RPAP1 nécessiterait une investigation plus poussée. La caractérisation de RPAP1 est actuellement en cours dans le laboratoire de Benoit Coulombe.

***Deux nouveaux partenaires d'interaction de XAB1 : C19orf2 (RMP), VDRIP***

La protéine C19orf2, aussi appelée RMP (*Rpb5 mediated protein*), agit de manière antagoniste à l'activité co-activatrice de HBx en compétitionnant pour la liaison à Rpb5 (Dorjsuren et al., 1998). HBx est une protéine du virus de l'hépatite B nécessaire à sa pathogénécité et capable de transactiver certains gènes cellulaires et viraux (Antunovic et al., 1993; Aufiero et Schneider, 1990; Benn et Schneider, 1994; Haviv et al., 1996). C19orf2 module négativement les fonctions de l'ARNP II en se liant à la sous-unité Rpb5, ce qui bloque les sites de liaison de TFIIB et de la protéine HBx (Dorjsuren et al., 1998). Cette même interaction de C19orf2 avec Rpb5 a été identifiée par Gstaiger *et al* (Gstaiger et al., 2003). L'étude démontrait que URI/RMP est un régulateur négatif de la transcription, contrôlé par la kinase TOR. URI/RMP est présent dans un complexe avec STAP1 (*SKP2-associated alpha PFD1*), deux autres membres des protéines PFDs (prefoldins) de classe  $\beta$  et Rpb5. Nos résultats de TAP, en utilisant XAB1 pour appât, montrent aussi cette interaction entre Rpb5 et C19orf2. Puisque C19orf2 module négativement la transcription via des interactions avec Rpb5 (Dorjsuren et al., 1998) (Delgermaa et al., 2004), sa présence dans l'éluat de XAB1-TAP comprenant trois sous-unités de l'ARNP II (Rpb1, Rpb2 et Rpb5) est cohérente.

D'autre part, le domaine CC de C19orf2 lie la protéine DMAP1 (*DNA methyltransferase 1-associated protein*), un co-répresseur transcriptionnel. DMAP1 facilite la localisation nucléaire et l'activité co-répresseur de C19orf2 de manière dose-dépendante (Delgermaa et al., 2004). La protéine DMAP1 est reliée au complexe d'acétylation et de méthylation de l'ADN par ses interactions avec DNMT1 (*DNA methyltransferase 1*) et HDAC2. L'interaction de DMAP1 et de C19orf2 facilite le recrutement de DNMT1 au niveau des régions promotrices, ce qui serait important pour la régulation négative de la transcription, par la modification de la structure de la chromatine (Delgermaa et al., 2004). Ces études

suggèrent donc un deuxième mécanisme de répression transcriptionnelle par C19orf2. Il a été montré que XAB1 peut inhiber MBD2, une composante du complexe MeCP1 qui inclut HDAC1/2, MBD2/3 et RbAp46/48, impliqué dans le *silencing* des gènes (Lembo et al, 2003). L'interaction que nous avons nouvellement identifiée entre XAB1 et C19orf2 pourraient suggérer un mécanisme de dé-répression transcriptionnel semblable dans lequel XAB1, par sa liaison à C19orf2, l'empêcherait d'exercer ses fonctions de répresseur transcriptionnel.

La protéine VDRIP (*vitamin D receptor interacting protein*) a été identifiée dans les éluats de XAB1-TAP. VDRIP est une protéine du médiateur correspondant à MED4 (Sato et al., 2004). Le médiateur est un complexe multiprotéique qui est exprimé de manière ubiquitaire chez les eucaryotes et ce dernier est requis pour l'activation de la transcription par l'ARNP II (voir section 1.3.2.3).

La purification de sous-unités du médiateur en association avec XAB1 n'est pas surprenante étant donnée l'interaction de XAB1 avec certaines sous-unités de l'ARNP II. Toutefois, une seule protéine du médiateur a été retrouvée, mais nos analyses ne sont pas finalisées. La liaison entre le médiateur et XAB1 est probablement indirecte, via l'ARNP II, expliquant le faible nombre de sous-unité du médiateur identifiées. De plus, le médiateur est un grand complexe qui peut être facilement perdu lors des étapes de purification.

Toutefois, les protéines C19orf2 et VDRIP ont obtenu un faible pointage et n'ont pas encore été validé par une deuxième identification en spectrométrie de masse, ni par d'autres méthodes. Leur association à XAB1 demeure donc à être confirmée.

### ***Identification des partenaires protéiques de DDX5-TAP***

Un seul polypeptide interagissant avec DDX5 a été purifié en utilisant la méthode TAP (figure 30). Il s'agit de IPO11, une importine impliquée dans l'importation nucléaire des protéines. Les importines possèdent un domaine de liaison aux nucléoporines des protéines formant les pores nucléaires. L'importation se fait par la liaison du complexe importine-cargo à une nucléoporine. Suite à cette liaison, le complexe est transloqué à travers le pore nucléaire (Wozniak et al., 1998). Une fois du côté nucléoplasmique, la protéine Ran se lie à l'importine, permettant au complexe importine-cargo de se dissocier. L'importine pourra ensuite être retournée vers le cytoplasme avec l'hydrolyse du GTP et le relâchement de la protéine Ran. L'importine 11 est responsable de l'importation de UbcM2 (*E2-type ubiquitin-conjugating enzyme*), une protéine impliquée dans l'ubiquitination et de RPL12 (*ribosomal protein L12*), une protéine ribosomale (Plafker et Macara, 2002; Plafker et Macara, 2000). Dans des essais *in vivo* et *in vitro*, RPL12 et UbcM2 entrent en compétition pour lier l'importine 11.

L'interaction de IPO11 avec DDX5 suggère que DDX5 peut être un nouveau substrat de l'importine 11. Pour réaliser leurs fonctions nucléaires, certaines protéines, traduites dans le cytoplasme, doivent être rapatriées au noyau. Les nombreux rôles de DDX5 dans la régulation de l'épissage et sa localisation nucléaire (PSORT – *prediction of protein sorting signals and localization sites* - 52% nucléaire) corrélerent avec la liaison de DDX5 à une importine pour permettre son transport au noyau.

