

Direction des bibliothèques

AVIS

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

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

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

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

NOTICE

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

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

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

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

Université de Montréal

**Analyse des complexes protéiques interagissant avec la machinerie
de l'ARN polymérase II dans les cellules humaines**

**par
Célia Jeronimo**

Programme de Biochimie
Faculté de Médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade
Philosophiae Doctor (Ph.D.)
en biochimie

Avril 2008
© Célia Jeronimo, 2008



Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :
Analyse des complexes protéiques interagissant avec la machinerie
de l'ARN polymérase II dans les cellules humaines

Présentée par :
Célia Jeronimo

a été évaluée par un jury composé des personnes suivantes :

Sylvie Mader, présidente-rapporteur

Benoit Coulombe, directeur de recherche

Jacques Côté, membre du jury

Anne-Claude Gingras, examinatrice externe

Jacques Archambault, représentant du doyen de la FES

RÉSUMÉ

L'expression, la réplication et la réparation du génome humain impliquent l'action concertée de diverses machineries moléculaires formant un réseau complexe d'interactions protéine-protéine, protéine-ADN et protéine-ARN. La première étape du processus d'expression d'un gène, soit sa transcription en ARNm par l'ARN polymérase II (ARN Pol II), est régulée par une myriade de protéines, dont les facteurs généraux de transcription (FGT). Des recherches antérieures ont montré que ces facteurs existent dans la cellule sous forme de complexes protéiques capables de s'associer à l'ADN pour exercer leur rôle régulateur. Toutefois, peu est connu sur l'organisation de la machinerie transcriptionnelle dans la fraction nucléaire soluble de la cellule, avant son association à l'ADN ou après son relâchement suite à la transcription.

Dans cette thèse, nous avons purifié par double chromatographie d'affinité à l'aide de protéines étiquetées (TAP-tag) exprimées dans des lignées cellulaires HEK 293, des complexes protéiques solubles contenant des FGT ainsi que l'ARN Pol II afin de définir des réseaux d'interactions protéine-protéine impliquées dans la fonction de la machinerie de l'ARN Pol II. À l'aide de la lignée Rpb11-TAP, nous avons purifié pour la première fois un complexe comprenant l'ARN Pol II, les FGT TFIIB et TFIIF, Fcp1 (la phosphatase de l'ARN Pol II), et un polypeptide à fonction inconnue, que nous avons nommé *RNA polymerase II-associated protein 1* (RPAP1). Nous avons montré que l'ARN Pol II purifiée par la méthode TAP était sous la forme hypophosphorylée et qu'elle était active en transcription *in vitro* et *in vivo*. De plus, nous avons proposé un rôle pour RPAP1 dans la régulation de l'ARN Pol II.

Notre stratégie de purification de complexes protéiques couplée à la spectrométrie de masse et à des analyses bioinformatiques s'est avérée très efficace pour définir des réseaux régulateurs transcriptionnels dans les cellules humaines. Trente-deux protéines étiquetées ont permis de générer un réseau de 805 interactions protéiques validées bioinformatiquement. Tous les facteurs étiquetés que nous avons isolés et testés sont

actifs dans des réactions de transcription *in vitro* et ils ont été localisés le long de gènes humains par la technique d'immunoprécipitation de la chromatine (ChIP) *in vivo*, indiquant que des complexes protéiques fonctionnels sont purifiés par cette méthode. De plus, nous avons purifié XAB1, un polypeptide identifié précédemment en association avec le facteur de réparation XPA, et plusieurs polypeptides à fonction inconnue, dont deux associés à l'ARN Pol II, que nous avons nommés RPAP2 et RPAP3. Nos résultats montrent que les protéines XAB1-RPAPs sont physiquement et fonctionnellement associées à l'ARN Pol II, formant une interface entre l'enzyme et des protéines chaperons et d'échafaudage.

Dans cette thèse, le rôle de BCDIN3, un polypeptide possédant un domaine caractéristique des méthyltransférases, a aussi été étudié. BCDIN3 a été trouvé en association avec le facteur d'élongation P-TEFb ainsi que le petit ARN nucléaire (snARN) 7SK, un snARN jouant un rôle important dans la régulation transcriptionnelle de P-TEFb. Nous avons montré que cette méthyltransférase dont la fonction était à ce jour inconnue est responsable de l'ajout de la coiffe en 5' de 7SK, ce qui lui confère de la stabilité dans les cellules. BCDIN3 a été nommée *methylphosphate capping enzyme* (MEPCE). Dans le même réseau protéique, nous avons aussi identifié LARP7, un polypeptide à fonction inconnue possédant les domaines LA et RRM de liaison à l'ARN. Nous avons pu montrer que LARP7 est un composant stable de la ribonucléoprotéine 7SK•HEXIM•MEPCE•P-TEFb. Finalement, par des expériences d'interférence par l'ARN double brin (siRNA), nous avons pu observer que LARP7 joue un rôle dans la stabilité de 7SK dans les cellules humaines, contribuant ainsi au maintien de P-TEFb dans un état transcriptionnel inactif.

Mots clés : Expression génique, synthèse des ARN messagers, ARN polymérase II, facteurs généraux de transcription, machinerie de transcription, interactions protéine-protéine, petit ARN nucléaire, enzyme de coiffage.

ABSTRACT

The expression, replication and repair of the human genome involve the concerted action of various molecular machineries which form an intricate network of protein-protein, protein-DNA and protein-RNA interactions. The first step of the gene expression process, the transcription of genes by RNA polymerase II (RNAP II), is regulated by a myriad of proteins, including the general transcription factors (GTFs). Previous studies have shown that, in the cell, these factors assemble with other proteins into complexes able to bind DNA to exert their regulatory role. However, very little is known about the organization of the transcription machinery in the soluble nuclear compartment of the cell, prior to or following its association with genomic DNA.

In the following thesis, we have programmed HEK 293 cells to express RNAP II and GTFs subunits carrying tandem affinity purification (TAP) tags. In order to define the protein interaction networks involved in the function of the RNAP II machinery, soluble protein complexes were purified by double affinity chromatography. By using the Rpb11-TAP subunit of RNAP II, we isolated for the first time a complex containing RNAP II, the GTFs TFIIB and TFIIF, the RNAP II phosphatase Fcp1, and a novel polypeptide of unknown function that we named *RNA polymerase II-associated protein 1* (RPAP1). We showed that the TAP-tagged RNAP II complex contains the hypophosphorylated form of RNAP II and is functionally active both *in vitro* and *in vivo*. In addition, we have proposed a role for RPAP1 in the regulation of RNAP II.

Our strategy for the purification of protein complexes coupled to mass spectrometry allowed for the simple and efficient survey of soluble human protein complexes containing components of the transcription machinery. Thirty-two tagged polypeptides yielded a network of 805 high-confidence interactions. Of note, all the tagged transcription factors that we purified and tested were active in transcription reactions *in vitro* and were found in association with transcribed regions using chromatin immunoprecipitation (ChIP) experiments *in vivo*, indicating that *bona fide*, functional

protein complexes were purified. Remarkably, the network is significantly enriched in previously uncharacterized proteins for which we have inferred functions. XAB1, a polypeptide previously identified in association with the DNA repair factor XPA, and two proteins we named RPAP2 and RPAP3, are physically and functionally associated with RNAP II, forming an interface between the enzyme and chaperone/scaffolding proteins.

The role of BCDIN3, a previously uncharacterized protein containing a consensus domain of methyltransferases, was also studied in this thesis. BCDIN3 was found in association with the elongation factor P-TEFb and the 7SK small nuclear RNA (snRNA), an important regulator of P-TEFb. We have been able to determine that BCDIN3 is the 7SK snRNA *methylphosphate capping enzyme* (MEPCE). Our results suggest that MEPCE stabilizes the 7SK snRNA through the addition of a methylphosphate cap at its 5' end. In the same protein network, we have also identified LARP7, a polypeptide of unknown function containing the LA and RRM domains, which are involved in RNA binding and recognition. We determined that LARP7 is a stable component of the 7SK•HEXIM•MEPCE•P-TEFb ribonucleoprotein. Finally, we showed that the knockdown of LARP7 decreases the steady-state level of 7SK snRNA, indicating that LARP7 stabilizes the 7SK in human cells, thereby contributing to the maintenance of P-TEFb in an inactive state.

Key words : Gene expression, mRNA synthesis, RNA polymerase II, general transcription factors, transcription machinery, protein-protein interactions, small nuclear RNA, capping enzyme.

TABLE DES MATIÈRES

RÉSUMÉ DE LA THÈSE.....	iii
ABSTRACT.....	v
TABLE DES MATIÈRES.....	vii
LISTE DES TABLEAUX.....	xi
LISTE DES FIGURES.....	xii
LISTE DES ABRÉVIATIONS.....	xvii
REMERCIEMENTS.....	xxi
CHAPITRE 1.....	1
INTRODUCTION.....	1
1.1 Mécanismes de régulation de la transcription.....	2
1.1.1 La chromatine.....	3
1.1.1.1 La méthylation de l'ADN.....	4
1.1.1.2 Les complexes de remodelage de la chromatine.....	6
1.1.1.3 Les modifications post-traductionnelles des histones.....	9
1.1.1.4 L'incorporation des variants d'histones.....	12
1.1.2 Les facteurs liant l'ADN.....	14
1.1.2.1 Les séquences d'ADN régulatrices.....	15
1.1.2.2 Les activateurs transcriptionnels.....	18
1.1.2.3 Les répresseurs transcriptionnels.....	20
1.1.2.4 Le complexe Médiateur.....	22
1.1.3 Les ARN non-codants.....	25
1.1.3.1 Le snARN 7SK.....	27
1.1.3.2 Le snARN U1.....	30
1.1.3.3 Le ncARN B2.....	31
1.2 La transcription par l'ARN polymérase II.....	33
1.2.1 La machinerie transcriptionnelle.....	33

1.2.1.1 L'ARN Pol II.....	33
1.2.1.2 Les facteurs généraux de transcription.....	39
1.2.1.2.1 TFIID.....	40
1.2.1.2.2 TFIIA.....	42
1.2.1.2.3 TFIIB.....	44
1.2.1.2.4 TFIIF.....	45
1.2.1.2.5 TFIIE.....	46
1.2.1.2.6 TFIIH.....	47
1.2.1.3 Les facteurs d'élongation.....	50
1.2.1.3.1 Les facteurs d'élongation modulant l'activité de l'ARN Pol II.....	50
1.2.1.3.2 Les facteurs d'élongation modifiant la structure de la chromatine.....	54
1.2.1.3.3 Rôle du facteur d'élongation P-TEFb dans la transcription du VIH-1.....	56
1.2.1.4 Les facteurs de terminaison.....	57
1.2.2 Le cycle transcriptionnel.....	58
1.2.2.1 Le promoteur basal.....	59
1.2.2.2 L'assemblage du complexe de pré-initiation au promoteur.....	61
1.2.2.3 L'initiation de la synthèse de l'ARNm.....	63
1.2.2.4 L'élongation de l'ARNm.....	64
1.2.2.5 La terminaison de la transcription.....	65
1.3 Couplage entre la transcription et la maturation de l'ARNm.....	66
1.3.1 Les étapes de la maturation de l'ARNm.....	68
1.3.1.1 L'ajout d'une structure coiffe en 5'.....	68
1.3.1.2 L'épissage alternatif.....	69
1.3.1.3 Le clivage et la polyadénylation en 3'.....	71
1.3.2 Le rôle du CTD de l'ARN Pol II.....	73
1.4 Objectifs de recherche.....	74
CHAPITRE 2.....	75
RPAP1, UNE NOUVELLE PROTÉINE ASSOCIÉE À L'ARN POLYMERASE II PURIFIÉE À PARTIR DE SOUS-UNITÉS RECOMBINANTES SAUVAGES ET MUTANTES.....	75
2.1 Préambule.....	75
2.2 Article.....	77

CHAPITRE 3.....	94
L'ANALYSE SYSTÉMATIQUE DU RÉSEAU D'INTERACTIONS PROTÉIQUES DE LA MACHINERIE DE TRANSCRIPTION RÉVÈLE L'IDENTITÉ DE L'ENZYME DE COIFFAGE DU snARN 7SK.....	94
3.1 Préambule.....	94
3.2 Article.....	96
CHAPITRE 4.....	120
LOCALIZATION GÉNOMIQUE DE LA MACHINERIE GÉNÉRALE DE L'ARN POLYMÉRISE II : ÉVIDENCE D'UN RÔLE POUR TFIIF ET RPB7 AUX STADES D'INITIATION ET D'ÉLONGATION DE LA TRANSCRIPTION.....	120
4.1 Préambule.....	120
4.2 Article.....	122
CHAPITRE 5.....	132
LARP7 EST UNE COMPOSANTE STABLE DE LA snRNP 7SK ALORS QUE P-TEFb, HEXIM1 ET hnRNP A1 SONT ASSOCIÉS DE FAÇON RÉVERSIBLE.....	132
5.1 Préambule.....	132
5.2 Article	133
CHAPITRE 6.....	145
DISCUSSION.....	145
6.1 Étude protéomique de la machinerie de l'ARN polymérase II : purification de complexes protéiques et établissement de réseaux d'interactions protéine-protéine.....	146
6.2 Complexes protéiques contenant l'ARN Pol II en association avec des régulateurs impliqués dans l'assemblage des protéines : les protéines XAB1-RPAPs.....	155
6.3 Complexe ribonucléoprotéique 7SK snARN•HEXIM•P-TEFb : rôle des protéines BCDIN3/MEPCE et LARP7.....	157
6.3.1 Identification de l'enzyme de coiffage de 7SK.....	157
6.3.2 Rôle des protéines BCDIN3/MEPCE et LARP7 dans la stabilité du snARN 7SK.....	159

6.3.3 Rôle des protéines BCDIN3/MEPCE et LARP7 dans la régulation de l'activité de P-TEFb.....	161
CONCLUSIONS ET PERSPECTIVES.....	164
BIBLIOGRAPHIE.....	167

LISTE DES TABLEAUX

CHAPITRE 1

I. Les facteurs de remodelage de la chromatine.....	8
II. Les domaines protéiques impliqués dans la reconnaissance des modifications des histones	11
III. Les facteurs généraux de transcription humains.....	39

CHAPITRE 4

I. Sequences of the primers used to analyse the location of various polypeptides.....	126
---	-----

CHAPITRE 6

I. Liste des protéines étiquetées et des partenaires d'interaction co-purifiés selon leur GO (<i>Gene Ontology</i>).....	154
---	-----

LISTE DES FIGURES

CHAPITRE 1

1. Mécanismes de régulation de l'expression génique.....	2
2. Éléments de régulation distaux.....	16
3. Régulation de la réaction transcriptionnelle de l'ARN Pol II par les petits ARN nucléaires (snARN) 7SK, U1 et B2.....	26
4. Modèle de régulation du niveau nucléaire de P-TEFb actif par la réorganisation dynamique et réversible des snRNP contenant 7SK.....	29
5. Modèle illustrant le mécanisme de répression transcriptionnelle par le ncARN B2 après le choc thermique dans les cellules de souris.....	33
6. Structure cristallographique de l'ARN polymérase II de levure manquant Rpb4 et Rpb7.....	36
7. Structure de l'ARN polymérase II à 12 sous-unités.....	37
8. Modulation de l'étape d'élongation de l'ARN Pol II par les facteurs P-TEFb, NELF, DSIF, TFIIF, Elongins, ELL et TFIIS.....	51
9. P-TEFb est nécessaire pour la transcription du VIH-1.....	56
10. Le cycle transcriptionnel de l'ARN Pol II.....	59
11. Représentation schématique du couplage entre la transcription et	

la maturation de l'ARNm.....	67
------------------------------	----

CHAPITRE 2

1. TAP of human transcription factors.....	80
2. Purification of TAP-tagged human RNAPII complex.....	81
3. Regions of homology between <i>Homo sapiens</i> RPAP1, <i>D. melanogaster</i> CG32104-PB, and <i>S. cerevisiae</i> Ydr527wp/RBA50, as determined by BLAST analyses.....	83
4. TAP-tagged human RNAPII assembles on promoter DNA both <i>in vitro</i> and <i>in vivo</i> and binds to acetylated histones.....	84
5. TAP-tagged human RNAPII complex can initiate transcription <i>in vitro</i>	85
6. Purification and functional analysis of TAP-tagged human RNAPII with a mutation in fork loop 1.....	87
7. Purification and functional analysis of TAP-tagged human RNAPII with a mutation in switch 3.....	89
8. Association of mutant forms of the RNAPII complex with chromatin <i>in vivo</i>	90