Les études de spectrométrie de masse n'ont pas révélé la présence de TFIIS dans les éluats purifiés de DDX5-TAP. Pour confirmer son absence, des analyses d'immunobuvardage ont été réalisées. La figure 31 montre que l'immunobuvardage n'a pas révélé la présence de TFIIS dans l'éluat de DDX5-TAP. Toutefois, la quantité de DDX5 dans les éluats de TFIIS est relativement faible, suggérant une interaction indirecte ou faible entre ces deux protéines. De plus, l'ajout de l'étiquette TAP-Tag, du côté C-terminal de la protéine, peut

interférer avec l'interaction de TFIIIS ou, encore, modifier le repliement de DDX5. Conséquemment, il peut-être utile d'avoir recours à une étiquette située du côté N-terminale pour identifier les partenaires agissant du côté C-terminal de la protéine. Par contre, le très faible nombre de protéines identifiées en complexe avec DDX5 ne prête pas beaucoup d'intérêt à une analyse plus poussée.

## 5- CONCLUSION ET PERSPECTIVES

Les travaux de photopontage réalisés dans le laboratoire de Benoit Coulombe ont permis de déterminer la topologie PIC (Forget et al., 1997; Robert et al., 1996; Douziech et al., 2000). Toutefois, la méthode de photopontage classique permet la formation de sous-complexes pouvant interférer avec les résultats. Dans le but d'étudier un complexe TBP-TFIIB-TFIIE-TFIIF-ARNP II-promoteur purifié, nous avons mis en place une méthode qui permet d'isoler les complexes en conditions natives avant de procéder au pontage. Pour la première partie de mon projet de maîtrise, des analyses de photopontage en gel ont été réalisées et ont permis de révéler avec précision la du PIC et la trajectoire de l'ADN dans le complexe.

Les résultats de photopontage en gel avec l'ARNP II et les facteurs TBP, TFIIB, TFIIE et TFIIF montrent relativement peu de différences avec les études précédentes utilisant le photopontage classique. Toutefois, les facteurs de transcription TFIIF et TFIIE montrent des pontages beaucoup moins nombreux le long du promoteur et se localisent à des régions beaucoup plus définies. Le phénomène d'enroulement de l'ADN autour de l'ARNP II peut être mis en évidence. Le modèle élaboré permet de montrer un enroulement de l'ADN vers la gauche autour du *clamp* de l'ARNP II. Cet enroulement permettrait d'ouvrir partiellement l'ADN pour donner accès aux hélicases de TFIIF. Finalement, le pontage de TFIIE56, identifiée directement en amont de la bulle transcriptionnelle, a permis de suggérer un rôle de ce facteur dans l'ouverture du promoteur et l'initiation de la transcription. Des essais de transcription, utilisant un promoteur artificiellement ouvert, ont révélé que le contact de l'ADN à la position -10 par TFIIE56, stabilise l'ouverture du promoteur, ce qui pourrait stimuler la transcription. Toutes ces données ont permis de mieux comprendre le phénomène d'initiation de la transcription pour établir une structure détaillée des interactions nucléoprotéiques du PIC.

Les étapes futures, envisagées pour la continuité de ce projet, impliquent des analyses de photopontage avec le facteur TFIIH, en présence et en absence d'ATP, dans le but d'obtenir un modèle de la localisation du promoteur d'ADN dans le PIC ouvert. Ces données permettront de réaliser un schéma dynamique du mécanisme d'ouverture de la bulle transcriptionnelle.

Pour la seconde partie de mon projet de maîtrise, la méthode TAP a été utilisée afin de déterminer les partenaires protéiques de XAB1 et DDX5. Ces polypeptides ont été identifiés antérieurement dans le laboratoire en complexe avec des protéines de la machinerie transcriptionnelle basale. La méthode TAP est très puissante pour étudier le réseau d'interaction des protéines à grande échelle. Cette méthode nous a permis de confirmer l'interaction de XAB1 avec l'ARNP II et RPAP1, ainsi que d'identifier deux nouveaux partenaires d'interactions de XAB1; C19orf2 et VDRIP.

L'investigation de XAB1 se poursuit dans le laboratoire de Benoit Coulombe, en utilisant, entre autre, une étiquette d'affinité du côté N-terminal pour tenter d'identifier d'autres partenaires d'interaction. De plus, ces études additionnelles permettront de valider la présence des protéines C19orf2 et VDRIP dans l'éluat XAB1-TAP. La caractérisation de RPAP1 est en cours pour tenter d'élucider les rôles de cette protéine dans la transcription. D'autre part, puisque aucune protéine reliée à la régulation transcriptionnelle n'a été identifiée dans le complexe DDX5-TAP, l'investigation sur cette protéine ne sera pas poursuivie.

**6-LISTE DE RÉFÉRENCES**

1. Acker,J., de Graaff,M., Cheynel,I., Khazak,V., Kedinge,r,C., and Vigneron,M. (1997). Interactions between the human RNA polymerase II subunits. *J. Biol. Chem.* 272, 16815-16821.
2. Albright,S.R. and Tjian,R. (2000). TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* 242, 1-13.
3. Andel,F., III, Ladurner,A.G., Inouye,C., Tjian,R., and Nogales,E. (1999). Three-dimensional structure of the human TFIID-IIA-IIB complex. *Science* 286, 2153-2156.
4. Antunovic,J., Lemieux,N., and Cromlish,J.A. (1993). The 17 kDa HBx protein encoded by hepatitis B virus interacts with the activation domains of Oct-1, and functions as a coactivator in the activation and repression of a human U6 promoter. *Cell Mol Biol Res.* 39, 463-482.
5. Aoyagi,S., Narlikar,G., Zheng,C., Sif,S., Kingston,R.E., and Hayes,J.J. (2002). Nucleosome remodeling by the human SWI/SNF complex requires transient global disruption of histone-DNA interactions. *Mol. Cell Biol.* 22, 3653-3662.
6. Archambault,J., Chambers,R.S., Kobor,M.S., Ho,Y., Cartier,M., Bolotin,D., Andrews,B., Kane,C.M., and Greenblatt,J. (1997). 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.
7. Archambault,J. and Friesen,J.D. (1993). Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiol. Rev.* 57, 703-724.
8. Archambault,J., Pan,G., Dahmus,G.K., Cartier,M., Marshall,N., Zhang,S., Dahmus,M.E., and Greenblatt,J. (1998). 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.
9. Armache,K.J., Kettenberger,H., and Cramer,P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A* 100, 6964-6968.
10. Asturias,F.J., Jiang,Y.W., Myers,L.C., Gustafsson,C.M., and Kornberg,R.D. (1999). Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* 283, 985-987.
11. Auble,D.T., Hansen,K.E., Mueller,C.G., Lane,W.S., Thorner,J., and Hahn,S. (1994). Mot1, a global repressor of RNA polymerase II transcription, inhibits

- TBP binding to DNA by an ATP-dependent mechanism. *Genes Dev.* **8**, 1920-1934.
12. **Aufiero,B. and Schneider,R.J. (1990).** The hepatitis B virus X-gene product trans-activates both RNA polymerase II and III promoters. *EMBO J.* **9**, 497-504.
  13. **Bader,G.D., Donaldson,I., Wolting,C., Ouellette,B.F., Pawson,T., and Hogue,C.W. (2001).** BIND--The Biomolecular Interaction Network Database. *Nucleic Acids Res.* **29**, 242-245.
  14. **Bagby,S., Kim,S., Maldonado,E., Tong,K.I., Reinberg,D., and Ikura,M. (1995).** Solution structure of the C-terminal core domain of human TFIIB: similarity to cyclin A and interaction with TATA-binding protein. *Cell* **82**, 857-867.
  15. **Bartholomew,B., Kassavetis,G.A., Braun,B.R., and Geiduschek,E.P. (1990).** The subunit structure of *Saccharomyces cerevisiae* transcription factor III<sub>C</sub> probed with a novel photocrosslinking reagent. *EMBO J.* **9**, 2197-2205.
  16. **Bell,S.D., Brinkman,A.B., van der,O.J., and Jackson,S.P. (2001).** The archaeal TFIIE $\alpha$  homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep.* **2**, 133-138.
  17. **Bengal,E., Flores,O., Krauskopf,A., Reinberg,D., and Aloni,Y. (1991).** Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. *Mol. Cell Biol.* **11**, 1195-1206.
  18. **Benn,J. and Schneider,R.J. (1994).** Hepatitis B virus HBx protein activates Ras-GTP complex formation and establishes a Ras, Raf, MAP kinase signaling cascade. *Proc. Natl. Acad. Sci. U. S. A* **91**, 10350-10354.
  19. **Bensaude,O., Bonnet,F., Casse,C., Dubois,M.F., Nguyen,V.T., and Palancade,B. (1999).** Regulated phosphorylation of the RNA polymerase II C-terminal domain (CTD). *Biochem. Cell Biol* **77**, 249-255.
  20. **Blackwood,E.M. and Kadonaga,J.T. (1998).** Going the distance: a current view of enhancer action. *Science* **281**, 61-63.
  21. **Bradsher,J., Coin,F., and Egly,J.M. (2000).** Distinct roles for the helicases of TFIIF in transcript initiation and promoter escape. *J. Biol. Chem.* **275**, 2532-2538.
  22. **Bulger,M. and Groudine,M. (1999).** Looping versus linking: toward a model for long-distance gene activation. *Genes & Development* **13**, 2465-2477.



23. Buratowski,S., Hahn,S., Sharp,P.A., and Guarente,L. (1988). Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* 334, 37-42.
24. Burke,T.W. and Kadonaga,J.T. (1996). *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* 10, 711-724.
25. Bushnell,D.A., Westover,K.D., Davis,R.E., and Kornberg,R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* 303, 983-988.
26. Butler,J.E. and Kadonaga,J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* 16, 2583-2592.
27. Cairns,B.R., Lorch,Y., Li,Y., Zhang,M., Lacomis,L., Erdjument-Bromage,H., Tempst,P., Du,J., Laurent,B., and Kornberg,R.D. (1996). RSC, an essential, abundant chromatin-remodeling complex. *Cell* 87, 1249-1260.
28. Cavallini,B., Faus,I., Matthes,H., Chipoulet,J.M., Winsor,B., Egly,J.M., and Chambon,P. (1989). Cloning of the gene encoding the yeast protein BTF1Y, which can substitute for the human TATA box-binding factor. *Proc. Natl. Acad. Sci. U. S. A* 86, 9803-9807.
29. Chalkley,G.E. and Verrijzer,C.P. (1999). DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J.* 18, 4835-4845.
30. Chambers,R.S., Wang,B.Q., Burton,Z.F., and Dahmus,M.E. (1995). 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.
31. Chan,H.M. and La Thangue,N.B. (2001). p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J. Cell Sci.* 114, 2363-2373.
32. Chen,B.S., Mandal,S.S., and Hampsey,M. (2004). High-resolution protein-DNA contacts for the yeast RNA polymerase II general transcription machinery. *Biochemistry* 43, 12741-12749.
33. Coleman,R.A., Taggart,A.K., Burma,S., Chicca,J.J., and Pugh,B.F. (1999). TFIIA regulates TBP and TFIID dimers. *Mol. Cell* 4, 451-457.
34. Conaway,R.C. and Conaway,J.W. (1993). General initiation factors for RNA polymerase II. *Annu. Rev. Biochem.* 62, 161-190.

35. Cooper,G.M. (1999). *La cellule. Une approche moléculaire.*, I.Sinauer Associates, ed., p. -706.
36. Corden,J.L. (1990). Tails of RNA polymerase II. *Trends Biochem. Sci.* *15*, 383-387.
37. Coulombe,B. and Burton,Z.F. (1999). DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. *Microbiol. Mol. Biol. Rev.* *63*, 457-478.
38. Coulombe,B., Jeronimo,C., Langelier,M.F., Cojocar,M., and Bergeron,D. (2004). Interaction networks of the molecular machines that decode, replicate and maintain the integrity of the human genome. *Mol. Cell Proteomics.* *3*, 851-856.
39. Coulombe,B., Li,J., and Greenblatt,J. (1994). Topological localization of the human transcription factors IIA, IIB, TATA box-binding protein, and RNA polymerase II-associated protein 30 on a class II promoter. *J. Biol. Chem.* *269*, 19962-19967.
40. Cox,J.M., Hayward,M.M., Sanchez,J.F., Gegnas,L.D., van der,Z.S., Dennis,J.H., Sigler,P.B., and Schepartz,A. (1997). Bidirectional binding of the TATA box binding protein to the TATA box. *Proc. Natl. Acad. Sci. U. S. A* *94*, 13475-13480.
41. Craig,M.L., Tsodikov,O.V., McQuade,K.L., Schlax,P.E., Jr., Capp,M.W., Saecker,R.M., and Record,M.T., Jr. (1998). DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. *J. Mol. Biol.* *283*, 741-756.
42. Cramer,P. (2004). RNA polymerase II structure: from core to functional complexes. *Curr. Opin. Genet. Dev.* *14*, 218-226.
43. Cramer,P., Bushnell,D.A., Fu,J., Gnatt,A.L., Maier-Davis,B., Thompson,N.E., Burgess,R.R., Edwards,A.M., David,P.R., and Kornberg,R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* *288*, 640-649.
44. Cramer,P., Bushnell,D.A., and Kornberg,R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* *292*, 1863-1876.
45. Crowley,T.E., Hoey,T., Liu,J.K., Jan,Y.N., Jan,L.Y., and Tjian,R. (1993). A new factor related to TATA-binding protein has highly restricted expression patterns in *Drosophila*. *Nature* *361*, 557-561.