CHAPITRE 3

1. A method for the analysis of protein complexes in human cells.....	99
---	----

2. A network of protein complexes involving the RNAP II basal transcription machinery.....	100
3. Previously uncharacterized protein complexes containing RNAP II in association with regulators of protein complex assembly.....	101
4. A complex comprising the elongation factor P-TEFb, the 7SK and U6 snRNAs, and the previously uncharacterized protein BCDIN3.....	103
5. BCDIN3 is a conserved methyltransferase that targets the 7SK snRNA.....	104
6. BCDIN3 is the 7SK snRNA MePCE.....	105
7. BCDIN3 silencing decreases the steady-state level of cellular 7SK.....	106
S1. A method for the purification of protein complexes in human cells.....	118
S2. Computational validation of protein interactions.....	119

CHAPITRE 4

1. TAP of the RNAPII general transcription machinery.....	127
2. Occupancy of regions proximal to the TIS and further downstream TRs of selected human genes by the GTF and RNAPII subunits obtained by ChIP.....	128
3. Occupancy of regions proximal to the TIS and further downstream TRs of selected human genes by the TFIIF subunit RAP30 and the RNAPII subunit Rpb11 as determined by ChIP.....	129

4. Occupancy of regions proximal to the TIS and further downstream TRs of selected human genes by Rpb7 as determined by ChIP.....	130
--	-----

CHAPITRE 5

1. LARP7 is associated with P-TEFb, HEXIM, MEPCE, 7SK snRNA and U6 snRNA.....	137
2. LARP7 co-sediments with the 7SK snRNP before and after P-TEFb and HEXIM release.....	138
3. Immunoprecipitation of LARP7 confirms that it is a stable component of the 7SK snRNP.....	139
4. Knockdown of LARP7 decreases 7SK levels, increases free P-TEFb, and enhances Tat transactivation.....	140
5. Direct knockdown of 7SK disrupts the large form of P-TEFb and decreases total P-TEFb.....	141
6. Model of the 7SK snRNP before and after release of P-TEFb.....	141

CHAPITRE 6

1. Procédure générale pour définir des réseaux d'interactions protéine-protéine et protéine-ADN.....	147
2. Stratégie pour la caractérisation de réseaux d'interactions protéine-protéine dans les cellules humaines	150
3. Carte d'un réseau à haute densité d'interactions protéine-protéine impliquant des facteurs de transcription de l'ARN Pol II et de maturation de l'ARNm.....	152
4. Modèle de la participation de MEPCE et LARP7 à la stabilisation du snARN 7SK, contribuant au maintien de P-TEFb dans un état transcriptionnel inactif.....	162

LISTE DES ABRÉVIATIONS

ActD	Actinomycine D
ADN	Acide désoxyribonucléique
ADP	Adénosine diphosphate
ARC	<i>Activator-recruited factor</i>
ARN	Acide ribonucléique
ARNm	ARN messenger
ARN Pol	ARN polymérase
ARNr	ARN ribosomiaux
ARNt	ARN de transfert
Asf1	<i>Anti-silencing function 1</i>
ATP	Adénosine triphosphate
ATPase	ATP hydrolase
Brd4	<i>Bromodomain protein 4</i>
BRE	<i>TFIIB recognition element</i>
CAF-1	<i>Chromatin assembly factor 1</i>
CAK	<i>Cyclin-activating kinase complex</i>
CBP	<i>CREB-binding protein et Calmodulin-binding peptide</i>
CDD	<i>Charge clustered domain</i>
CDK	<i>Cyclin dependant kinase</i>
CENP-A	<i>Centromeric protein A</i>
CFIm	<i>Cleavage factor I</i>
CFIIm	<i>Cleavage factor II</i>
CHD	<i>Chromodomain SWI/SNF-like helicase/ATPase domain and DNA binding domain</i>
ChIP	Immunoprécipitation de la chromatine
CoTC	<i>Co-transcriptional cleavage</i>
CPF	<i>Cleavage/polyadenylation factor</i>
CPSF	<i>Cleavage and polyadenylation specificity factor</i>

CRSP	<i>Cofactor required for Sp1 activation</i>
CS	Syndrome de Cockayne
CstF	<i>Cleavage stimulatory factor</i>
CTD	<i>Carboxy-terminal domain</i>
C-terminal	Carboxy-terminal
DCE	<i>Downstream core element</i>
DES	<i>Downstream sequence element</i>
DPE	<i>Downstream promoter element</i>
DRB	5,6-dichloro-1- β -D-ribofuranosylbenzimidazole
DRIP	<i>Vitamin-D-receptor interacting proteins</i>
DSIF	<i>DRB-sensitivity inducing factor</i>
ELL	<i>Eleven-nineteen lysine-rich in leukemia</i>
ESI	<i>Electrospray ionization</i>
FACT	<i>Facilitates chromatin transcription</i>
Fcp1	<i>TFIIF-associated CTD phosphatase</i>
FGT	Facteurs généraux de transcription
GCN	<i>General control of amino acid synthesis</i>
GO	<i>Gene ontology</i>
GT	ARN guanylyltransférase
HAT	<i>Histone acétyltransférase</i>
HEXIM1	<i>Hexamethylene bisacetamide-induced protein-1</i>
HIRA	<i>Histone regulating protein A</i>
His	Histidine
HDAC	Histone désacétylase
HMG	<i>High mobility group</i>
HMT	Histone méthyltransférase
Inr	Initiateur
ISWI	<i>Imitation of SWI</i>
LCR	<i>Locus control region</i>
MALDI	<i>Matrix-assisted laser desorption/ionization</i>
MAT	Ménage à trois

MBD	<i>Methyl-CpG binding protein</i>
MEPCE	<i>Methylphosphate capping enzyme</i>
MS	<i>Mass spectrometry</i>
MT	ARN méthyltransférase
MudPIT	<i>Multidimensional protein identification technology</i>
MTE	<i>Motif ten element</i>
NAP-1	<i>Nucleosome assembly protein 1</i>
NAT	<i>Negative regulator of activated transcription</i>
ncARN	ARN non-codant
NCoR	<i>Nuclear receptor corepressor</i>
NELF	<i>Negative elongation factor</i>
NER	<i>Nucleotide excision repair</i>
N-terminal	Amino-terminal
NTP	Ribonucléotide triphosphate
NuRD	<i>Nucleosome remodeling factor</i>
PABPN1	<i>Poly A binding protein</i>
PAP	Poly(A) polymérase
pb	Paire de base
PCAF	<i>p300/CBP-associated factor</i>
PC2	<i>Positive cofactor 2</i>
P-TEFb	<i>Positive transcription elongation factor b</i>
RAP	<i>RNA polymerase II-associated protein</i>
RHA	ARN hélicase A
rNTP	Ribonucléotides triphosphates
RPAP	<i>RNA polymerase II-associated protein</i>
RRM	<i>RNA recognition motif</i>
RT	ARN triphosphatase
SAGA	SPT-ADA-GCN5 acétyltransférase
SCP	<i>Small CTD phosphatases</i>
Ser	Sérine
siARN	<i>Short interference RNA</i>

SINES	<i>Short interspersed repetitive elements</i>
SKIP	<i>c-Ski-interacting protein</i>
SMCC	<i>SRB-Med-containing cofactor</i>
snARN	<i>Small nuclear RNA</i>
SNF	<i>Sucrose non fermenting</i>
snoARN	<i>Small nucleolar RNA</i>
snRNP	<i>Small nuclear ribonucleoprotein</i>
Spt	<i>Supressor of Ty insertion</i>
SR	<i>Serine/arginine-rich</i>
Srb	<i>Supressor of RNA polymerase B</i>
SREBP	<i>Sterol-response-element-binding protein</i>
STAGA	<i>Supressor of Ty insertion-TAF-Gcn5-acetyltransferase</i>
SWI	<i>Switching mating-type</i>
TAF	<i>TBP-associated factor</i>
TAP	<i>Tandem affinity purification</i>
TBP	<i>TATA binding protein</i>
TCR	<i>Transcription coupled repair</i>
TF	<i>Transcription factor</i>
TFTC	<i>TBP-free TAF-containing complex</i>
TEV	<i>Tobacco etch virus</i>
TRAP	<i>Thyroid hormone receptor</i>
TRF	<i>TBP related factor</i>
TTD	<i>Trichothiodystrophie</i>
UAS	<i>Upstream activating sequence</i>
UV	<i>Ultraviolet</i>
VIH	<i>Virus de l'immunodéficience acquise humaine</i>
VP16	<i>Herpes simplex virion protein 16</i>
XP	<i>Xeroderma pigmentosum</i>

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de recherche, Benoit Coulombe, pour m'avoir accepté dans son laboratoire pour réaliser mon projet de doctorat alors qu'il ne me connaissait pas. Je le remercie pour sa grande disponibilité, ses précieux conseils et son soutien lorsque les résultats se faisaient attendre. Je le remercie aussi d'avoir en toutes occasions pris le temps de m'écouter et de me comprendre.

Je remercie les membres du jury, le Dr Sylvie Mader, le Dr Anne-Claude Gingras, le Dr Jacques Côté et le Dr Jacques Archambault, d'avoir accepté d'évaluer mon travail.

J'aimerais remercier également tous les membres du laboratoire qui m'ont accompagné au cours de mon doctorat et qui ont contribué de près ou de loin à cette thèse. Je les remercie aussi pour les encouragements, leur gentillesse et la bonne humeur. Un remerciement spécial est fait à Diane Forget et Annie Bouchard pour tous leurs conseils et suggestions ainsi que pour leur support technique. Je remercie Marie-France Langelier et Marilena Cojocarú d'avoir collaboré efficacement pour la réalisation de ce travail.

Je remercie mes amis et ma famille pour m'avoir soutenue tout au long de ces études. En particulier ma mère qui m'a toujours encouragé dans cette voie et Telmo de Oliveira pour sa grande compréhension et son soutien moral.

CHAPITRE 1

INTRODUCTION

Dans la cellule, l'exécution fidèle des divers processus biologiques, tels que le développement, la prolifération et l'apoptose, exige un ensemble précis et hautement orchestré d'étapes qui dépendent de l'expression appropriée des gènes. La régulation de la transcription, soit la synthèse d'un ARN à partir d'une matrice d'ADN, est une des étapes les plus importantes pour le contrôle de la croissance et de la différenciation cellulaires. La transcription est catalysée par les enzymes ARN polymérases (ARN Pol). Chez les organismes procaryotes, les divers types d'ARN cellulaires sont synthétisés par une seule ARN Pol, alors que chez les eucaryotes, il existe trois ARN polymérases (ARN Pol I, II et III). L'ARN Pol I est impliquée dans la synthèse des ARN ribosomiaux (ARNr), l'ARN Pol II, des ARN messagers (ARNm) et des petits ARN nucléaires (snARN) U1-U5 et l'ARN Pol III, des ARN de transfert (ARNt), des snARN U6 et 7SK ainsi que de l'ARNr 5S.

La transcription par l'ARN Pol II est un processus étroitement régulé par des complexes protéiques qui interviennent séquentiellement afin de permettre la synthèse de l'ARNm. La machinerie basale de transcription de l'ARN Pol II est composée d'un ensemble de complexes protéiques comprenant l'ARN Pol II et les facteurs généraux de transcription (FGT) TFIIA, TFIIB, TFIID, TFIIE, TFIIIF et TFIIH. D'autres facteurs assistent l'ARN Pol II lors des différentes étapes du cycle transcriptionnel, tels les facteurs d'élongation et les facteurs de terminaison de la transcription. Dans cette thèse, nous nous sommes intéressés à l'analyse des complexes protéiques interagissant avec la machinerie de l'ARN Pol II dans les cellules humaines. L'introduction présentée décrit les mécanismes de régulation de la transcription par l'ARN Pol II, le rôle des différents composants de la machinerie de transcription dans les étapes du cycle transcriptionnel et le couplage entre la transcription et la maturation de l'ARNm.

1.1 Mécanismes de régulation de la transcription

La régulation de l'expression génique chez les eucaryotes est un processus complexe. Chaque cellule doit réguler au moins deux aspects de la réaction transcriptionnelle: l'accessibilité de l'ARN Pol II à la région promotrice et la transcription subséquente du gène cible (Dunn et Kingston, 2007). Des progrès marquants ont permis de démontrer que ces deux étapes peuvent être régulées par des facteurs liant directement des séquences régulatrices spécifiques sur l'ADN, notamment des activateurs et des répresseurs transcriptionnels, ainsi que leurs cofacteurs, comme par exemple le complexe Médiateur (figure 1) (Kornberg, 2005; Conaway et al., 2005; Bjorklund et Gustafsson, 2005; Kadonaga, 2004; Roeder, 2005). Toutefois, il est devenu de plus en plus évident au cours de la dernière décennie que la réaction transcriptionnelle peut aussi être régulée par d'autres mécanismes moléculaires incluant la chromatine ainsi que les petits ARN non-codants (figure 1) (Li et al., 2007; Goodrich et Kugel, 2006). Ces différents facteurs régulateurs du processus transcriptionnel seront discutés dans la section qui suit.

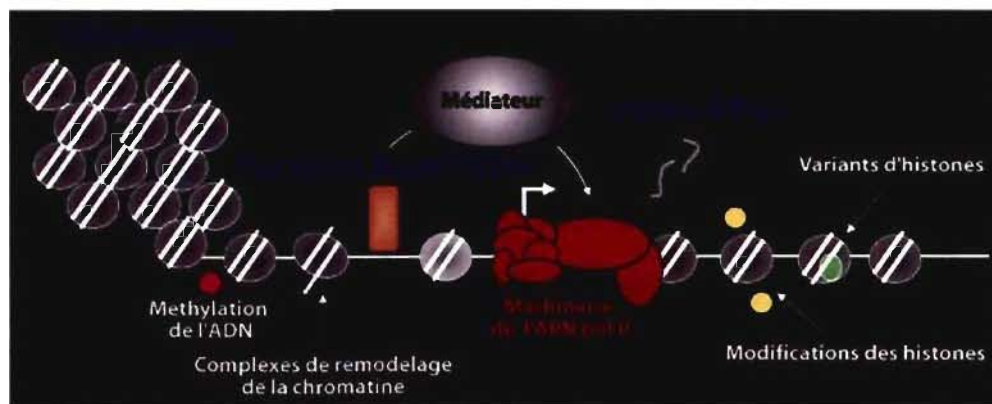


Figure 1 : Mécanismes de régulation de l'expression génique. Les facteurs régulateurs (en bleu) de la transcription par l'ARN Pol II sont représentés sous forme schématique. Dans cette représentation, les différentes altérations de la chromatine sont indiquées : la méthylation de l'ADN est représentée par un cercle rose ; les modifications des histones par des cercles jaunes et les variants d'histones par un cercle vert. La machinerie de l'ARN Pol II recrutée au promoteur du gène cible est indiquée en rouge et le complexe Médiateur en gris. Figure modifiée de (Dunn et Kingston, 2007).

1.1.1 La chromatine

Dans le noyau des cellules eucaryotes, le génome est hautement compacté et organisé sous la forme d'une structure appelée chromatine (revu par (Olins et Olins, 2003; Woodcock, 2006)). La chromatine est structurée en domaines fonctionnels, notamment l'euchromatine et l'hétérochromatine (Sarma et Reinberg, 2005). L'euchromatine consiste en un état chromatinien décondensé connu pour contenir des régions transcriptionnellement actives. Quant à l'hétérochromatine, cette forme de la chromatine est hautement compacte et elle possède des régions transcriptionnellement silencieuses. L'unité de base de la chromatine, le nucléosome, est formé d'une particule cœur et d'une région internucléosomique (Kornberg et Lorch, 1999). La particule cœur est composée de 147 paires de bases (pb) d'ADN enroulées 1,65 fois autour d'un octamère de protéines comprenant deux copies de chacune des histones H2A, H2B, H3 et H4 (Luger et al., 1997; Davey et al., 2002; Richmond et Davey, 2003). Au sein du nucléosome, les histones H3 et H4 forment un tétramère, (H3-H4)₂, entouré de part et d'autre par un dimère H2A-H2B. Les histones sont des protéines basiques constituées d'un domaine globulaire formé de trois hélices α . Ce domaine, appelé *histone-fold*, est très conservé au cours de l'évolution et permet la dimérisation des histones. Les extrémités amino- et carboxy-terminales entourant le domaine globulaire des histones émergent à la surface du nucléosome et peuvent interagir avec de nombreux régulateurs protéiques, l'ADN internucléosomique et l'ADN des nucléosomes avoisinants. Ces extrémités des histones sont la cible de multiples modifications post-traductionnelles (Kouzarides, 2007; Bhaumik et al., 2007). Des données récentes montrent que les histones de la particule cœur du nucléosome peuvent être remplacées par des variants qui confèrent une spécialisation fonctionnelle de la chromatine à des endroits précis du génome (Sarma et Reinberg, 2005; Henikoff et al., 2004). Quant à la région internucléosomique, elle est composée d'environ 20 pb d'ADN (ADN *linker*) liées par l'histone H1 (Woodcock et al., 2006). En s'associant avec l'ADN présent entre chaque nucléosome, l'histone H1 joue un rôle clé dans la formation et le maintien d'un état chromatinien plus compact : la fibre de chromatine de 30 nm. Les fibres de chromatine

peuvent ensuite se replier pour former une structure de compaction ultime de 100-300 nm observée au niveau du chromosome lors de la division cellulaire (Tremethick, 2007).

Au-delà d'un rôle structural, cette organisation compacte du matériel génétique en chromatine influence profondément tous les événements moléculaires faisant intervenir l'ADN, tels la réplication, la transcription, la réparation et la recombinaison (Groth et al., 2007; Li et al., 2007; Grewal et Elgin, 2007). Pour réguler adéquatement ces processus biologiques nucléaires, et en particulier, la transcription des gènes, des changements dans la structure globale de la chromatine sont nécessaires (Li et al., 2007). Ainsi, en réponse à des stimuli physiologiques et environnementaux d'activation ou de répression, la structure de la chromatine doit être modifiée afin de favoriser ou de réprimer la transcription de gènes cibles. Pour moduler la transcription à partir de la chromatine, la cellule met en œuvre plusieurs mécanismes incluant la méthylation de l'ADN, le recrutement de complexes de remodelage de la chromatine, les modifications post-traductionnelles des histones et l'incorporation des variants d'histones dans la chromatine (figure 1) (Henikoff, 2008). Ces modes de régulation sont dits « épigénétiques » car ils modifient l'expression des gènes de manière héritable au cours des divisions cellulaires, sans affecter directement la séquence d'ADN (Goldberg et al., 2007).

1.1.1.1 La méthylation de l'ADN

La méthylation de l'ADN est probablement la mieux caractérisée des modifications de la chromatine (Goldberg et al., 2007). Chez les mammifères, cette modification est localisée presque exclusivement sur les résidus cytosines au sein de dinucléotides CpG. Dans le génome, les domaines riches en ces dinucléotides sont appelés îlots CpG et ils sont positionnés au niveau du promoteur et/ou du premier exon de plus de 60% des gènes humains (Bird, 2002). La méthylation de l'ADN est catalysée par un groupe de protéines, les méthyltransférases de l'ADN (DNMTs), qui transfèrent de groupements méthyles sur la portion 5' des cytosines (Goll et Bestor, 2005; Gal-Yam et al., 2007). Il existe au moins trois DNMTs dans la cellule: DNMT1, DNMT3a et

DNMT3b. DNMT1 est l'enzyme capable de la transmission du profil de méthylation de la cellule mère aux cellules filles au cours de la réplication de l'ADN et elle est appelée la DNMT de maintien (Groth et al., 2007). Quant aux protéines DNMT3a et DNMT3b, ces méthyltransférases dites *de novo* catalysent la méthylation des cytosines des îlots de CpGs qui n'ont pas été préalablement méthylés (Okano et al., 1999).

La méthylation de l'ADN est la marque d'une chromatine transcriptionnellement silencieuse et elle constitue une modification épigénétique essentielle au développement normal (Bird, 2002). En effet, les études montrent que la méthylation de l'ADN est cruciale pour plusieurs processus biologiques associés au développement, comme le silence de séquences d'ADN répétitives adjacentes aux centromères, l'inactivation du chromosome X chez les mammifères femelles ou encore l'empreinte génomique parentale (Goll et Bestor, 2005). L'inactivation de l'un des gènes responsables de catalyser cette méthylation induit chez la souris de nombreuses anomalies du développement, voire une létalité embryonnaire (Okano et al., 1999; Li et al., 1992). Généralement, les îlots CpGs ne sont pas méthylés au cours des divers stades du développement, permettant ainsi l'expression d'un gène spécifique de tissus en particulier lorsque les facteurs de transcription appropriés sont présents et que la structure de la chromatine leur est favorable (Esteller, 2007). Par contre, dans certaines circonstances, les îlots de CpGs peuvent devenir méthylés et résulter dans un silence génique. D'autre part, une méthylation aberrante de l'ADN semble jouer un rôle important dans le processus de tumorigenèse (Jones et Baylin, 2007).

Des progrès considérables réalisés au cours des dernières années ont permis de mieux comprendre comment la méthylation de l'ADN réprime l'expression génique. Deux modèles de répression transcriptionnelle par la méthylation de l'ADN ont été proposés (Klose et Bird, 2006). Le premier est basé sur l'encombrement stérique engendré par l'ajout du groupement méthyle (Watt et Molloy, 1988). Cette modification de l'ADN peut ainsi réprimer la transcription en empêchant la fixation d'activateurs transcriptionnels aux séquences régulatrices spécifiques sur l'ADN méthylé. Le second modèle, plus récent, est quant à lui davantage complexe et repose sur le recrutement et la

coopération de plusieurs facteurs modifiant la structure de la chromatine (revu par (Esteller, 2007)). Selon ce modèle, l'information retenue dans les îlots CpGs méthylés est interprétée par des facteurs ayant une forte affinité pour l'ADN méthylé, les protéines MBDs (*methyl-CpG binding proteins*) (revu par (Goll et Bestor, 2005)). Cette famille de protéines est composée de cinq membres bien caractérisés : MeCP2, MBD1, MBD2, MBD3 et MBD4 (Ballestar et Esteller, 2005). Les MBDs jouent un rôle clé dans le maintien de la répression des gènes méthylés par le recrutement de complexes ayant des activités de modifications des histones, comme la désacétylation et la méthylation, entraînant l'extinction de la transcription.

1.1.1.2 Les complexes de remodelage de la chromatine

Les complexes de remodelage de la chromatine consistent en des complexes multi-protéiques qui utilisent l'énergie libérée par l'hydrolyse de l'ATP afin d'altérer les interactions histones-ADN, rendant l'ADN nucléosomique plus accessible à la machinerie de transcription (revu par (Li et al., 2007; Saha et al., 2006)). Ces complexes fonctionnent en modifiant le positionnement, la composition et la structure des nucléosomes. Les complexes de remodelage peuvent affecter le nucléosome d'au moins quatre façons (Cairns, 2007). Ainsi, ils peuvent (i) induire le glissement ou modifier le positionnement de l'octamère d'histones, ce qui expose l'ADN; (ii) éjecter ou complètement repositionner l'octamère afin de rendre accessible l'ADN; (iii) enlever les dimères H2A-H2B, laissant seulement le tétramère (H3-H4)₂ au centre, ce qui expose l'ADN et déstabilise le nucléosome; et enfin, (iv) remplacer les dimères des histones de la particule cœur par des variants d'histones (Cairns, 2007). Les complexes de remodelage ont tous en commun la présence d'une sous-unité catalytique ATP hydrolase (ATPase) appartenant à la superfamille des ADN hélicases/ATPases *SNF2* (Lusser et Kadonaga, 2003). Au cours des dernières années, des travaux chez la levure ont révélé que les sous-unités ATPases de certains complexes de remodelage sont en fait des translocases à ADN ATP-dépendantes (Saha et al., 2002; Whitehouse et al., 2003).

Chez les eucaryotes, les divers facteurs de remodelage ont été classifiés en quatre grandes familles basées sur leurs homologies de séquence et leurs activités biochimiques: SWI/SNF, ISWI, CHD/NuRD et INO80/SWR1 (tableau I) (Langst et Becker, 2004). La caractéristique distinguant SWI/SNF des autres familles est la présence dans la partie C-terminale d'un bromodomaine, un domaine protéique pouvant lier des lysines acétylées sur les extrémités des histones (Winston et Allis, 1999; Hassan et al., 2002). Les complexes SWI/SNF sont principalement impliqués dans la réorganisation du positionnement des nucléosomes afin de favoriser la liaison des facteurs de transcription entraînant une activation transcriptionnelle (Martens et Winston, 2003). Cependant, les travaux du groupe de Winston chez la levure démontrent que ces complexes de remodelage peuvent également induire la liaison de répresseurs transcriptionnels au niveau de certains gènes (Martens et Winston, 2002). De plus, les membres de la famille SWI/SNF sont capables de faire éjecter les nucléosomes de l'ADN (Lorch et al., 2006). De leur côté, les complexes de remodelage ISWI sont caractérisés par la présence de domaines SANT à leur extrémité C-terminale (Grune et al., 2003). Le domaine SANT est similaire aux domaines de liaison à l'ADN retrouvés dans le corépresseur NCoR et le facteur de transcription TFIIB. La fonction principale des complexes de remodelage ISWI est d'organiser et d'ordonner le positionnement des nucléosomes, conduisant à la répression transcriptionnelle (Deuring et al., 2000; Goldmark et al., 2000). D'autre part, les complexes ISWI sont capables de structurer la chromatine afin de promouvoir l'élongation transcriptionnelle (Martens et Winston, 2003). Les études montrent que tous les membres des familles SWI/SNF et ISWI peuvent faire éjecter les dimères d'histones (Bruno et al., 2003; Yang et al., 2007). Quant à la famille CHD/NuRD, elle est caractérisée par la présence de deux chromodomaines à leur extrémité N-terminale (Marfella et Imbalzano, 2007). Les modules chromodomaines agissent comme des surfaces d'interaction avec divers composants de la chromatine comme l'ADN et l'histone H3 méthylée (Brehm et al., 2004; Flanagan et al., 2005). Le membre le plus étudié de cette famille, l'ATPase Mi-2, appartient au complexe NuRD possédant aussi une activité histone désacétylase (Feng et Zhang, 2003). Ces complexes catalysent le glissement du nucléosome *in vitro* (Brehm et al., 2000). Finalement, les membres de la famille INO80/SWR1 comportent plusieurs sous-unités dont les ADN hélicases Rvb1p et

Rvb2p (Tip49a et Tip49b chez les mammifères) (Bao et Shen, 2007). Les complexes INO80 et SWR1 sont tous les deux impliqués dans la réparation de l'ADN. De plus, le complexe INO80 favorise l'activation transcriptionnelle (Gangaraju et Bartholomew, 2007). D'autre part, le complexe SWR1 chez la levure, appelé SRCAP et TIP60 chez l'humain, permet l'échange, au sein du nucléosome, de l'histone H2A par son variant, l'histone H2A.Z (Kobor et al., 2004; Krogan et al., 2003b; Mizuguchi et al., 2004).

Tableau I. Les facteurs de remodelage de la chromatine

Famille	Domaine	Caractéristiques
SWI/SNF	Bromodomaine	Lie les lysines acétylées, est impliqué dans l'activation et la répression de certains gènes
ISWI	SANT	Lie l'ADN, est impliqué dans la répression transcriptionnelle, peut promouvoir l'élongation
CHD/NuRD	Double Chromodomaine	Lie l'ADN et l'histone H3 méthylée, est impliqué dans la répression transcriptionnelle
INO80/SWR1		Favorise l'activation transcriptionnelle, est impliqué dans la réparation de l'ADN

Une régulation efficace des complexes de remodelage de la chromatine est essentielle afin d'assurer que la structure chromatinienne soit modifiée au bon endroit et au bon moment de manière appropriée dans le noyau (Hogan et Varga-Weisz, 2007). Il existe plusieurs processus permettant de réguler l'activité des complexes de remodelage des nucléosomes incluant les modifications post-traductionnelles de la sous-unité catalytique, les modifications post-traductionnelles des histones et l'incorporation des variants d'histones (Hogan et Varga-Weisz, 2007). D'autre part, des protéines interagissant avec les histones, connues sous le nom de chaperons d'histones, jouent un rôle clé dans l'assemblage et les dynamiques du nucléosome pendant la transcription des gènes (Li et al., 2007; Akey et Luger, 2003). Par exemple, les chaperons nucléophosmine, FACT (*facilitates chromatin transcription*), NAP-1 (*nucleosome assembly protein 1*) et Asf1 (*anti-silencing function 1*) forment des complexes avec les

dimères d'histones H2A-H2B et H3-H4, permettant leur transfert au cours de la réaction transcriptionnelle (De Koning et al., 2007; Mousson et al., 2007; Lorch et al., 2006).

1.1.1.3 Les modifications post-traductionnelles des histones

Les extrémités amino-terminales des histones subissent de nombreuses modifications post-traductionnelles (revu par (Kouzarides, 2007; Bhaumik et al., 2007; Berger, 2007)). Dans une moindre mesure, des modifications ont aussi été identifiées dans les extrémités C-terminales ainsi que dans le domaine globulaire des histones (Hyland et al., 2005). Ces modifications covalentes ont lieu sur des résidus spécifiques pour chaque histone. Au moins huit classes de modifications covalentes différentes ont été caractérisées et plusieurs résidus ont été identifiés pour chaque classe de modification grâce à l'utilisation d'anticorps spécifiques des diverses modifications ou à la technique de spectrométrie de masse (*mass spectrometry*, MS) (Kouzarides, 2007). Plusieurs modifications ont été décrites à ce jour et elles incluent l'acétylation, la méthylation, l'ubiquitinylation, la phosphorylation, l'ADP-ribosylation, la sumoylation, la glycosylation et la biotinylation. Il a été proposé que l'ensemble des modifications post-traductionnelles des histones formerait un code, appelé le «code histone», qui permettrait d'associer à chaque combinaison de modifications un état transcriptionnel de la chromatine (Strahl et Allis, 2000; Jenuwein et Allis, 2001; Turner, 2002). Selon cette hypothèse, le code est établi par des enzymes modifiant la chromatine et il est interprété par des facteurs interagissant spécifiquement avec l'histone modifiée, ce qui entraîne l'expression ou le silence des gènes en réponse à des stimuli extracellulaires (Taverna et al., 2007). De plus, l'existence de plusieurs modifications sur une même histone permet une combinaison de modifications et une régulation croisée de ces modifications, dictant un état particulier de la chromatine et la manière de régler l'expression d'un gène (Latham et Dent, 2007).

Les modifications post-traductionnelles des histones les plus étudiées restent aujourd'hui l'acétylation et la méthylation. L'acétylation consiste en une neutralisation de

la charge positive de la lysine cible, et elle a lieu sur des lysines spécifiques au sein des quatre histones de la particule cœur (Bhaumik et al., 2007). Cette modification est catalysée par des histones acétyltransférases (HATs) via le transfert d'un acétyle à partir de l'acétyl-coenzyme A vers le groupement ϵ -amine des résidus lysines cibles (Lee et Workman, 2007). Plusieurs coactivateurs transcriptionnels, tels Gcn5/PCAF, CBP/p300 and SRC-1, possèdent une activité HAT intrinsèque. L'acétylation des histones a été associée depuis longtemps à l'activation de l'expression génique (Kouzarides, 2007). Quant à la méthylation des histones, cette modification cible les résidus lysines ou arginines (Kouzarides, 2007; Bhaumik et al., 2007). Chez les eucaryotes supérieurs, la méthylation des arginines peut avoir lieu sous la forme mono- ou di-méthylée de façon symétrique ou asymétrique à des positions spécifiques sur les histones H3 et H4. Le coactivateur transcriptionnel CARM1 et la protéine arginine méthyltransférase PRMT1 sont capables de méthyler des résidus arginine spécifiques des histones H3 et H4 (Shilatifard, 2006). Les lysines peuvent être mono-, di- ou tri-méthylées sur les lysines 4, 9, 27, 36 et 79 de l'histone H3 et sur la lysine 20 de l'histone H4. La majorité des histones lysines méthyltransférases caractérisées à date possèdent un domaine SET et ces enzymes peuvent catalyser la méthylation de lysines spécifiques des histones H3 et H4 (Bhaumik et al., 2007). Dot1, une histone lysine méthyltransférase d'une nouvelle classe ne possédant pas de domaine SET, ne méthyle que des histones dans un contexte nucléosomal (Ng et al., 2002; van Leeuwen et al., 2002; Lacoste et al., 2002; Feng et al., 2002). La méthylation des lysines des histones a été liée à l'activation et à la répression transcriptionnelle, dépendamment du site et du degré de la méthylation (Shilatifard, 2006).