46. Dantoni, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* 389, 399-402.
47. Davis, J.A., Takagi, Y., Kornberg, R.D., and Asturias, F.A. (2002). Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol. Cell* 10, 409-415.
48. de Laat, W.L., Jaspers, N.G., and Hoeijmakers, J.H. (1999). Molecular mechanism of nucleotide excision repair. *Genes Dev.* 13, 768-785.
49. DeJong, J., Bernstein, R., and Roeder, R.G. (1995). Human general transcription factor TFIIA: characterization of a cDNA encoding the small subunit and requirement for basal and activated transcription. *Proc. Natl. Acad. Sci. U. S. A* 92, 3313-3317.
50. DeJong, J. and Roeder, R.G. (1993). A single cDNA, hTFIIA/alpha, encodes both the p35 and p19 subunits of human TFIIA. *Genes Dev.* 7, 2220-2234.
51. Delgermaa, L., Hayashi, N., Dorjsuren, D., Nomura, T., Thuy, I.T., and Murakami, S. (2004). Subcellular localization of RPB5-mediating protein and its putative functional partner. *Mol Cell Biol* 24, 8556-8566.
52. Dion, V. and Coulombe, B. (2003). Interactions of a DNA-bound transcriptional activator with the TBP-TFIIA-TFIIB-promoter quaternary complex. *J. Biol Chem.* 278, 11495-11501.
53. Dorjsuren, D., Lin, Y., Wei, W., Yamashita, T., Nomura, T., Hayashi, N., and Murakami, S. (1998). RMP, a novel RNA polymerase II subunit 5-interacting protein, counteracts transactivation by hepatitis B virus X protein. *Mol. Cell Biol.* 18, 7546-7555.
54. Douziech, M., Coin, F., Chipoulet, J.M., Arai, Y., Ohkuma, Y., Egly, J.M., and Coulombe, B. (2000). Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking. *Mol. Cell Biol.* 20, 8168-8177.
55. Douziech, M., Forget, D., Greenblatt, J., and Coulombe, B. (1999). Topological localization of the carboxyl-terminal domain of RNA polymerase II in the initiation complex. *J. Biol. Chem.* 274, 19868-19873.
56. Dubrovskaya, V., Lavigne, A.C., Davidson, I., Acker, J., Staub, A., and Tora, L. (1996). Distinct domains of hTAFII100 are required for functional interaction with transcription factor TFIIIF beta (RAP30) and incorporation into the TFIID complex. *EMBO J.* 15, 3702-3712.

57. Edwards,A.M., Darst,S.A., Feaver,W.J., Thompson,N.E., Burgess,R.R., and Kornberg,R.D. (1990). Purification and lipid-layer crystallization of yeast RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A* 87, 2122-2126.
58. Egly,J.M. (2001). The 14th Datta Lecture. TFIIH: from transcription to clinic. *FEBS Lett.* 498, 124-128.
59. Elsby,L.M. and Roberts,S.G. (2004). The role of TFIIB conformation in transcriptional regulation. *Biochem. Soc. Trans.* 32, 1098-1099.
60. Fairley,J.A., Evans,R., Hawkes,N.A., and Roberts,S.G. (2002). Core promoter-dependent TFIIB conformation and a role for TFIIB conformation in transcription start site selection. *Mol. Cell Biol.* 22, 6697-6705.
61. Fang,S.M. and Burton,Z.F. (1996). RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding. *J. Biol. Chem.* 271, 11703-11709.
62. Fields,S. and Song,O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* 340, 245-246.
63. Flores,O., Ha,I., and Reinberg,D. (1990). Factors involved in specific transcription by mammalian RNA polymerase II. Purification and subunit composition of transcription factor IIF. *J. Biol. Chem.* 265, 5629-5634.
64. Flores,O., Lu,H., Killeen,M., Greenblatt,J., Burton,Z.F., and Reinberg,D. (1991). The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proc. Natl. Acad. Sci. U. S. A* 88, 9999-10003.
65. Flores,O., Maldonado,E., and Reinberg,D. (1989). Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J. Biol. Chem.* 264, 8913-8921.
66. Forget,D. and Coulombe,B. (2003). Site-specific protein-DNA photo-cross-linking of purified complexes: Topology of the RNA polymerase II transcription initiation complex. In *RNA Polymerase and Associated Factors, Part C*, S.L.Adhya and S.Garges, eds. (San Diego, Calif.: Elsevier Academic Press), pp. 701-712.
67. Forget,D., Langelier,M.F., Therien,C., Trinh,V., and Coulombe,B. (2004). Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. *Mol Cell Biol* 24, 1122-1131.
68. Forget,D., Robert,F., Grondin,G., Burton,Z.F., Greenblatt,J., and Coulombe,B. (1997). RAP74 induces promoter contacts by RNA polymerase