Au cours de la dernière décennie, des expériences biochimiques ont permis d'identifier différents domaines protéiques conservés impliqués dans la reconnaissance spécifique des modifications des histones (tableau II) (revu par (Taverna et al., 2007)). Ainsi, les lysines acétylées des histones interagissent spécifiquement avec des protéines à bromodomaine (Dhalluin et al., 1999; Zeng et Zhou, 2002). D'autre part, les histones méthylées sont liées par des facteurs spécifiques possédant des chromodomaines ou des domaines Tudor, MBT, PHD et WD40 (Taverna et al., 2007).

Tableau II. Les domaines protéiques impliqués dans la reconnaissance des modifications des histones

Domaine	Modifications des histones
Bromodomaine	Plusieurs lysines acétylées
Chromodomaine	Lysines 4, 9 et 27 méthylées de l'histone H3
Tudor	Lysine 4 méthylée de l'histone H3 Lysine 20 méthylée de l'histone H4
MBT	Lysines 4 et 9 méthylées de l'histone H3 Lysine 20 méthylée de l'histone H4
PHD	Lysines 4, 9 et 36 méthylées de l'histone H3
WD40	Lysine 4 méthylée de l'histone H3 Arginine 2 méthylée de l'histone H3

La modification des histones est un processus dynamique. En effet, plusieurs des modifications des histones sont enzymatiquement réversibles et les enzymes capables d'enlever ces modifications ont été identifiées. Les histones désacétylases (HDACs) permettent la désacétylation des histones (Bhaumik et al., 2007). De plus, divers corépresseurs transcriptionnels, tels les complexes mSin3a, NCoR/SMRT et NuRD/Mi-2, possèdent des sous-unités avec une activité HDAC (Shahbazian et Grunstein, 2007; Denslow et Wade, 2007). La méthylation des lysines est une modification réversible grâce à deux familles d'enzymes, les amines oxydases telles LSD1 et les hydroxylases de la famille JmjC (Klose et Zhang, 2007; Shi et Whetstine, 2007). Dépendamment des protéines auxquelles LSD1 est associée, cette enzyme peut déméthyliser les lysines 4 et 9 de l'histone H3 (Klose et Zhang, 2007; Shi et Whetstine, 2007). Plusieurs membres de la famille JmjC ont des substrats uniques, et des déméthylases spécifiques pour divers résidus lysines de l'histone H3 ont été décrites (Shi et Whetstine, 2007).

Au cours des dernières années, l'incidence globale des modifications des histones a été déterminée grâce à la technique d'immunoprécipitation de la chromatine (ChIP) en

utilisant des anticorps spécifiques des différentes modifications couplée à la technologie de puces à ADN (ChIP-CHIP) (revu par (Rando, 2007; Millar et Grunstein, 2006)). Ces analyses à l'échelle du génome, effectuées chez la levure, montrent que les modifications ne sont pas uniformément distribuées (Pokholok et al., 2005; Liu et al., 2005). Certaines caractéristiques communes ont été mises en lumière concernant la composition et l'enrichissement des modifications au niveau des gènes activement transcrits. Ainsi, l'acétylation est enrichie à des sites spécifiques au promoteur et dans l'extrémité 5' des régions codantes. Au promoteur, deux nucléosomes entourant le site d'initiation transcriptionnel sont non seulement hypoacétylés sur certaines lysines mais aussi enrichis du variant Hzt1 de l'histone H2A (Liu et al., 2005; Zhang et al., 2005a; Raisner et al., 2005; Millar et al., 2006). Quant au site d'initiation de la transcription comme tel, il est dépourvu de nucléosomes sur une longueur moyenne de 150 paires de bases (Yuan et al., 2005). De plus, la lysine tri-méthylée est enrichie au niveau de la région codante et cette modification corrèle avec le taux de transcription {Pokholok, 2005 6995 /id}. Dans le cas de la lysine 36 de l'histone H3, la tri-méthylation entraîne un environnement hypoacétylé au niveau la région codante (Carrozza et al., 2005; Joshi et Struhl, 2005). Enfin, chacun des sites de méthylation connus chez la levure (lysines 4, 36 et 79 de l'histone H3) possède un patron de distribution spécifique. Ce patron de modifications des histones observé chez la levure semble être aussi conservé chez les mammifères (Bernstein et al., 2005).

1.1.1.4 L'incorporation des variants d'histones

Les variants d'histones, des formes protéiques alternatives des histones principales du nucléosome, sont codés par des gènes différents et ils sont présents chez tous les organismes eucaryotes (Malik et Henikoff, 2003; Kamakaka et Biggins, 2005). À ce jour, des variants ont été identifiés pour les histones H2A, H2B et H3. Tandis que l'expression des histones principales est augmentée lors de la phase S du cycle cellulaire afin qu'elles soient assemblées en chromatine au cours de la réplication de l'ADN, les variants d'histones sont exprimés indépendamment du cycle cellulaire et ils peuvent être

incorporées dans la chromatine de manière indépendante de la synthèse d'ADN (Ahmad et Henikoff, 2002a). L'échange des histones principales par des variants d'histones au sein du nucléosome peut affecter la structure et les fonctions spécifiques de la chromatine (Sarma et Reinberg, 2005; Henikoff et al., 2004). En effet, des résultats récents montrent que les variants d'histones jouent un rôle important dans la modulation de plusieurs processus biologiques nucléaires dont la réparation de l'ADN et la transcription des gènes (revu par (Polo et Almouzni, 2006)).

Les variants les mieux caractérisés sont ceux des histones H2A et H3 (Sarma et Reinberg, 2005; Henikoff et al., 2004; Polo et Almouzni, 2006). Trois variants majeurs de l'histone H2A ont été identifiés : H2A.X, H2A.Z et macroH2A. Le variant H2A.X peut subir différentes modifications post-traductionnelles dont la phosphorylation (Rogakou et al., 1998). La forme phosphorylée de ce variant, appelée γ H2A.X, est présente au niveau des sites de cassure de l'ADN double brin et semble être impliquée dans la réparation de l'ADN. L'histone H2A.Z (nommé Htz1 chez *Saccharomyces cerevisiae*), le variant le plus conservé entre les différentes espèces, joue un rôle essentiel dans le développement embryonnaire précoce chez la souris (Faast et al., 2001). Selon plusieurs études, le variant H2A.Z semble important pour la formation de régions transcriptionnellement actives, alors que d'autres travaux lient fonctionnellement ce variant à la répression transcriptionnelle (revu par (Raisner et Madhani, 2006)). Enfin, macroH2A, un variant spécifique des vertébrés, est présent dans des régions silencieuses du génome, comme le chromosome X inactif chez les femelles de mammifères (Costanzi et Pehrson, 1998). Quatre variants différents de l'histone H3 ont été décrits : H3.1, H3.2, H3.3 et CENP-A (*centromeric protein A*) (Henikoff et al., 2004; Polo et Almouzni, 2006). Les variants H3.1 et H3.2 ne diffèrent que par quelques acides aminés de H3.3. Le variant H3.3 a été localisé dans des régions transcriptionnellement actives (Ahmad et Henikoff, 2002b). Le variant CENP-A a, quant à lui, été retrouvé spécifiquement dans les régions centromériques et ce variant semble jouer un rôle indispensable dans la ségrégation chromosomique (Sullivan, 2001; Dalal et al., 2007).

La cellule doit diriger la localisation des variants d'histones dans la chromatine à des endroits précis où ceux-ci peuvent exercer leurs différents rôles. Des études récentes montrent que plusieurs complexes multi-protéiques sont responsables du dépôt ciblé des variants d'histones au sein du génome (revu par (Jin et al., 2005; De Koning et al., 2007)). En effet, comme mentionné précédemment, le complexe SWR1 chez la levure catalyse le remplacement spécifique de l'histone H2A par son variant Htz1 (l'homologue de l'histone H2A.Z chez l'humain) (Kobor et al., 2004; Krogan et al., 2003b; Mizuguchi et al., 2004). De plus, CAF-1 (*chromatin assembly factor 1*) a été identifié comme le complexe de chaperons permettant le remplacement du variant H3.1 dans le nucléosome (Tagami et al., 2004). Pour sa part, le variant H3.3 est incorporé dans la chromatine grâce au chaperon d'histones HIRA (*histone regulating protein A*) (Tagami et al., 2004; Loppin et al., 2005). Enfin, récemment, Zhang et al. ont montré que les chaperons HIRA et ASF1a sont capables d'incorporer le variant macroH2A au sein de l'hétérochromatine (Zhang et al., 2005b).

1.1.2 Les facteurs liant l'ADN

Un objectif important dans le domaine de la régulation transcriptionnelle consiste à comprendre le mécanisme de transcription dépendant des facteurs liant l'ADN. Chez les eucaryotes, il existe des dizaines de milliers de gènes codant pour des protéines, lesquels possèdent leur propre programme spécifique de contrôle transcriptionnel. La spécificité de ces programmes est effectuée, en grande partie, par des facteurs transcriptionnels qui se fixent sur des séquences régulatrices spécifiques sur l'ADN (revu par (Kadonaga, 2004; Maston et al., 2006)). Ces facteurs contiennent un module de liaison à l'ADN associé à un ou plusieurs modules d'activation ou de répression, ainsi que possiblement un module de multimérisation et un module régulateur (Kadonaga, 2004). Les facteurs liant l'ADN, appelés activateurs et répresseurs transcriptionnels, sont importants pour l'activation et la répression de la transcription. Ainsi, ces facteurs transcriptionnels interprètent l'information génétique codée par la séquence d'ADN et la transmettent à d'autres facteurs, appelés cofacteurs, lesquels interagissent avec la

machinerie transcriptionnelle. En somme, les facteurs liant l'ADN agissent, de façon générale, comme des éléments clé d'interface entre l'information de régulation contenue sur l'ADN et la machinerie de transcription de l'ARN Pol II.

1.1.2.1 Les séquences d'ADN régulatrices

Afin de comprendre les mécanismes moléculaires gouvernant les patrons d'expression génique chez les eucaryotes, il est important d'identifier les éléments de régulation transcriptionnelle associés à chaque gène. Un des objectifs de la recherche génomique consiste à identifier tous les éléments fonctionnels dans le génome, incluant ceux impliqués dans la régulation de l'expression des gènes. Typiquement, les gènes transcrits par l'ARN Pol II contiennent deux familles distinctes de séquences d'ADN régulatrices (Maston et al., 2006). La première, laquelle sera discutée plus loin, consiste du promoteur d'un gène, lui-même formé d'un promoteur basal et d'éléments régulateurs proximaux. Quant à la seconde famille, elle est composée d'éléments de régulation distaux, incluant les *enhancers*, les *silencers*, les *insulators* et les LCRs (*locus control regions*) (figure 2) (Maston et al., 2006). Ces séquences d'ADN régulatrices contiennent des sites de reconnaissance pour des facteurs transcriptionnels liant l'ADN, lesquels agissent afin d'activer ou de réprimer la transcription.

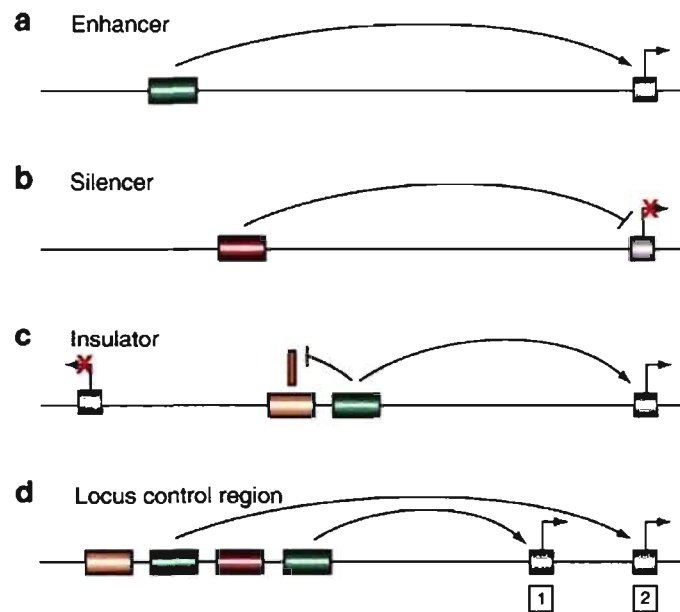


Figure 2 : Éléments de régulation distaux. (a, b) Les *enhancers* et les *silencers* agissent respectivement afin d'activer et de réprimer la transcription. (c) Les *insulators* fonctionnent en isolant les effets des éléments de régulation transcriptionnelle avoisinants sur des gènes en particulier. (d) Les LCRs (*locus control regions*) regroupent plusieurs éléments régulateurs impliqués dans la régulation d'un locus entier. Figure tirée de (Maston et al., 2006).

Les *enhancers* consistent en des séquences régulatrices chez les eucaryotes supérieurs liées par des activateurs transcriptionnels et pouvant activer leurs gènes cibles sur une longue distance (figure 2a) (Blackwood et Kadonaga, 1998). En effet, les *enhancers* peuvent agir jusqu'à plusieurs centaines de kpb en amont ou en aval d'un promoteur, au niveau d'un intron ou encore, au-delà de l'extrémité 3' d'un gène. Des données récentes suggèrent un modèle, le *DNA-looping*, pour expliquer ce mode d'activation à distance des *enhancers*. Selon ce modèle, l'*enhancer* et le promoteur basal d'un gène sont rapprochés suite à la formation d'une boucle ou *loop* de l'ADN qui les sépare (Vilar et Saiz, 2005). Chez la levure, ces séquences régulatrices sont appelées UAS (*upstream activation sequences*) et elles ne fonctionnent qu'en amont des gènes, à quelques centaines de paires de bases en 5' du promoteur génique (Guarente, 1988).

De leur part, les *silencers* consistent en des séquences régulatrices conférant un effet négatif (répression) sur la transcription d'un gène cible et ils sont liés par des répresseurs transcriptionnels (figure 2b). Comme les *enhancers* de mammifères, les *silencers* influencent la transcription indépendamment de leur orientation et de la distance les séparant du promoteur basal (Maston et al., 2006).

Les *insulators* sont des séquences d'ADN capables d'isoler les effets des *enhancers* et des *silencers* sur des gènes en particulier, empêchant ces séquences régulatrices de moduler la transcription génique de manière inappropriée (figure 2c) (Wallace et Felsenfeld, 2007; Capelson et Corces, 2004). Les *insulators* ont été divisés en deux groupes. Le premier groupe rassemble les agents bloquant les *enhancers*, empêchant ainsi les *enhancers* d'interagir avec le promoteur d'un gène cible. Le second groupe d'*insulators*, quant à lui, est capable d'engendrer une barrière contre la propagation de l'hétérochromatine.

Les LCRs regroupent plusieurs éléments régulateurs impliqués dans la régulation d'un locus entier, incluant des *enhancers*, des *silencers* et des *insulators* (figure 2d) (revu par (Li et al., 2002; Dean, 2006)). Ces éléments sont liés par des facteurs transcriptionnels liant l'ADN (activateurs et répresseurs), des coactivateurs et des facteurs modifiant la chromatine (Maston et al., 2006). Les LCRs peuvent réguler l'expression des gènes à longue distance, indépendamment de leur position par rapport aux gènes cibles. Bien que différents modèles ont été proposés afin d'expliquer le mode de régulation des LCRs, le modèle de *DNA-looping* observé également chez les *enhancers*, tel que discuté auparavant, a été retenu par plusieurs chercheurs à la lumière de récentes études (Cook, 2003; Maston et al., 2006). Ainsi, les contacts physiques formant une boucle d'ADN résultent dans le rapprochement de séquences d'ADN régulatrices actives, jouant un rôle de premier ordre dans l'établissement d'une chromatine plus ouverte (Tolhuis et al., 2002). Ces interactions à longue distance ne sont observées que lorsque un locus est transcriptionnellement actif, suggérant que les LCRs jouent un rôle dans l'activation de la transcription (Maston et al., 2006).