- II upstream and downstream of a DNA bend centered on the TATA box. Proc. Natl. Acad. Sci. U. S. A 94, 7150-7155.**
69. **Gaiser,F., Tan,S., and Richmond,T.J. (2000). Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. J. Mol. Biol. 302, 1119-1127.**
  70. **Galas,D.J. and Schmitz,A. (1978). DNase footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. 5, 3157-3170.**
  71. **Garner,M.M. and Revzin,A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids Res. 9, 3047-3060.**
  72. **Gavin,A.C., Bosche,M., Krause,R., Grandi,P., Marzioch,M., Bauer,A., Schultz,J., Rick,J.M., Michon,A.M., Cruciat,C.M., Remor,M., Hofert,C., Schelder,M., Brajenovic,M., Ruffner,H., Merino,A., Klein,K., Hudak,M., Dickson,D., Rudi,T., Gnau,V., Bauch,A., Bastuck,S., Huhse,B., Leutwein,C., Heurtier,M.A., Copley,R.R., Edelman,A., Querfurth,E., Rybin,V., Drewes,G., Raida,M., Bouwmeester,T., Bork,P., Seraphin,B., Kuster,B., Neubauer,G., and Superti-Furga,G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141-147.**
  73. **Ghazy,M.A., Brodie,S.A., Ammerman,M.L., Ziegler,L.M., and Ponticelli,A.S. (2004). Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. Mol Cell Biol 24, 10975-10985.**
  74. **Gnatt,A. (2002). Elongation by RNA polymerase II: structure-function relationship. Biochim. Biophys. Acta 1577, 175.**
  75. **Gnatt,A.L., Cramer,P., Fu,J., Bushnell,D.A., and Kornberg,R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. Science 292, 1876-1882.**
  76. **Greenblatt,J. (1997). RNA polymerase II holoenzyme and transcriptional regulation. Curr. Opin. Cell Biol. 9, 310-319.**
  77. **Gstaiger,M., Luke,B., Hess,D., Oakeley,E.J., Wirbelauer,C., Blondel,M., Vigneron,M., Peter,M., and Krek,W. (2003). Control of nutrient-sensitive transcription programs by the unconventional prefoldin URI. Science 302, 1208-1212.**
  78. **Guarente,L. and Hoar,E. (1984). 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.
79. Guil,S., Gattoni,R., Carrascal,M., Abian,J., Stevenin,J., and Bach-Elias,M. (2003). Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol Cell Biol* *23*, 2927-2941.
  80. Ha,I., Lane,W.S., and Reinberg,D. (1991). Cloning of a human gene encoding the general transcription initiation factor IIB. *Nature* *352*, 689-695.
  81. Ha,I., Roberts,S., Maldonado,E., Sun,X., Kim,L.U., Green,M., and Reinberg,D. (1993). Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II. *Genes Dev.* *7*, 1021-1032.
  82. Hahn,S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol Biol* *11*, 394-403.
  83. Hampsey,M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* *62*, 465-503.
  84. Haviv,I., Vaizel,D., and Shaul,Y. (1996). pX, the HBV-encoded coactivator, interacts with components of the transcription machinery and stimulates transcription in a TAF-independent manner. *EMBO J.* *15*, 3413-3420.
  85. He,H. and Lehming,N. (2003). Global effects of histone modifications. *Brief. Funct. Genomic. Proteomic.* *2*, 234-243.
  86. Ho,Y., Gruhler,A., Heilbut,A., Bader,G.D., Moore,L., Adams,S.L., Millar,A., Taylor,P., Bennett,K., Boutilier,K., Yang,L., Wolting,C., Donaldson,I., Schandorff,S., Shewnarane,J., Vo,M., Taggart,J., Goudreault,M., Muskat,B., Alfarano,C., Dewar,D., Lin,Z., Michalickova,K., Willems,A.R., Sassi,H., Nielsen,P.A., Rasmussen,K.J., Andersen,J.R., Johansen,L.E., Hansen,L.H., Jaspersen,H., Podtelejnikov,A., Nielsen,E., Crawford,J., Poulsen,V., Sorensen,B.D., Matthiesen,J., Hendrickson,R.C., Gleeson,F., Pawson,T., Moran,M.F., Durocher,D., Mann,M., Hogue,C.W., Figeys,D., and Tyers,M. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* *415*, 180-183.
  87. Hodo,H.G., III and Blatti,S.P. (1977). Purification using polyethylenimine precipitation and low molecular weight subunit analyses of calf thymus and wheat germ DNA-dependent RNA polymerase II. *Biochemistry* *16*, 2334-2343.
  88. Hoiby,T., Mitsiou,D.J., Zhou,H.Q., Erdjument-Bromage,H., Tempst,P., and Stunnenberg,H.G. (2004). Cleavage and proteasome-mediated degradation of the basal transcription factor TFIIA. *EMBO J.* *23*, 3083-3091.

89. Holstege,F.C., Fiedler,U., and Timmers,H.T. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.* *16*, 7468-7480.
90. Holstege,F.C., Tantin,D., Carey,M., van der Vliet,P.C., and Timmers,H.T. (1995). The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO J.* *14*, 810-819.
91. Hu,S.L. and Manley,J.L. (1981). DNA sequence required for initiation of transcription in vitro from the major late promoter of adenovirus 2. *Proc. Natl. Acad. Sci. U. S. A* *78*, 820-824.
92. Inostroza,J., Flores,O., and Reinberg,D. (1991). Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of general transcription factor IIE. *J. Biol. Chem.* *266*, 9304-9308.
93. Ito,T., Chiba,T., Ozawa,R., Yoshida,M., Hattori,M., and Sakaki,Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A* *98*, 4569-4574.
94. 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., and Coulombe,B. (2004). RPAP1, a novel human RNA polymerase II-associated protein affinity purified with recombinant wild-type and mutated polymerase subunits. *Mol Cell Biol* *24*, 7043-7058.
95. Jupp,R., Flores,O., Nelson,J.A., and Ghazal,P. (1993). The Dna-Binding Subunit of Human Transcription Factor-Iid Can Interact with the Tata Box As A Multimer. *J. Biol. Chem.* *268*, 16105-16108.
96. Kamenski,T., Heilmeier,S., Meinhart,A., and Cramer,P. (2004). Structure and mechanism of RNA polymerase II CTD phosphatases. *Mol Cell* *15*, 399-407.
97. Kays,A.R. and Schepartz,A. (2002). Gal4-VP16 and Gal4-AH increase the orientational and axial specificity of TATA box recognition by TATA box binding protein. *Biochemistry* *41*, 3147-3155.
98. Kayukawa,K., Makino,Y., Yogosawa,S., and Tamura,T. (1999). A serine residue in the N-terminal acidic region of rat RPB6, one of the common subunits of RNA polymerases, is exclusively phosphorylated by casein kinase II in vitro. *Gene* *234*, 139-147.
99. Kettenberger,H., Armache,K.J., and Cramer,P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* *114*, 347-357.