1.1.2.2 Les activateurs transcriptionnels

Les activateurs sont des protéines liant l'ADN sur des séquences spécifiques localisées généralement en amont du promoteur basal (revu par (Ptashne et Gann, 1997)). Plusieurs classes d'activateurs ont été distinguées sur la base de leurs domaines de liaison à l'ADN. Il existe plusieurs modules de liaison à l'ADN, incluant le motif hélice-tour-hélice, l'homéodomaine, le doigt à zinc, la fermeture-éclair à leucine, l'hélice-boucle-hélice, ETS ainsi que les domaines HMG1 et Pit-Oct-Unc (POU) (Pabo et Sauer, 1992). À l'inverse des domaines de liaison à l'ADN bien définis, les modules d'activation transcriptionnelle sont structurellement plus élusifs. Ainsi, certains activateurs contiennent une hélice alpha amphipatique avec une face chargée négativement pouvant agir comme un module d'activation transcriptionnelle (Ptashne, 1988). En outre, d'autres motifs ont été trouvés dans les régions d'activation, comme des feuillets bêta hydrophobiques et des motifs riches en acides aminés glutamine ainsi qu'en acides aminés proline (Kadonaga, 2004). Les domaines d'activation établissent des contacts directs protéine-protéine avec les facteurs associés à l'ARN Pol II et ils semblent stimuler la transcription en guidant le recrutement de l'enzyme aux régions du promoteur.

Vers le milieu des années 80, plusieurs groupes de recherche ont utilisé la technique de chromatographie d'affinité afin de purifier les facteurs liant l'ADN, incluant SP1, USF, CTF/NF-I, HSF, AP-1, AP-2, Oct-2 et CREB (Kadonaga, 2004). D'autre part, grâce à des méthodes immunologiques, des récepteurs nucléaires, comme par exemple les récepteurs aux hormones stéroïdes (le récepteur aux glucocorticoïdes et le récepteur des estrogènes), ont aussi été identifiés comme des facteurs liant des séquences spécifiques sur l'ADN. Ces divers facteurs appartiennent à de grandes familles multi-protéiques. À l'intérieur de chaque famille, les membres possèdent relativement les mêmes propriétés de liaison à l'ADN mais des fonctions d'activation distinctes. Les sites de liaison des activateurs consistent en des courtes (6-12 paires de bases) séquences d'ADN (Maston et al., 2006). Les séquences d'ADN régulatrices telles la boîte CCAAT, la boîte GC, l'élément du choc thermique et l'élément de réponse des glucocorticoïdes ont été identifiées comme des régions liées par des activateurs transcriptionnels (Kadonaga,

2004). De plus, plusieurs activateurs forment des hétérodimères et/ou des homodimères, et par conséquent, leurs sites de liaison sont généralement composés de deux moitiés de sites. Des études montrent que l'orientation de ces moitiés de sites de liaison, ainsi que l'espace les séparant, jouent un rôle majeur en modulant l'action régulatrice des dimères d'activateurs liés sur ces séquences (Claessens et Gewirth, 2004).

Les modes de stimulation de la synthèse d'ARNm par les activateurs transcriptionnels sont basés sur trois mécanismes fondamentaux (revu par (Kadonaga, 2004; Maston et al., 2006)). Tout d'abord, les activateurs peuvent stimuler le recrutement de la machinerie transcriptionnelle basale, notamment l'ARN Pol II et les facteurs généraux de transcription, au promoteur basal en interagissant directement avec les composants de cette machinerie (Orphanides et al., 1996). Dans le second, selon un mécanisme de « post-recrutement », les activateurs stimulent l'activité de la machinerie de transcription basale dans une étape ultérieure à son association avec le promoteur pour former le complexe de pré-initiation, notamment l'initiation, l'élongation ou la ré-initiation transcriptionnelle (Lee et Young, 2000). Dans le troisième mécanisme, les activateurs peuvent modifier la matrice d'ADN en recrutant des facteurs capables de modifier la chromatine au promoteur d'un gène (ces facteurs ont été décrits plus haut) (Lemon et Tjian, 2000). Les altérations locales résultantes peuvent stimuler la formation du complexe de pré-initiation, ainsi que l'étape de post-recrutement de la machinerie transcriptionnelle de l'ARN Pol II. D'autre part, les activateurs peuvent stimuler la transcription de manière synergétique et cet effet découle de la coopération entre plusieurs copies du même facteur ou de la coopération entre différents facteurs (Maston et al., 2006). Le mécanisme expliquant ce phénomène de synergie transcriptionnelle des activateurs demeure vague (Green, 2005). Une des possibilités est que chaque activateur interagisse et recrute simultanément différents facteurs généraux de transcription ou des cofacteurs. Une autre possibilité est que différents activateurs peuvent avoir des fonctions distinctes : certains peuvent modifier la structure de la chromatine, tandis que d'autres peuvent réguler différentes étapes de la réaction transcriptionnelle, telles le dégagement du promoteur ou l'élongation (Maston et al., 2006).

La fonction des activateurs peut être modulée par un groupe de facteurs essentiels à la transcription chez les eucaryotes : les coactivateurs (revu par (Lonard et O'Malley, 2005; Spiegelman et Heinrich, 2004)). Typiquement, les coactivateurs ne possèdent pas la capacité de lier des séquences spécifiques sur l'ADN. Par conséquent, les coactivateurs sont recrutés par des activateurs liés sur l'ADN via des interactions protéine-protéine. De la même manière que les activateurs transcriptionnels, les coactivateurs agissent en stimulant la formation du complexe de pré-initiation transcriptionnel ou en modifiant la chromatine (Lonard et O'Malley, 2005; Spiegelman et Heinrich, 2004). En effet, certains coactivateurs possèdent une activité histone acétyltransférase, tels SRC-1, p300 et CBP. D'autres coactivateurs ont une fonction histone méthyltransférase, comme par exemple CARM1 et PRMT1. Enfin, le coactivateur Brg1 peut remodeler la chromatine. En modulant la capacité des activateurs de réguler positivement ou négativement la transcription, les coactivateurs jouent un rôle important dans la réponse régulatrice de l'expression génique (Lemon et Tjian, 2000).

1.1.2.3 Les répresseurs transcriptionnels

Tous les gènes sont maintenus dans un état de répression transcriptionnelle par des facteurs appelés répresseurs (revu par (Gaston et Jayaraman, 2003)). Certains répresseurs, comme par exemple le facteur de doigt-de-zinc Krüppel et la protéine homéodomaine *Engrailed*, lient l'ADN sur des séquences spécifiques. D'autres répresseurs, comme les protéines MeCP2 et MBD2, peuvent lier l'ADN méthylé. Les répresseurs NCoR, Rb et MDM2 forment, quant à eux, des interactions protéine-protéine avec des activateurs transcriptionnels, bloquant l'activité de ces derniers. Enfin, un autre groupe de répresseurs séquestre des activateurs, coactivateurs ou des composants de la machinerie transcriptionnelle dans des complexes inactifs. Des exemples de ce groupe de répresseurs incluent IκB, Mot1 et la protéine d'adénovirus E1A243R (Gaston et Jayaraman, 2003).

Les répresseurs sont divisés en deux catégories selon le degré de répression génique qu'ils exercent : une répression de longue distance ou de courte distance (Maston et al., 2006). Dans le cas de la répression de longue distance, un répresseur module le promoteur afin que celui-ci contrecarre l'influence de tous les *enhancers*, même si ces derniers sont localisés à des centaines de paires de bases du site de liaison sur l'ADN du répresseur. Ce type de répression est souvent appelé *silencing* transcriptionnel car un locus chromosomique entier devient inactivé. En revanche, les répresseurs de courte distance ne modulent pas la transcription de tout un locus mais ils bloquent plutôt la fonction des activateurs avoisinants, sans interférer avec les activateurs plus distants. Chez la Drosophile, les répresseurs de courte distance résident à environ 100 paires de bases de leur gène cible afin d'exercer leur effet répressif (Maston et al., 2006). La présence de ces deux types de répresseurs dans la cellule apporte une flexibilité additionnelle à la régulation génique (Courey et Jia, 2001). Ainsi, la répression de longue-distance permet de réprimer tout un locus indépendamment du nombre de modules régulateurs qui peuvent contrôler l'activité de ce locus. De son côté, la répression de courte-distance permet de moduler l'activité d'un *enhancer* sans interférer avec l'activité des autres *enhancers*. Cette autonomie des *enhancers* semble être surtout importante au niveau des loci complexes contenant de multiples *enhancers*, où chacun est requis pour un patron d'expression génique particulier.

Divers modèles ont été proposés pour expliquer la fonction des répresseurs. Dans certains cas, les répresseurs semblent agir en bloquant la liaison d'un activateur avoisinant ou en étant en compétition directe pour le même site de liaison sur l'ADN (Harris et al., 2005; Li et al., 2004a). Alternativement, un répresseur peut empêcher des activateurs et/ou des facteurs généraux de transcription d'accéder au promoteur basal en engendrant une structure de la chromatine fermée via le recrutement de facteurs modifiant la chromatine (Srinivasan et Atchison, 2004). Enfin, il a été récemment suggéré qu'un répresseur peut bloquer l'activation transcriptionnelle en inhibant la formation du complexe de pré-initiation transcriptionnel (Chen et Widom, 2005). La capacité des répresseurs à fonctionner par de multiples mécanismes semble être primordiale à la régulation génique chez les eucaryotes (Courey et Jia, 2001). En effet,

cette multifonctionnalité peut permettre aux répresseurs de moduler l'expression d'un gène en particulier selon les besoins de la cellule. Par exemple, si le but est de réprimer la transcription de manière transitoire en réponse à un changement temporaire dans l'environnement, la répression via des interactions avec la machinerie transcriptionnelle, lesquelles devraient être rapidement réversibles, pourrait s'avérer l'option préférable. Au contraire, si le but est de générer un état répressif épigénétique, la répression via des modifications covalentes dans la structure de la chromatine pourrait être l'option appropriée. En effet, la redistribution semi-conservative des histones pendant la phase S permet la conservation de cette information épigénétique au fil des divisions cellulaires (Groth et al., 2007).

Les répresseurs peuvent recruter des cofacteurs négatifs, aussi appelés corépresseurs (Privalsky, 2004). Les corépresseurs exercent un effet opposé aux coactivateurs sur la chromatine, la rendant inaccessible à la liaison de facteurs transcriptionnels. Un grand nombre de corépresseurs possède une activité de désacétylation des histones (Struhl, 1998). Un des corépresseurs le plus étudié ayant une activité de désacétylation des histones est HDAC1. Ce facteur est souvent retrouvé dans plusieurs complexes multi-protéiques, incluant les complexes Sin3 et NuRD (Knoepfler et Eisenman, 1999). Les facteurs Dnmt3, MBD3, Tup1 et Groucho constituent d'autres exemples de corépresseurs (Gaston et Jayaraman, 2003).

1.1.2.4 Le complexe Médiateur

Le complexe Médiateur joue un rôle crucial dans l'activation et la répression de la synthèse de l'ARNm chez les eucaryotes. Plusieurs études révèlent que le Médiateur s'associe avec l'ARN Pol II et il est recruté aux *enhancers* ou aux *silencers* transcriptionnels par des activateurs ou des répresseurs *in vivo* (Kornberg, 2005; Bjorklund et Gustafsson, 2005; Malik et Roeder, 2005). Le Médiateur de *S. cerevisiae*, le premier à être identifié par l'équipe de Kornberg, est un complexe de haut poids moléculaire composé de plus de 20 sous-unités (Kim et al., 1994). Ces sous-unités sont

distribuées en trois sous-domaines distincts, nommés les modules « tête », « milieu » et « queue » (Conaway et al., 2005). Un autre module, lequel contient une kinase et une cycline, est associé à une partie des complexes Médiateur. Dans les cellules de levure en croissance exponentielle, ce module kinase-cycline a été impliqué dans la répression d'un groupe de gènes (Liao et al., 1995; Hengartner et al., 1998; Holstege et al., 1998; Carlson, 1997). Des études récentes ont permis de définir également la composition et les activités du Médiateur de mammifère, révélant que la structure et la fonction du complexe Médiateur sont hautement conservées de la levure à l'homme (Boube et al., 2002).

Chez les mammifères, divers complexes Médiateurs ont été isolés grâce à une variété de méthodes, incluant la chromatographie conventionnelle et la chromatographie d'affinité. Le premier complexe Médiateur de mammifère a été purifié par le groupe de Roeder. Ce complexe a été désigné le complexe TRAP (*thyroid hormone receptor-associated proteins*) car il a été isolé en association avec le récepteur de l'hormone thyroïde (Fondell et al., 1996; Fondell et al., 1999). Par la suite, d'autres complexes Médiateurs similaires au complexe TRAP ont été purifiés par plusieurs laboratoires à partir de cellules B de souris, de cellules HeLa humaines et de foie de rat (Jiang et al., 1998; Ryu et al., 1999; Ryu et Tjian, 1999; Malik et al., 2000; Brower et al., 2002; Gu et al., 1999; Sun et al., 1998; Naar et al., 1999; Boyer et al., 1999). Ces chercheurs ont nommé les différents complexes CRSP (*cofactor required for Sp1 activation*), PC2 (*positive cofactor 2*), DRIP (*vitamin D receptor-interacting proteins*), SMCC [*suppressor of RNA polymerase B (SRB)-mediator-containing cofactor*], NAT (*negative regulator of activated transcription*), SREBP (*sterol-response-element-binding protein*) ou encore, ARC (*activator-recruited factor*) (revu par (Conaway et al., 2005)). Plusieurs de ces complexes Médiateurs, incluant les complexes TRAP, SMCC, PC2, CRSP et DRIP, semblent stimuler la transcription dépendante des activateurs *in vitro*. D'autre part, les complexes TRAP et NAT ont été montrés capables de réprimer la transcription *in vitro* (Gu et al., 1999; Sun et al., 1998). Globalement, les divers complexes Médiateurs de mammifère purifiés par plusieurs laboratoires en utilisant différentes procédures de purification semblent posséder une composition en polypeptides distincte, bien que certaines sous-unités soient communes. Ces différences suggéreraient l'existence de

formes distinctes fonctionnelles du Médiateur dans les diverses cellules. Cependant, les résultats d'une analyse protéomique récente montrent qu'une partie des différences retrouvées entre les divers complexes Médiateur chez les mammifères découle de l'utilisation de procédures d'identification des protéines peu sensibles et/ou de la perte de certaines sous-unités lors de l'étape de purification (Sato et al., 2004). Dans ces expériences de Sato *et al.*, MudPIT (*multidimensional protein identification technology*), une méthode protéomique sensible qui couple la chromatographie 2D et la MS en tandem, a été utilisée pour comparer la composition de complexes purifiés par chromatographie d'immunoaffinité à partir de cellules HeLa ou de lignées cellulaires HEK293 exprimant de façon stable différentes sous-unités du Médiateur (Sato et al., 2004). L'étude de MudPIT a révélé que 37 protéines étaient communes à tous les complexes Médiateur déjà connus.

Les mécanismes par lesquels le complexe Médiateur contrôle la synthèse de l'ARNm n'ont pas encore été fermement établis. Toutefois, des évidences importantes montrent que le Médiateur active la transcription au moins en partie via des interactions directes avec (i) des activateurs transcriptionnels liant l'ADN au sein d'éléments en amont du promoteur et aux *enhancers*; (ii) l'ARN Pol II; et probablement, (iii) un ou plusieurs facteurs généraux de transcription associés au promoteur basal (Conaway et al., 2005). En effet, le complexe Médiateur est considéré un élément clé dans l'activation transcriptionnelle et représente possiblement la plus grande cible des activateurs. Plusieurs activateurs interagissent directement avec le complexe Médiateur, mais la sous-unité spécifique qui est contactée dépend de l'activateur. D'autre part, le Médiateur stimule la transcription basale *in vitro* et plusieurs sous-unités du Médiateur sont essentielles pour la transcription par l'ARN Pol II chez la levure (Thompson et Young, 1995; Holstege et al., 1998). De plus, le Médiateur s'associe avec l'ARN Pol II pour générer un complexe stable, parfois appelé holoenzyme de l'ARN Pol II (Malik et Roeder, 2000; Myers et Kornberg, 2000). À la lumière de ces données, le Médiateur est considéré comme un composant général et essentiel de la machinerie de l'ARN Pol II *in vivo*, jouant un rôle central pour la traduction des signaux d'activation envoyés par les activateurs liés sur les *enhancers* vers les facteurs généraux de transcription. Toutefois, le

Médiateur n'est pas nécessaire pour la transcription à partir d'un promoteur basal *in vitro*, et par conséquent, il n'est pas classifié comme un facteur général d'initiation (Struhl, 2005). De plus, des expériences de CHIP chez la levure ont révélé que le Médiateur n'est pas associé aux promoteurs de l'ARN Pol II hautement actifs (Fan et al., 2006). Plus précisément, ces auteurs ont montré que le Médiateur est recruté par certains activateurs impliqués dans la réponse au stress, mais pas par la majorité des activateurs jouant un rôle dans des conditions optimales de croissance. Les études montrent que le Médiateur est impliqué dans l'étape de recrutement de l'ARN Pol II, possiblement via la stabilisation d'un complexe intermédiaire de ré-initiation dépendant des activateurs, permettant ainsi que de multiples tours de transcription aient lieu à partir du complexe de pré-initiation (Struhl, 1996; Yudkovsky et al., 2000). Chez la levure, le Médiateur est nécessaire pour la formation et la stabilité des complexes de pré-initiation qui sont dépendants d'un activateur fonctionnel (Kuras et Struhl, 1999; Li et al., 1999b). De plus, *in vivo*, l'association du Médiateur à des promoteurs peut avoir lieu indépendamment et avant l'association de l'ARN Pol II (Cosma et al., 2001; Park et al., 2001). Par conséquent, le Médiateur semble servir de pont permettant de lier des activateurs associés à des régions spécifiques sur le génome avec la fonction de la machinerie basale de l'ARN Pol II au promoteur.

1.1.3 Les ARN non-codants

De multiples molécules d'ARN sont impliquées dans la catalyse et la régulation de plusieurs processus biologiques fondamentaux (Storz et al., 2005; Hannon et al., 2006; Mattick et Makunin, 2006). Généralement, ces ARN possédant des fonctions régulatrices ne codent pas pour des protéines et ils sont connus sous le nom d'ARN non-codants (ncARN) (Huttenhofer et al., 2005). Les ncARN sont transcrits par l'une des trois ARN Pol nucléaires (ARN Pol I, II ou III). Trois mécanismes principaux caractérisent le mode d'action des ncARN: l'alignement spécifique à un acide nucléique cible; la catalyse de réactions biologiques; ou enfin, la liaison à une protéine, suivie de la modulation de son activité (Goodrich et Kugel, 2006; Mattick et Makunin, 2006).

Les ncARN participent à des étapes clé de l'expression des gènes. En particulier, certains petits ARN nucléaires (snARN) interviennent dans l'épissage des ARN pré-messagers, les ARN ribosomiaux (ARNr) et les ARN de transfert (ARNt) catalysent la synthèse protéique, les micro-ARN antisens sont impliqués dans la stabilité des transcrits ou encore dans la répression de la synthèse de protéines spécifiques, et les petits ARN nucléolaires (snoARN) agissent comme des guides pour modifier des nucléotides sur les autres ARN (Storz et al., 2005). De plus, plusieurs ncARN peuvent moduler la transcription. C'est le cas de l'ARN *Xist* qui inactive la transcription de l'un des chromosomes X chez les souris femelles (Wutz et Gribnau, 2007; Ng et al., 2007). *Xist* accumule au niveau du chromosome X à être inactivé et initie la répression transcriptionnelle en recrutant des protéines PcG (*Polycomb Group*) responsables d'induire des modifications de la chromatine impliquées dans l'inactivation génique, telles la tri-méthylation de la lysine 27 de l'histone H3 et l'ubiquitinylation de la lysine 119 de l'histone H2A. D'autres ncARN ciblent directement la machinerie générale de transcription. Par exemple, le snARN 6S s'associe à l'ARN Pol bactérienne, l'empêchant de se lier au promoteur génique et l'inactivant (Wassarman, 2007). Chez les mammifères, une série de travaux récents montrent que les snARN 7SK et U1 régulent la fonction de deux facteurs généraux de transcription, P-TEFb (*positive transcription elongation factor b*) et TFIIH, tandis que B2 est un ncARN murin modulant l'activité de l'ARN Pol II (figure 3) (revu par (Goodrich et Kugel, 2006)).

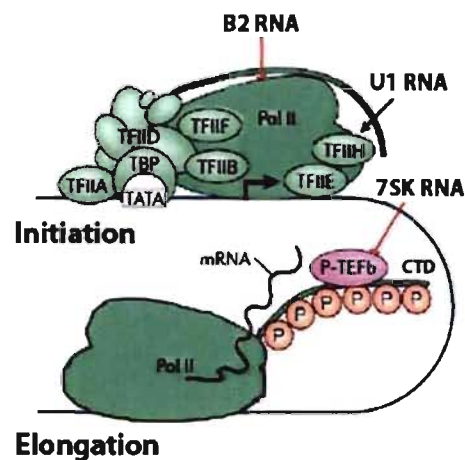


Figure 3 : Régulation de la réaction transcriptionnelle de l'ARN Pol II par les petits ARN nucléaires (snARN) 7SK, U1 et B2. L'étape d'initiation transcriptionnelle est ciblée par les snARN U1 et B2, qui régulent l'activité du facteur de transcription TFIIH et de l'ARN Pol II. Le

snARN 7SK, quant à lui, affecte l'étape d'élongation transcriptionnelle en modulant la fonction du facteur d'élongation P-TEFb. La flèche noire indique un effet stimulateur ; la ligne rouge indique un effet inhibiteur. CTD, domaine C-terminal de l'ARN Pol II ; P, groupement phosphorylé. Figure modifiée de (Goodrich et Kugel, 2006).

1.1.3.1 Le snARN 7SK

Transcrit par l'ARN Pol III, 7SK est un abondant (2×10^5 copies/cellule) snARN de 331 nucléotides hautement conservé chez les vertébrés (Egloff et al., 2006; Murphy et al., 1987; Zieve et al., 1977; Wassarman et Steitz, 1991; Ullu et al., 1982). Une caractéristique de certains transcrits de l'ARN Pol III, tel que 7SK, est la présence d'une structure coiffe particulière, soit un monométhyle sur le phosphate gamma à leur extrémité 5' (Shumyatsky et al., 1990; Singh et Reddy, 1989; Gupta et al., 1990). Dans les cellules HeLa, 7SK a été trouvé en complexe avec diverses protéines, formant une particule de 12S (Ullu et al., 1982). Les résultats obtenus par Wassarman *et al.* ont permis d'élucider la composition de cette ribonucléoprotéique (snRNP) de 7SK, notamment huit protéines de 83 à 13 kDa (Wassarman et Steitz, 1991).

Malgré une première description de ce snARN dans les années 1970, le rôle cellulaire de 7SK est resté énigmatique pendant plus de deux décennies (Zieve et al., 1977). En 2001, les équipes de Q. Zhou et O. Bensaude ont purifié le snARN 7SK en complexe avec le facteur d'élongation transcriptionnelle P-TEFb impliqué dans la phosphorylation du domaine carboxy-terminal (CTD, *Carboxy-terminal domain*) de l'ARN Pol II et ils ont révélé que 7SK inhibe l'activité kinase de ce facteur (figure 3) (Nguyen et al., 2001; Yang et al., 2001). En effet, leurs travaux ont montré que P-TEFb, composé des sous-unités CDK9 (*cyclin dependant kinase*) et sa cycline T1/T2, est présent dans les cellules humaines sous deux formes de complexes. Le petit complexe contenant P-TEFb libre, possède une activité kinase active tandis que le deuxième complexe, de grande taille, est catalytiquement inactif et il comprend un P-TEFb associé au snARN 7SK (Nguyen et al., 2001; Yang et al., 2001). La protéine Brd4 (*bromodomain protein 4*) interagit avec P-TEFb libre, le recrutant aux gènes cibles (Jang et al., 2005; Yang et al., 2005).

Des études ont montré que 7SK s'associe à la protéine HEXIM1 (*hexamethylene bisacetamide-induced protein-1*) ou à son homologue HEXIM2, la convertissant en un inhibiteur de P-TEFb (Barboric et al., 2005; Blazek et al., 2005; Michels et al., 2004; Yik et al., 2003; Byers et al., 2005; Li et al., 2005; Yik et al., 2005). Des essais deux-hybrides dans la levure et des expériences de chromatographie d'affinité (*pull-down*) *in vitro* ont permis de révéler que le domaine C-terminal de HEXIM1 peut interagir directement avec la sous-unité Cycline T1, montrant que cette interaction est responsable de l'inactivation de P-TEFb (Michels et al., 2003; Michels et al., 2004). Par contre, la protéine HEXIM1 ne peut interagir avec P-TEFb et, par conséquent, inhiber son activité kinase, qu'en présence du snARN 7SK, suggérant que 7SK joue un rôle important d'échafaudage dans l'assemblage du complexe inactif 7SK/HEXIM1/P-TEFb (Michels et al., 2004; Yik et al., 2004). D'autre part, HEXIM1 peut se fixer spécifiquement sur 7SK par une séquence de 18 acides aminés basiques caractéristique du motif de reconnaissance de l'ARN riche en résidu arginine retrouvée dans divers facteurs liant l'ARN (Michels et al., 2004; Yik et al., 2003; Weiss et Narayana, 1998). Finalement, des expériences de reconstitution *in vitro* ont montré que P-TEFb peut également interagir avec le snARN 7SK (Yik et al., 2003; Yik et al., 2004). Cependant, une phosphorylation de la sous-unité CDK9, possiblement sur le résidu conservé thréonine 186, est nécessaire pour l'association *in vitro* de P-TEFb avec 7SK (Chen et al., 2004; Li et al., 2005). Les travaux de l'équipe de T. Kiss ont récemment montré que deux éléments structuraux distincts de liaison protéique localisés dans les structures d'épingles à cheveux en 5' et en 3' de 7SK supportent le recrutement *in vivo* de HEXIM1 et de P-TEFb (Egloff et al., 2006).

Des conditions de stress cellulaire variées conduisant à un blocage de la transcription, notamment le traitement aux rayons ultraviolets (UV) et à l'actinomycine D (ActD) ou encore, des inhibiteurs de P-TEFb, tels le 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB) et le flavopiridol, déclenchent une dissociation du complexe 7SK/HEXIM1/P-TEFb, permettant le relâchement de P-TEFb et, par conséquent, la stimulation de l'élongation transcriptionnelle par ce facteur (Yik et al., 2003; Nguyen et al., 2001). Ces résultats suggèrent une régulation étroite de la transcription par l'interaction entre 7SK et P-TEFb. Des études très récentes montrent que lorsque 7SK

n'est plus associé avec P-TEFb, ce petit ARN nucléaire peut se lier à des protéines pour former d'autres snRNP (Barrandon et al., 2007; Van Herreweghe et al., 2007). Ainsi, dans les cellules HeLa, la fraction de 7SK qui n'est pas impliquée dans la formation du complexe 7SK/HEXIM1/P-TEFb interagit spécifiquement avec l'ARN hélicase A (RHA) et les hnRNPs A1, A2/B1, R et Q, des facteurs connus pour être impliqués dans le processus d'épissage (figure 4). De plus, l'inhibition de la transcription induit le désassemblage du complexe 7SK/HEXIM1/P-TEFb et une augmentation des snRNP 7SK contenant les protéines RHA, hnRNP A1, A2/B1, R et Q (7SK/hnRNPs). Par contre, quand les inhibiteurs transcriptionnels sont retirés du milieu de culture cellulaire, les niveaux initiaux des complexes 7SK/HEXIM1/P-TEFb et 7SK/hnRNPs sont rétablis. À la lumière de ces résultats, les chercheurs ont proposé un modèle selon lequel le niveau nucléaire de P-TEFb actif est contrôlé par la réorganisation dynamique et réversible des snRNP 7SK (figure 4) (Barrandon et al., 2007; Van Herreweghe et al., 2007).

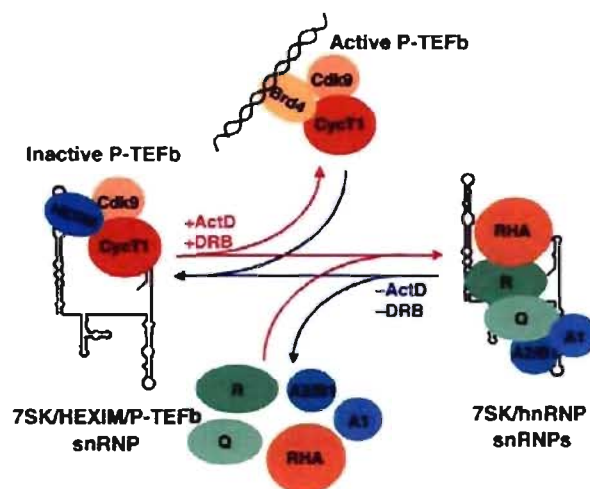


Figure 4 : Modèle de régulation du niveau nucléaire de P-TEFb actif par la réorganisation dynamique et réversible des snRNP contenant 7SK. Dans les cellules HeLa, environ 50% de P-TEFb est séquestré dans la snRNP 7SK/HEXIM1/P-TEFb et devient catalytiquement inactif, tandis que l'autre moitié s'associe avec la protéine Brd4 (*bromodomain protein 4*) capable de le recruter P-TEFb actif aux gènes cibles. L'inhibition de la transcription par l'actinomycine D (ActD) ou le traitement avec 5,6-dichloro-1-β-D-ribofuranosyl-benzimidazole (DRB) induit la dissociation de P-TEFb et HEXIM1 du snARN 7SK et, simultanément, facilite la liaison du 7SK avec l'ARN hélicase A (RHA) et les hnRNPs A1, A2/B1, R et Q, formant des snRNP 7SK/hnRNP. Ce processus est réversible lorsque les inhibiteurs de la transcription sont retirés. Figure tirée de (Van Herreweghe et al., 2007).

1.1.3.2 Le snARN U1

En cherchant des ncARN s'associant avec des facteurs généraux de transcription, Kwek *et al.* ont co-purifié un ARN d'environ 160 nucléotides avec le facteur transcriptionnel TFIIH (Kwek *et al.*, 2002). Tel qu'il sera discuté plus loin, le complexe multi-protéique TFIIH est impliqué, entre autres, dans l'étape d'initiation transcriptionnelle (Yudkovsky *et al.*, 2000; Feaver *et al.*, 1994; Coin *et al.*, 1998). Le séquençage de la bande co-purifiée avec TFIIH a révélé qu'il s'agissait du snARN U1, un petit ARN nucléaire transcrit par l'ARN Pol II (Kwek *et al.*, 2002). Le snARN U1 est un composant important du spliceosome et il joue un rôle essentiel dans l'épissage alternatif (Black, 2003). En effet, U1 participe dans la reconnaissance du site consensus donneur (5') d'épissage (Proudfoot *et al.*, 2002). De plus, le snARN U1 est impliqué dans le couplage entre l'épissage et l'initiation transcriptionnelle ainsi que dans le couplage entre l'épissage et la polyadénylation *in vivo* (Proudfoot *et al.*, 2002; Will *et al.*, 2001). Dans les cellules, le snARN U1 existe associé à divers facteurs, tels U1-70K, U1-A, U1-C, FBP11 et huit protéines Sm, formant la snRNP U1 {Jurica, 2003 5250 /id}. De manière intéressante, aucune protéine de la particule snRNP U1 n'a pu être retrouvée dans le complexe snARN U1/TFIIH identifié par l'équipe de A. Akoulitchev par des expériences de chromatographie d'affinité (*pull-down*) en utilisant un oligo d'ADN complémentaire à la séquence du snARN U1 (Kwek *et al.*, 2002).

Des expériences biochimiques additionnelles ont montré que le snARN U1 lie directement et de manière spécifique la sous-unité cycline H de TFIIH (Kwek *et al.*, 2002). Afin d'identifier les régions du snARN U1 interagissant avec la protéine cycline H, des essais de pontage aux rayons UV suivis d'une digestion tryptique et des analyses de MS ont été utilisés (O'Gorman *et al.*, 2005). Le snARN U1 peut stimuler non seulement l'initiation transcriptionnelle de l'ARN Pol II *in vitro*, mais également l'activité kinase de TFIIH, lequel phosphoryle le CTD de l'ARN Pol II (figure 3) (Kwek *et al.*, 2002). Cependant, les mécanismes moléculaires par lesquels U1 régule la transcription dans les cellules restent à déterminer. Il existe plusieurs possibilités (Manley, 2002). D'une part, le snARN U1 pourrait moduler l'activité kinase de TFIIH

via son interaction avec la cycline H. D'autre part, U1 pourrait jouer un rôle structural, en aidant à maintenir l'intégrité du complexe TFIIH. Enfin, Kwek *et al.* ont proposé un rôle possible de U1 comme régulateur général de TFIIH au niveau de tous les promoteurs, en liant la réaction de transcription à l'étape d'épissage (Kwek *et al.*, 2002).