100. Killeen,M., Coulombe,B., and Greenblatt,J. (1992). Recombinant TBP, transcription factor IIB, and RAP30 are sufficient for promoter recognition by mammalian RNA polymerase II. *J. Biol. Chem.* *267*, 9463-9466.
101. Killeen,M.T. and Greenblatt,J.F. (1992). The general transcription factor RAP30 binds to RNA polymerase II and prevents it from binding nonspecifically to DNA. *Mol. Cell Biol.* *12*, 30-37.
102. Kim,T.K., Lagrange,T., Wang,Y.H., Griffith,J.D., Reinberg,D., and Ebright,R.H. (1997). Trajectory of DNA in the RNA polymerase II transcription preinitiation complex. *Proc. Natl. Acad. Sci. U. S. A* *94*, 12268-12273.
103. Kim,Y., Geiger,J.H., Hahn,S., and Sigler,P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature* *365*, 512-520.
104. Klug,A. (2001). Structural biology. A marvellous machine for making messages. *Science* *292*, 1844-1846.
105. Kornberg,R.D. (1974). Chromatin structure: a repeating unit of histones and DNA  
111. *Science* *184*, 868-871.
106. Kou,H., Irvin,J.D., Huisinga,K.L., Mitra,M., and Pugh,B.F. (2003). Structural and Functional Analysis of Mutations along the Crystallographic Dimer Interface of the Yeast TATA Binding Protein. *Mol. Cell Biol.* *23*, 3186-3201.
107. Kuo,M.H. and Allis,C.D. (1999). In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* *19*, 425-433.
108. Kutach,A.K. and Kadonaga,J.T. (2000). 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.
109. Lagrange,T., Kapanidis,A.N., Tang,H., Reinberg,D., and Ebright,R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* *12*, 34-44.
110. Langelier,M.F., Forget,D., Rojas,A., Porlier,Y., Burton,Z.F., and Coulombe,B. (2001). Structural and functional interactions of transcription factor (TF) IIA with TFIIE and TFIIIF in transcription initiation by RNA polymerase II. *J. Biol. Chem.* *276*, 38652-38657.



111. Larochelle,S., Pandur,J., Fisher,R.P., Salz,H.K., and Suter,B. (1998). Cdk7 is essential for mitosis and for in vivo Cdk-activating kinase activity. *Genes Dev.* 12, 370-381.
112. Laybourn,P.J. and Dahmus,M.E. (1990). Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. *J. Biol. Chem.* 265, 13165-13173.
113. Lee,T.I., Rinaldi,N.J., Robert,F., Odom,D.T., Bar-Joseph,Z., Gerber,G.K., Hannett,N.M., Harbison,C.T., Thompson,C.M., Simon,I., Zeitlinger,J., Jennings,E.G., Murray,H.L., Gordon,D.B., Ren,B., Wyrick,J.J., Tagne,J.B., Volkert,T.L., Fraenkel,E., Gifford,D.K., and Young,R.A. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799-804.
114. Lee,T.I. and Young,R.A. (2000). Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* 34, 77-137.
115. Legube,G. and Trouche,D. (2003). Regulating histone acetyltransferases and deacetylases. *EMBO Rep.* 4, 944-947.
116. Lei,L., Ren,D., Finkelstein,A., and Burton,Z.F. (1998). Functions of the N- and C-terminal domains of human RAP74 in transcriptional initiation, elongation, and recycling of RNA polymerase II. *Mol. Cell Biol.* 18, 2130-2142.
117. Lembo,F., Pero,R., Angrisano,T., Vitiello,C., Iuliano,R., Bruni,C.B., and Chiariotti,L. (2003). MBDin, a novel MBD2-interacting protein, relieves MBD2 repression potential and reactivates transcription from methylated promoters. *Mol Cell Biol* 23, 1656-1665.
118. Liu,Z.R. (2002). p68 RNA Helicase Is an Essential Human Splicing Factor That Acts at the U1 snRNA-5' Splice Site Duplex. *Mol. Cell Biol.* 22, 5443-5450.
119. Lorch,Y., LaPointe,J.W., and Kornberg,R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* 49, 203-210.
120. Lu,H., Flores,O., Weinmann,R., and Reinberg,D. (1991). The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc. Natl. Acad. Sci. U. S. A* 88, 10004-10008.
121. Ma,D., Watanabe,H., Mermelstein,F., Admon,A., Oguri,K., Sun,X., Wada,T., Imai,T., Shiroya,T., Reinberg,D., and . (1993). Isolation of a cDNA encoding the largest subunit of TFIIA reveals functions important for activated transcription. *Genes Dev.* 7, 2246-2257.

122. **Martinez,E. (2002). Multi-protein complexes in eukaryotic gene transcription. *Plant Mol Biol* 50, 925-947.**
123. **Maxon,M.E., Goodrich,J.A., and Tjian,R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIF: a model for promoter clearance. *Genes Dev.* 8, 515-524.**
124. **Maxon,M.E. and Tjian,R. (1994). Transcriptional activity of transcription factor IIE is dependent on zinc binding. *Proc. Natl. Acad. Sci. U. S. A* 91, 9529-9533.**
125. **Meinhart,A., Blobel,J., and Cramer,P. (2003). An Extended Winged Helix Domain in General Transcription Factor E/IIE{alpha}. *J. Biol. Chem.* 278, 48267-48274.**
126. **Mohrmann,L., Langenberg,K., Krijgsveld,J., Kal,A.J., Heck,A.J., and Verrijzer,C.P. (2004). Differential targeting of two distinct SWI/SNF-related *Drosophila* chromatin-remodeling complexes. *Mol. Cell Biol.* 24, 3077-3088.**
127. **Myer,V.E. and Young,R.A. (1998). RNA polymerase II holoenzymes and subcomplexes. *J. Biol. Chem.* 273, 27757-27760.**
128. **Myers,L.C., Gustafsson,C.M., Bushnell,D.A., Lui,M., Erdjument-Bromage,H., Tempst,P., and Kornberg,R.D. (1998). The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev.* 12, 45-54.**
129. **Nakadai,T., Shimada,M., Shima,D., Handa,H., and Tamura,T.A. (2004). Specific interaction with transcription factor IIA and localization of the mammalian TATA-binding protein-like protein (TLP/TRF2/TLF). *J. Biol Chem.* 279, 7447-7455.**
130. **Nguyen,B.D., Chen,H.T., Kobor,M.S., Greenblatt,J., Legault,P., and Omichinski,J.G. (2003). Solution structure of the carboxyl-terminal domain of RAP74 and NMR characterization of the FCP1-binding sites of RAP74 and human TFIIB. *Biochemistry* 42, 1460-1469.**
131. **Nikolov,D.B. and Burley,S.K. (1994). 2.1 A resolution refined structure of a TATA box-binding protein (TBP). *Nat. Struct. Biol.* 1, 621-637.**
132. **Nikolov,D.B., Chen,H., Halay,E.D., Hoffman,A., Roeder,R.G., and Burley,S.K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci. U. S. A* 93, 4862-4867.**
133. **No,D., Yao,T.P., and Evans,R.M. (1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U. S. A* 93, 3346-3351.**