1.1.3.3 Le ncARN B2

L'ARN B2 murin est un ncARN originant d'une famille de courts (150-300 pb) éléments d'ADN répétés en tandem appelés SINES (*short interspersed repetitive elements*), lesquels sont abondamment dispersés dans le génome de souris (Kramerov et Vassetzky, 2005). Le nombre de copies des SINES est supérieur à 10^5 copies par génome chez les animaux multicellulaires et les plantes (Weiner, 2002). L'abondance élevée des SINES au sein des génomes et leur grande hétérogénéité a amené les chercheurs à considérer ces éléments comme des parasites moléculaires, ne possédant pas de fonction dans la cellule. Toutefois, des éléments typiques du promoteur de l'ARN Pol III ont été retrouvés à l'intérieur des SINES. En effet, les SINES B2 sont transcrits par l'ARN Pol III, engendrant des ARN B2 de 180 nucléotides (Kramerov et Vassetzky, 2005). Les études montrent que le niveau cellulaire du ncARN B2 est augmenté lors du développement, dans les cellules cancéreuses, suite à une infection virale et en réponse à certains stress, tels l'irradiation avec les rayons UV et le traitement à la cycloheximide (Bachvarova, 1988; White *et al.*, 1989; Kramerov *et al.*, 1990; Rudin et Thompson, 2001; Singh *et al.*, 1985; Liu *et al.*, 1995). De plus, le niveau de transcription de B2 par l'ARN Pol III augmente jusqu'à 100 fois en réponse au choc thermique. En somme, toutes ces observations suggèrent que la synthèse de ce petit ARN est régulée et que, possiblement, il peut jouer un rôle dans la cellule (Fornace, Jr. et Mitchell, 1986; Li *et al.*, 1999a; Liu *et al.*, 1995).

En 2002, les travaux de Kugel et Goodrich ont montré que l'ARN Pol II pouvait être régulée suite à la liaison d'un ARN synthétique (Kugel et Goodrich, 2002). Ces données ont conduit les auteurs à prédire que des ARN endogènes pouvaient agir de

façon similaire à celle de l'oligonucléotide synthétique. De plus, sachant que la transcription de plusieurs gènes par l'ARN Pol II est réprimée en réponse au choc thermique (Findly et Pederson, 1981; Gilmour et Lis, 1985; O'Brien et Lis, 1993; Sonna et al., 2002), les équipes de Kugel et Goodrich ont émis l'hypothèse que, dans les cellules de souris, l'augmentation de la synthèse de l'ARN B2 et l'inhibition de l'ARN Pol II sont corrélées (Allen et al., 2004; Espinoza et al., 2004). Une série d'expériences a d'abord révélé que l'activité de l'ARN Pol III est essentielle pour l'inhibition transcriptionnelle de l'ARN Pol II induite par le choc thermique, ce qui corrélait avec l'implication de l'ARN B2 dans ce processus (Allen et al., 2004). De plus, comparativement aux ARN contrôles, l'enlèvement ou l'ajout de B2 a permis de confirmer que ce petit ncARN est spécifiquement nécessaire pour l'inhibition de l'ARN Pol II lors du choc thermique (figure 3). Dans les cellules de souris ayant subi un choc thermique, les chercheurs ont, par ailleurs, montré que B2 co-immunoprécipite avec l'ARN Pol II (Allen et al., 2004).

Un système de transcription reconstitué a été utilisé afin d'étudier les aspects mécanistiques de l'inhibition de l'ARN Pol II par le ncARN B2 (Espinoza et al., 2004). Les auteurs ont trouvé que, *in vitro*, B2 interagit directement et avec une forte affinité avec l'ARN Pol II et que cette interaction inhibe la formation de complexes de pré-initiation stables, ce qui résulte en un blocage de la synthèse de l'ARNm (Espinoza et al., 2004). La figure 5 montre un modèle illustrant le mécanisme de répression transcriptionnelle de l'ARN Pol II par l'ARN B2 en réponse au choc thermique dans les cellules murines. Une fois synthétisé par l'ARN Pol III, l'ARN B2 réprime la transcription à partir de promoteurs généraux de l'ARN Pol II, comme par exemple ceux des gènes codant pour l'actine et l'hexokinase II (Allen et al., 2004). Par contre, soulignons que les promoteurs dirigeant l'expression de gènes du choc thermique (par exemple, *hsp70*) échappent à cette répression par le ncARN B2. Il reste à déterminer comment les gènes induits par le choc thermique sont transcrits par l'ARN Pol II en présence du répresseur B2.

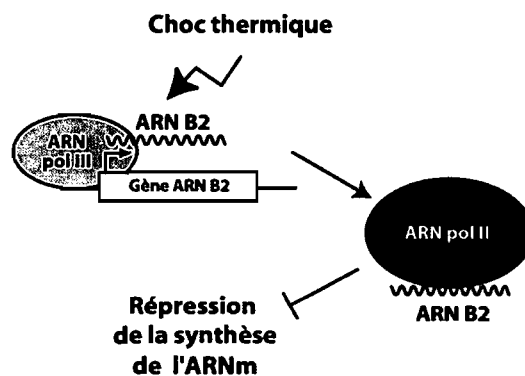


Figure 5 : Modèle illustrant le mécanisme de répression transcriptionnelle par le ncARN B2 après le choc thermique dans les cellules de souris. En réponse au choc thermique, l'ARN B2 qui est synthétisé par l'ARN Pol III, lie l'ARN Pol II et inhibe la formation de complexes de pré-initiation stables, bloquant la synthèse de l'ARNm. Figure modifiée de (Allen et al., 2004).

1.2 La transcription par l'ARN polymérase II

La synthèse de l'ARNm, catalysée par l'ARN Pol II chez les eucaryotes, est un processus biochimique élaboré comportant plusieurs étapes, telles l'initiation, l'élongation et la terminaison (revu par (Woychik et Hampsey, 2002; Hahn, 2004)). Lors du cycle transcriptionnel, l'ARN Pol II s'associe avec diverses protéines impliquées dans la régulation de ces étapes, incluant les facteurs généraux de transcription (FGT), les facteurs d'élongation et les facteurs jouant des rôles dans la formation de l'extrémité 3' du transcrit et la terminaison. Dans la section qui suit, les rôles des divers composants de la machinerie de transcription de l'ARN Pol II, ainsi que les différentes étapes du cycle transcriptionnel seront discutés.

1.2.1 La machinerie transcriptionnelle

1.2.1.1 L'ARN Pol II

L'ARN Pol II forme le cœur de la machinerie transcriptionnelle (Boeger et al., 2005). Hautement conservée entre les espèces eucaryotes, l'ARN Pol II est un complexe

multi-protéique constitué de 12 sous-unités (revu par (Trinh et al., 2006; Lee et Young, 2000)). Les sous-unités de l'ARN Pol II, appelées Rpb1 à Rpb12, ont des poids moléculaires variant de 7 à 220 kDa. Des études génétiques et biochimiques ont permis d'attribuer des fonctions aux différentes sous-unités de l'ARN Pol II. Ainsi, les deux plus grandes sous-unités, Rpb1 et Rpb2, forment le centre catalytique de l'enzyme et elles sont homologues des sous-unités β' et β de l'ARN Pol de bactérie (Coulombe et Burton, 1999). Cinq sous-unités, Rpb5, Rpb6, Rpb8, Rpb10, et Rpb12, sont aussi trouvées chez l'ARN Pol I et l'ARN Pol III. Les sous-unités Rpb3 et Rpb11 sont homologues de l'homodimère α_2 impliqué dans l'assemblage de l'ARN Pol bactérienne. Des travaux réalisés chez la levure *S. cerevisiae* indiquent que toutes les sous-unités de l'ARN Pol II, à l'exception de Rpb4 et 9, sont essentielles pour la viabilité cellulaire (Lee et Young, 2000). De plus, des études chez cet organisme ont aussi montré que les sous-unités Rpb4 et Rpb7 forment un sous-complexe pouvant se dissocier de l'enzyme suite à des changements des conditions environnementales (Choder et Young, 1993). L'hétérodimère Rpb4-Rpb7 semblerait jouer des rôles divers, tels l'exportation de l'ARNm, la transcription couplée à la réparation (TCR, *transcription coupled repair*) et la réponse au stress (revu par (Choder, 2004; Sampath et Sadhale, 2005)). Pour ce qui est de Rpb9, les études ont révélé que cette sous-unité possède de multiples fonctions. Ainsi, Rpb9 joue un rôle dans l'étape d'élongation, favorisant la sortie de l'ARN Pol II des sites de pause (Awrey et al., 1997; Hemming et al., 2000; Van, V et al., 2002). De plus, il a été démontré que cette sous-unité est impliquée dans le choix du site d'initiation ainsi que dans la fidélité de la transcription (Hull et al., 1995; Sun et al., 1996; Furter-Graves et al., 1994; Nesser et al., 2006). Rpb9 est aussi importante pour le processus de TCR (Li et Smerdon, 2002; Li et al., 2006). Enfin, des travaux récents révèlent que Rpb9 joue un rôle clé dans l'ubiquitinylation et la dégradation de la sous-unité Rpb1 en réponse aux dommages à l'ADN (Chen et al., 2007).

Les structures cristallographiques à haute résolution de l'ARN Pol II ont permis d'obtenir des informations au niveau atomique sur le fonctionnement de l'enzyme catalytiquement active et sur le mécanisme transcriptionnel (revu par (Kornberg, 2007)). Cette avancée spectaculaire a valu à Roger D. Kornberg le prix Nobel de chimie 2006

(Landick, 2006; Svejstrup et al., 2006; Cramer, 2006). Les structures cristallographiques de l'ARN Pol II de la levure *S. cerevisiae* obtenues par le groupe de Kornberg servent de référence pour toutes les formes de l'enzyme (Boeger et al., 2005). Tel que mentionné précédemment, l'enzyme de levure, qui a un poids moléculaire de 0.5 MDa, peut être dissociée en deux parties, une particule cœur de 10 sous-unités et un hétérodimère formé par les sous-unités Rpb4 et Rpb7 (Edwards et al., 1991). Les premières structures de l'ARN Pol II à 10 sous-unités ont été déterminées sous deux formes : l'enzyme libre et l'enzyme en élongation sur l'ADN (figure 6) (Cramer et al., 2001; Gnatt et al., 2001). Ces structures cristallographiques ont montré que les plus grandes sous-unités, Rpb1 et Rpb2, forment le cœur de l'enzyme dans lequel se retrouve un canal principal ou sillon (*cleft*) chargé positivement (Cramer, 2004). Ce canal accueille les acides nucléiques, lesquels sont retenus par des domaines protéiques, dont les mâchoires supérieures et inférieures (*upper and lower jaws*) qui contactent la matrice d'ADN. Un côté du canal est constitué d'un domaine protéique de Rpb1, la pince (*clamp*) mobile qui adopte une conformation dite ouverte dans la structure de l'enzyme libre. Le centre catalytique de l'enzyme est formé entre la pince, une hélice de pontage (*bridge helix*) de Rpb1 qui traverse le canal principal et un élément protéique de Rpb2, le mur ou *wall*, qui ferme l'extrémité du canal. Une ouverture en forme d'entonnoir inversé (*funnel*), appelée canal secondaire ou pore, s'ouvre dans le canal principal près du centre catalytique et sert à accueillir les ribonucléotides triphosphates (rNTP) arrivants. Le bord de ce pore contient une boucle (*loop*) de Rpb1 qui lie un ion Mg^{2+} (appelé métal A) via trois résidus aspartate. Un deuxième ion métallique, métal B, peut lier faiblement plus loin dans le pore (Cramer, 2004). La structure de l'enzyme en élongation a révélé, pour sa part, que l'ADN est déenroulé le long de neuf paires de bases, formant le duplexe hybride ADN-ARN juste en haut du métal A et du pore (Cramer et al., 2001; Gnatt et al., 2001). Dans cette structure, la pince est dans une conformation fermée.

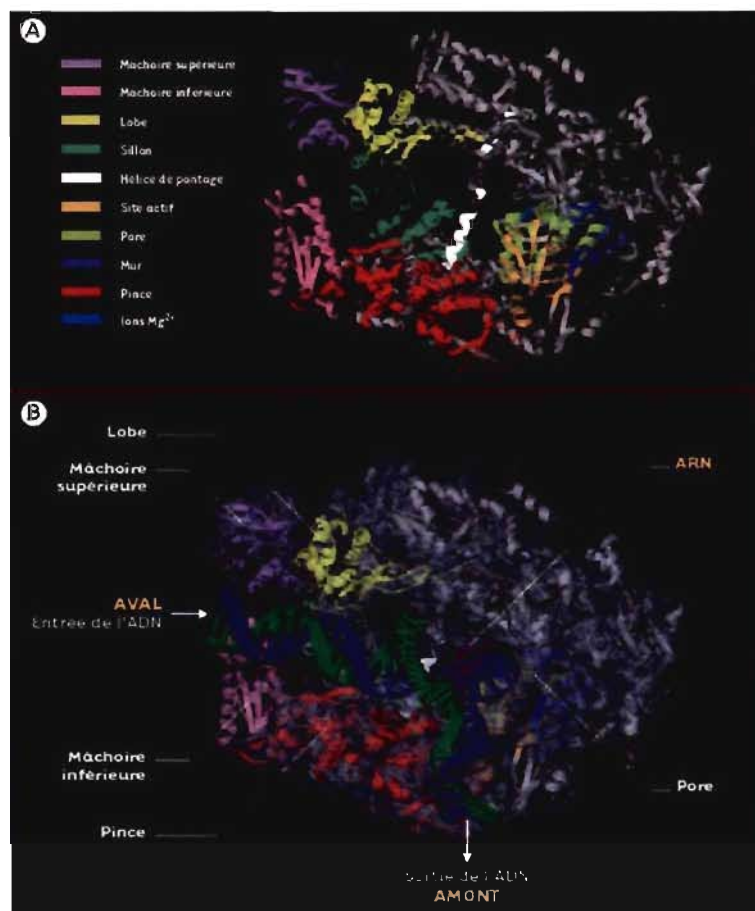


Figure 6 : Structure cristallographique de l'ARN polymérase II de levure manquant Rpb4 et Rpb7. (A) Les principaux éléments structuraux de l'enzyme libre, à une résolution de 2.8 Å, sont indiqués et colorés selon le code indiqué. (B) Les domaines protéiques de l'enzyme en élongation sur la matrice d'ADN à une résolution de 3.3 Å sont colorés comme en A. Les flèches montrent les sites d'entrée et de sortie de l'ADN. Figure tirée de (Langelier et al., 2002).

La particule cœur de l'ARN Pol II est catalytiquement active mais nécessite le sous-complexe Rpb4-Rpb7 pour l'initiation de la transcription à partir d'un promoteur (Edwards et al., 1991; Orlicky et al., 2001). En 2004, des structures cristallographiques à haute définition de l'ARN Pol II comprenant les 12 sous-unités ont été obtenues de manière indépendante par deux groupes (figure 7) (Armache et al., 2003; Bushnell et Kornberg, 2003). Leurs modèles confirment que Rpb4-Rpb7 s'attachent à la base de la pince via le domaine N-terminal de Rpb7 et que la pince est dans une conformation fermée. Ces résultats dévoilent également l'étroite surface d'interaction de Rpb4-Rpb7

avec l'enzyme, ce qui est compatible avec la dissociabilité du complexe. Ces structures révèlent ainsi que la majorité de la surface du sous-complexe Rpb4-Rpb7 est exposée et accessible pour des interactions avec des protéines ou des acides nucléiques. Les études montrent que Rpb4-Rpb7 peuvent apparemment recruter des facteurs à l'ARNm synthétisé sortant ou au CTD de l'ARN Pol II. En effet, chez la levure, le facteur Nrd1 connu pour lier l'ARN et la phosphatase du CTD, Fcp1 (*TFIIF-associated CTD phosphatase*), interagissent directement avec le sous-complexe Rpb4-Rpb7 (Mitsuzawa et al., 2003; Kimura et al., 2002). Par ailleurs, selon ces structures cristallographiques, une région de Rpb7 pouvant lier les acides nucléiques est dirigée vers le canal de sortie de l'ARNm. Une étude récente montre, en effet, que Rpb7 interagit avec le transcrit émergent du site actif de l'enzyme (Ujvari et Luse, 2006). De plus, ces observations pourraient expliquer le rôle possible de Rpb4 dans l'exportation de l'ARNm au cytoplasme (Farago et al., 2003). Très récemment, les travaux de l'équipe de Buratowski ont révélé que Rpb4 participe au recrutement des facteurs impliqués dans la maturation de l'extrémité 3' du transcrit (Runner et al., 2008). Enfin, une étude montre que Rpb4 serait également impliquée dans la dégradation de certains ARNm (Lotan et al., 2005).