134. Nogales,E. (2000). Recent structural insights into transcription preinitiation complexes. *J. Cell Sci.* *113 Pt 24*, 4391-4397.
135. Nonet,M., Sweetser,D., and Young,R.A. (1987). Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell* *50*, 909-915.
136. O'Shea-Greenfield,A. and Smale,S.T. (1992). Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J. Biol. Chem.* *267*, 1391-1402.
137. Oelgeschlager,T., Chiang,C.M., and Roeder,R.G. (1996). Topology and reorganization of a human TFIID-promoter complex. *Nature* *382*, 735-738.
138. Ogilvie,V.C., Wilson,B.J., Nicol,S.M., Morrice,N.A., Saunders,L.R., Barber,G.N., and Fuller-Pace,F.V. (2003). The highly related DEAD box RNA helicases p68 and p72 exist as heterodimers in cells. *Nucleic Acids Res.* *31*, 1470-1480.
139. Ohbayashi,T., Makino,Y., and Tamura,T.A. (1999). Identification of a mouse TBP-like protein (TLP) distantly related to the drosophila TBP-related factor. *Nucleic Acids Res.* *27*, 750-755.
140. Ohkuma,Y. (1997). Multiple functions of general transcription factors TFIIE and TFIIH in transcription: possible points of regulation by trans-acting factors. *J. Biochem. (Tokyo)* *122*, 481-489.
141. Ohkuma,Y., Hashimoto,S., Wang,C.K., Horikoshi,M., and Roeder,R.G. (1995). Analysis of the role of TFIIE in basal transcription and TFIIH-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-alpha. *Mol. Cell Biol.* *15*, 4856-4866.
142. Ohkuma,Y. and Roeder,R.G. (1994). Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature* *368*, 160-163.
143. Ohkuma,Y., Sumimoto,H., Hoffmann,A., Shimasaki,S., Horikoshi,M., and Roeder,R.G. (1991). Structural motifs and potential sigma homologies in the large subunit of human general transcription factor TFIIE. *Nature* *354*, 398-401.
144. Orphanides,G. and Reinberg,D. (2002). A unified theory of gene expression. *Cell* *108*, 439-451.
145. Ozer,J., Lezina,L.E., Ewing,J., Audi,S., and Lieberman,P.M. (1998a). Association of transcription factor IIA with TATA binding protein is required for transcriptional activation of a subset of promoters and cell cycle progression in *Saccharomyces cerevisiae*. *Mol Cell Biol* *18*, 2559-2570.

146. Ozer,J., Mitsouras,K., Zerby,D., Carey,M., and Lieberman,P.M. (1998b). Transcription factor IIA derepresses TATA-binding protein (TBP)-associated factor inhibition of TBP-DNA binding. *J. Biol. Chem.* *273*, 14293-14300.
147. Ozer,J., Moore,P.A., Bolden,A.H., Lee,A., Rosen,C.A., and Lieberman,P.M. (1994). Molecular cloning of the small (gamma) subunit of human TFIIA reveals functions critical for activated transcription. *Genes Dev.* *8*, 2324-2335.
148. Palancade,B. and Bensaude,O. (2003). Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur. J. Biochem.* *270*, 3859-3870.
149. Patikoglou,G.A., Kim,J.L., Sun,L., Yang,S.H., Kodadek,T., and Burley,S.K. (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* *13*, 3217-3230.
150. Peterson,C.L. (2002). Chromatin remodeling enzymes: taming the machines: Third in review series on chromatin dynamics. *EMBO Rep.* *3*, 319-322.
151. Plafker,S.M. and Macara,I.G. (2000). Importin-11, a nuclear import receptor for the ubiquitin-conjugating enzyme, UbcM2. *EMBO J.* *19*, 5502-5513.
152. Plafker,S.M. and Macara,I.G. (2002). Ribosomal protein L12 uses a distinct nuclear import pathway mediated by importin 11. *Mol Cell Biol* *22*, 1266-1275.
153. Polyakov,A., Severinova,E., and Darst,S.A. (1995). Three-dimensional structure of *E. coli* core RNA polymerase: promoter binding and elongation conformations of the enzyme. *Cell* *83*, 365-373.
154. Puig,O., Caspary,F., Rigaut,G., Rutz,B., Bouveret,E., Bragado-Nilsson,E., Wilm,M., and Seraphin,B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* *24*, 218-229.
155. Ranish,J.A., Lane,W.S., and Hahn,S. (1992). Isolation of two genes that encode subunits of the yeast transcription factor IIA. *Science* *255*, 1127-1129.
156. Ranish,J.A., Yudkovsky,N., and Hahn,S. (1999). Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes Dev.* *13*, 49-63.
157. Ren,D., Lei,L., and Burton,Z.F. (1999). A region within the RAP74 subunit of human transcription factor IIF is critical for initiation but dispensable for complex assembly. *Mol. Cell Biol.* *19*, 7377-7387.

158. Rigaut,G., Shevchenko,A., Rutz,B., Wilm,M., Mann,M., and Seraphin,B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* *17*, 1030-1032.
159. Rivetti,C., Guthold,M., and Bustamante,C. (1999). Wrapping of DNA around the E.coli RNA polymerase open promoter complex. *EMBO J.* *18*, 4464-4475.
160. Robert,F., Douziech,M., Forget,D., Egly,J.M., Greenblatt,J., Burton,Z.F., and Coulombe,B. (1998). Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. *Mol. Cell* *2*, 341-351.
161. Robert,F., Forget,D., Li,J., Greenblatt,J., and Coulombe,B. (1996). Localization of subunits of transcription factors IIE and IIF immediately upstream of the transcriptional initiation site of the adenovirus major late promoter. *J. Biol Chem.* *271*, 8517-8520.
162. Rossow,K.L. and Janknecht,R. (2003). Synergism between p68 RNA helicase and the transcriptional coactivators CBP and p300. *Oncogene* *22*, 151-156.
163. Ruppert,S. and Tjian,R. (1995). Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. *Genes Dev.* *9*, 2747-2755.
164. Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U., and Eisenberg,D. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* *32*, D449-D451.
165. Sambrook,J., Fritsch,E.F., and Maniatis,T. (1989). *Molecular Cloning: A Laboratory Manual.*, Cold Spring Harbor Laboratory, ed. (Cold Spring Harbor, NY).
166. Sato,S., Tomomori-Sato,C., Parmely,T.J., Florens,L., Zybaylov,B., Swanson,S.K., Banks,C.A., Jin,J., Cai,Y., Washburn,M.P., Conaway,J.W., and Conaway,R.C. (2004). A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Mol Cell* *14*, 685-691.
167. Schultz,P., Fribourg,S., Poterszman,A., Mallouh,V., Moras,D., and Egly,J.M. (2000). Molecular structure of human TFIIF. *Cell* *102*, 599-607.
168. Shen,W.C., Bhaumik,S.R., Causton,H.C., Simon,I., Zhu,X., Jennings,E.G., Wang,T.H., Young,R.A., and Green,M.R. (2003). Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly. *EMBO J.* *22*, 3395-3402.
169. Shykind,B.M., Kim,J., and Sharp,P.A. (1995). Activation of the TFIID-TFIIA complex with HMG-2. *Genes Dev.* *9*, 1354-1365.

170. Sims,R.J., III, Belotserkovskaya,R., and Reinberg,D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* *18*, 2437-2468.
171. Smale,S.T. and Kadonaga,J.T. (2003). The RNA polymerase II core promoter. *Annu. Rev. Biochem.* *72*, 449-479.
172. Solow,S.P., Lezina,L., and Lieberman,P.M. (1999). Phosphorylation of TFIIA stimulates TATA binding protein-TATA interaction and contributes to maximal transcription and viability in yeast. *Mol. Cell Biol.* *19*, 2846-2852.
173. Struhl,K. (1995). Yeast transcriptional regulatory mechanisms. *Annu. Rev. Genet.* *29*, 651-674.
174. 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., and Sugano,S. (2001). Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res.* *11*, 677-684.
175. Uetz,P., Giot,L., Cagney,G., Mansfield,T.A., Judson,R.S., Knight,J.R., Lockshon,D., Narayan,V., Srinivasan,M., Pochart,P., Qureshi-Emili,A., Li,Y., Godwin,B., Conover,D., Kalbfleisch,T., Vijayadamodar,G., Yang,M., Johnston,M., Fields,S., and Rothberg,J.M. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* *403*, 623-627.
176. Vickers,E.R. and Sharrocks,A.D. (2002). The use of inducible engrailed fusion proteins to study the cellular functions of eukaryotic transcription factors. *Methods* *26*, 270-280.
177. Wassarman,D.A. and Sauer,F. (2001). TAF(II)250: a transcription toolbox. *J. Cell Sci.* *114*, 2895-2902.
178. Watanabe,T., Hayashi,K., Tanaka,A., Furumoto,T., Hanaoka,F., and Ohkuma,Y. (2003). The carboxy terminus of the small subunit of TFIIE regulates the transition from transcription initiation to elongation by RNA polymerase II. *Mol. Cell Biol.* *23*, 2914-2926.
179. Wei,W., Dorjsuren,D., Lin,Y., Qin,W., Nomura,T., Hayashi,N., and Murakami,S. (2001). Direct interaction between the subunit RAP30 of transcription factor IIF (TFIIF) and RNA polymerase subunit 5, which contributes to the association between TFIIF and RNA polymerase II. *J. Biol. Chem.* *276*, 12266-12273.
180. Weis,L. and Reinberg,D. (1997). Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Mol. Cell Biol.* *17*, 2973-2984.

181. Werner,M.H. and Burley,S.K. (1997). Architectural transcription factors: proteins that remodel DNA. *Cell* 88, 733-736.
182. Westover,K.D., Bushnell,D.A., and Kornberg,R.D. (2004). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* 303, 1014-1016.
183. Wilson,B.J., Bates,G.J., Nicol,S.M., Gregory,D.J., Perkins,N.D., and Fuller-Pace,F.V. (2004). The p68 and p72 DEAD box RNA helicases interact with HDAC1 and repress transcription in a promoter-specific manner. *BMC. Mol Biol* 5, 11.
184. Wilson,C.J., Chao,D.M., Imbalzano,A.N., Schnitzler,G.R., Kingston,R.E., and Young,R.A. (1996). RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* 84, 235-244.
185. Wong,J.M. and Bateman,E. (1994). TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs. *Nucleic Acids Res.* 22, 1890-1896.
186. Woychik,N.A. and Hampsey,M. (2002). The RNA Polymerase II Machinery. Structure Illuminates Function. *Cell* 108, 453-463.
187. Woychik,N.A. and Young,R.A. (1989). RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. *Mol. Cell Biol.* 9, 2854-2859.
188. Wozniak,R.W., Rout,M.P., and Aitchison,J.D. (1998). Karyopherins and kissing cousins. *Trends Cell Biol* 8, 184-188.
189. Xu,L.C., Thali,M., and Schaffner,W. (1991). Upstream box/TATA box order is the major determinant of the direction of transcription. *Nucleic Acids Res.* 19, 6699-6704.
190. Yan,Q., Moreland,R.J., Conaway,J.W., and Conaway,R.C. (1999). Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J. Biol. Chem.* 274, 35668-35675.
191. Yeo,M., Lin,P.S., Dahmus,M.E., and Gill,G.N. (2003). A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. *J. Biol Chem.* 278, 26078-26085.
192. Yudkovsky,N., Ranish,J.A., and Hahn,S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408, 225-229.
193. Zawel,L. and Reinberg,D. (1995). Common themes in assembly and function of eukaryotic transcription complexes. *Annu. Rev. Biochem.* 64, 533-561.

194. Zheng,H., Ji,C., Gu,S., Shi,B., Wang,J., Xie,Y., and Mao,Y. (2005). Cloning and characterization of a novel RNA polymerase II C-terminal domain phosphatase. *Biochem. Biophys. Res. Commun.* *331*, 1401-1407.
195. Zheng,L., Hoeflich,K.P., Elsby,L.M., Ghosh,M., Roberts,S.G., and Ikura,M. (2004). FRET evidence for a conformational change in TFIIB upon TBP-DNA binding. *Eur. J. Biochem.* *271*, 792-800.
196. Zhou,T. and Chiang,C.M. (2001). The intronless and TATA-less human TAF(II)55 gene contains a functional initiator and a downstream promoter element. *J. Biol. Chem.* *276*, 25503-25511.
197. Zurita,M. and Merino,C. (2003). The transcriptional complexity of the TFIID complex. *Trends Genet.* *19*, 578-584.