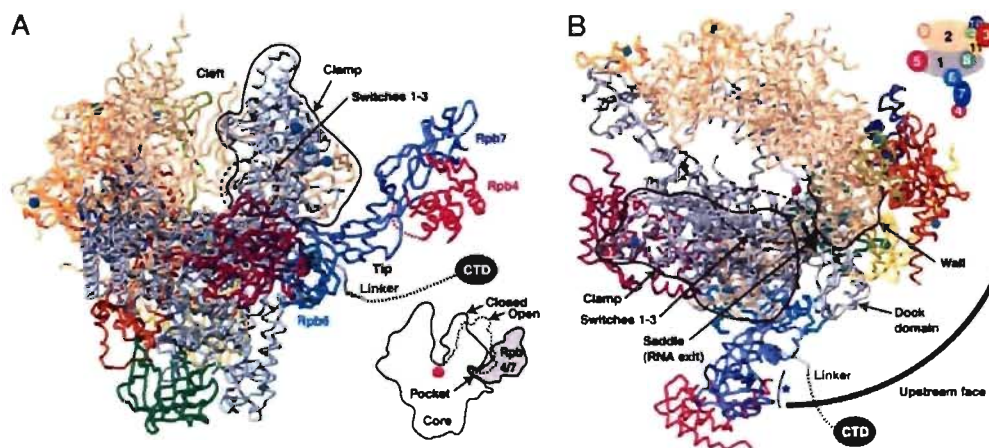


Figure 7 : Structure de l'ARN polymérase II à 12 sous-unités. A) Vue de l'ARN Pol II à 12 sous-unités montrant le site de liaison de l'hétérodimère Rpb4-Rpb7 et la pince (*clamp*) dans les conformations ouverte et fermée. La sphère rose indique le centre actif de l'enzyme. Les sous-unités sont colorées selon le code couleur indiqué sur le schéma en haut à droite en B. B) Vue de l'ARN Pol II complète montrant la localisation de Rpb4-Rpb7 à la base de la pince, à proximité du canal de sortie de l'ARNm. La direction de l'ARNm sortant est indiquée par la flèche noire.

L'astérisque bleu indique le domaine proposé de liaison à l'ARN de Rpb7. Figure tirée de (Armache et al., 2003).

Deux récentes structures cristallographiques à haute résolution de l'ARN Pol II rendent compte des mécanismes de sélection des nucléotides et de translocation après chaque ajout d'un nucléotide à la chaîne d'ARN synthétisé durant la polymérisation (Westover et al., 2004b; Westover et al., 2004a). Tout d'abord, la structure de l'enzyme dans l'état de post-translocation a révélé qu'une boucle appelée *lid*, formée par des résidus de Rpb1, est impliquée dans la séparation locale des brins de l'ADN matrice et de l'ARN synthétisé, participant ainsi au maintien de la bulle transcriptionnelle (Westover et al., 2004a). De plus, avec l'aide de deux autres éléments structuraux, la boucle *rudder* de Rpb1 et la boucle *fork-loop 1* de Rpb2, la boucle *lid* semble jouer un rôle dans la séparation de l'hybride ADN-ARN. La seconde structure cristallographique de l'ARN Pol II a, quant à elle, révélé que la sélection du nucléotide à être incorporé dans la chaîne d'ARN a lieu en deux étapes (Westover et al., 2004b). Ainsi, le rNTP arrivant entre d'abord par le pore et est positionné dans une orientation inversée au site E (*entry site*). Par la suite, une rotation du rNTP arrivant au site A (*addition site*) permet à l'enzyme de sélectionner les nucléotides formant la paire de base appropriée avec la matrice d'ADN. Lorsque l'appariement approprié est établi, un nouveau lien phosphodiester peut être formé par un mécanisme catalytique à deux ions métalliques (Steitz, 1998). Selon ce mécanisme, un premier ion magnésium, le métal A, est fermement lié au site A alors que le second cation, le métal B, favoriserait le départ du pyrophosphate relâché suite à la catalyse (Westover et al., 2004b). Un modèle a été proposé pour l'ajout de nucléotides : le rNTP entre au site E, et suite à la formation du lien phosphodiester, l'élément structural *bridge helix* subit une courbe, permettant la translocation des acides nucléiques. En reprenant sa position initiale, le *bridge helix* libère le site A, ce qui permet l'arrivée du nucléotide suivant (Westover et al., 2004b; Westover et al., 2004a).

Le CTD de la plus grande sous-unité de l'ARN Pol II est un domaine structural unique, constitué par une répétition heptapeptidique dont la séquence consensus est Y¹S²P³T⁴S⁵P⁶S⁷ (Dahmus, 1996). Bien que hautement conservé chez les différentes

espèces, le nombre de répétitions de l'heptapeptide varie d'une espèce à l'autre. Ainsi, le CTD de l'ARN Pol II de mammifères contient 52 répétitions alors que celui de la levure *S. cerevisiae* en contient 26-27. Le rôle clé du CTD dans la régulation de l'expression génique sera discuté dans une section suivante (1.3.2).

1.2.1.2 Les facteurs généraux de transcription

L'ARN Pol II catalyse la synthèse de l'ARNm mais elle se révèle incapable de reconnaître un promoteur et d'initier la transcription en absence de facteurs auxiliaires, appelés les facteurs généraux de transcription (FGT) (tableau III) (revu par (Woychik et Hampsey, 2002; Hahn, 2004)). Les FGT TFIID, TFIIA, TFIIB, TFIIIE, TFIIF et TFIIH interviennent dans la reconnaissance des séquences du promoteur, la réponse aux facteurs régulateurs ainsi que les changements conformationnels essentiels à l'activité de l'ARN Pol II pendant le cycle transcriptionnel (Orphanides et al., 1996; Reese, 2003).

Tableau III. Les facteurs généraux de transcription humains

Facteur	Sous-unité	Caractéristiques
TFIID	TBP TAF1 à 15	Lie la boîte TATA, induit une courbure de l'ADN, commun aux ARN pol I, II et III Lient l'Inr, le DPE et le DCE, impliqués dans la coactivation et la répression
TFIIA	TFIIA α TFIIA β TFIIA γ	TFIIA α et TFIIA β résultent d'un clivage protéolytique d'un précurseur, lie TBP et stabilise l'interaction boîte TATA-TBP, coactivation, anti-répression
TFIIB		Lie TBP, stabilise l'interaction boîte TATA-TBP, lie le BRE, participe au recrutement de l'ARN pol II-TFIIF, est impliqué dans le choix du site d'initiation
TFIIF	RAP74 RAP30	Lie l'ARN pol II, est impliqué dans le recrutement de l'ARN pol II au promoteur, enroulement du promoteur, dégagement du promoteur, stimule l'élongation
TFIIIE	TFIIIE34 TFIIIE56	Impliqué dans l'ouverture du promoteur indépendamment de TFIIH, participe au recrutement de TFIIH et stimule ses activités catalytiques, dégagement du promoteur
TFIIH	XPB XPD p62 p52 p44 p34 CDK7 Cycline H MAT1 TFB5	Hélicase 3'-5', ouverture et dégagement du promoteur, impliquée dans le NER Hélicase 5'-3', stimule la transcription, impliquée dans le NER Impliquée dans le NER, est la cible d'activateurs Impliquée dans le NER Stimule l'activité de XPD, impliquée dans le NER Impliquée dans le NER Phosphorylation du CTD Cycline de CDK7, lie le snARN U1 Interagit avec CDK7-Cycline H, impliquée dans le NER Stabilise le complexe TFIIH, impliquée dans le NER

1.2.1.2.1 TFIID

Le complexe TFIID est constitué par la protéine liant la boîte TATA appelée TBP (*TATA-box binding protein*) et 15 TAF (*TBP-associated factors*) (Burley et Roeder, 1996; Lee et Young, 1998). Ce complexe, d'environ 750 kDa, est hautement conservé de la levure jusqu'à l'homme. La structure cristallographique de TBP a révélé que cette protéine de 38 kDa possède une face concave en forme de selle à cheval permettant de contacter l'ADN et une surface convexe impliquée dans les interactions avec les facteurs TFIIA, TFIIB et les TAFs (revue par (Albright et Tjian, 2000)). Les études ont révélé qu'en se liant à la boîte TATA, TBP induit une courbure de l'ADN (Nikolov et Burley, 1997). Ce facteur joue ainsi un rôle clé dans l'initiation et l'assemblage du complexe de pré-initiation transcriptionnel.

Jusqu'à récemment, TBP était considéré comme un facteur transcriptionnel universel, notamment essentiel pour l'initiation de la transcription par l'ARN Pol I, II et III. Cette conclusion est maintenant remise en cause avec la découverte de protéines apparentées à TBP au niveau de la séquence, les *TBP-related factors* (TRF1, TRF2 et TRF3) (Reina et Hernandez, 2007; Hochheimer et Tjian, 2003; Davidson, 2003). TRF2 présente 60% d'acides aminés similaires et 40% d'acides aminés identiques à TBP alors que TRF1 et TRF3 sont encore plus proches de ce facteur. N'existant que chez la drosophile, TRF1 agit comme un facteur de transcription spécifique pour l'expression d'un sous-ensemble de gènes (Holmes et Tjian, 2000). Quant à TRF2, ce facteur a été identifié en complexe avec des facteurs de remodelage de la chromatine chez la drosophile (Hochheimer et al., 2002). TRF2 ne semble pas lier la boîte TATA *in vitro*, mais il s'associe avec les facteurs TFIIA et TFIIB (Moore et al., 1999; Rabenstein et al., 1999). TRF3, pour sa part, a été caractérisé dans les cellules humaines, de souris et de *Xenopus* (Persengiev et al., 2003; Deato et Tjian, 2007; Jallow et al., 2004). Ce facteur est capable de lier la boîte TATA en présence de TFIIA et de TFIIB, et il peut promouvoir la transcription *in vitro* (Bartfai et al., 2004).

Pour ce qui est des TAF, appelés TAF1 à TAF15, ces facteurs semblent réguler la transcription à différents niveaux (revu par (Albright et Tjian, 2000; Gangloff et al., 2001; Tora, 2002; Davidson et al., 2005)). Ainsi, ils agissent comme des coactivateurs, établissant le lien entre les activateurs transcriptionnels et la machinerie basale de l'ARN Pol II. De plus, ils jouent un rôle stabilisateur du complexe TFIID au promoteur en influençant positivement l'interaction de TBP avec la boîte TATA. Enfin, ils peuvent induire des modifications de la chromatine, comme par exemple TAF250 qui possède une activité histone acétyltransférase, de phosphorylation de l'histone H2B et de mono-ubiquitinylation de l'histone H1 (Mizzen et al., 1996; Maile et al., 2004; Pham et Sauer, 2000). Plusieurs TAF contiennent un motif *histone fold* retrouvé sur les histones, contribuant à la reconnaissance des éléments du promoteur (Gangloff et al., 2001; Selleck et al., 2001). Ce motif permet aussi aux TAF d'interagir entre eux et de former un octamère semblable à l'octamère d'histone.

D'autre part, le complexe TFIID a depuis longtemps été considéré comme le seul facteur reconnaissant le promoteur basal de l'ARN Pol II (revu par (Tora, 2002)). Toutefois, il existe des complexes TFIID dépourvus de certains TAF, engendrant un large spectre de variantes de ce complexe (Muller et al., 2007). Plusieurs études ont aussi montré qu'un sous-ensemble de TAF (TAF4, 5, 6, 8, 9, 10 et 12) forme un complexe TFIID minimal, ne contenant pas TBP (Wright et al., 2006; Leurent et al., 2004; Demeny et al., 2007). Ce dernier complexe semble pouvoir s'associer avec plusieurs combinaisons de TAF et TBP. En accord avec un modèle selon lequel les divers sous-modules du complexe TFIID reconnaissent différents éléments du promoteur basal, certains TAF ont été impliqués dans la reconnaissance du promoteur entourant le site d'initiation de la transcription (site +1) (Oelgeschlager et al., 1996). Finalement, plusieurs TAF entrent aussi dans la composition de complexes dépourvus de la protéine TBP. En effet, ces TAF sont des composants de complexes multiprotéiques possédant une activité histone acétyltransférase, tels SAGA (*Spt-Ada-Gcn5 Acetylase coactivator complex*) chez la levure ou les complexes TFTC (*TBP-Free TAF-Containing complex*), STAGA (*SPT3-TAF9-GCN5 containing complex*) et PCAF (*P300/CREB-Binding Protein Associated Factor complex*) chez les eucaryotes supérieures (Nagy et Tora, 2007; Baker et Grant,

2007; Martinez et al., 1998). Ces complexes jouent un rôle clé dans l'acétylation globale de la chromatine.

1.2.1.2.2 TFIIA

Le complexe TFIIA de levure est constitué de deux sous-unités, TOA1 et TOA2 (Ranish et Hahn, 1991). En contraste, chez l'humain et la drosophile, le complexe TFIIA est composé de trois sous-unités, TFIIA α (35 kDa), β (21 kDa) et γ (12 kDa) (DeJong et Roeder, 1993; Kraemer et al., 2001; Ma et al., 1993). Le clonage de TFIIA a révélé que les deux plus grandes sous-unités, α et β , étaient codées par un gène unique, TFIIA $\alpha\beta$ (55 kDa). En effet, les sous-unités α et β sont produites par clivage post-traductionnel de ce précurseur (Hoiby et al., 2004). Autant la forme non-clivée que la forme clivée de TFIIA $\alpha\beta$ peut s'associer avec la sous-unité TFIIA γ *in vivo* (Mitsiou et Stunnenberg, 2003). De plus, les deux formes interagissent avec TBP sur l'ADN et elles supportent la transcription *in vitro* (Hoiby et al., 2004; Sun et al., 1994). TFIIA $\alpha\beta$ et TFIIA γ sont homologues aux sous-unités TOA1 et TOA2 de levure (DeJong et Roeder, 1993; Kraemer et al., 2001; Ma et al., 1993).

Le rôle de TFIIA est demeuré éluif pendant des années. Les travaux initiaux de D. Reinberg ont décrit TFIIA comme un facteur général de transcription, essentiel pour reconstituer la réaction de transcription basale *in vitro* (Reinberg et al., 1987). Cette fonction de TFIIA a par la suite été mise en cause étant donné l'observation de résultats contradictoires. En effet, TFIIA n'est pas nécessaire pour l'assemblage du complexe de pré-initiation transcriptionnel (Sawadogo et Roeder, 1985; Van Dyke et al., 1988). Ces différences de données pourraient s'expliquer par des impuretés au niveau des facteurs ou des fractions protéiques utilisés dans les essais de transcription *in vitro*. D'autre part, les études cristallographiques ont révélé que TFIIA interagit directement avec TBP (Tan et al., 1996; Geiger et al., 1996; Kraemer et al., 2001). Ainsi, TFIIA stabilise la liaison de TBP au promoteur en établissant des contacts avec l'ADN en amont de la boîte TATA ainsi qu'avec TBP et TAF11. De plus, plusieurs études ont montré que TFIIA augmente

l'affinité de TBP pour l'ADN, favorisant ainsi l'assemblage du complexe de pré-initiation (Buratowski et al., 1989; Imbalzano et al., 1994; Kang et al., 1995). Toutefois, lorsque des facteurs hautement purifiés sont utilisés, TFIIA ne semble pas stimuler la transcription dépendante de TBP à partir d'un promoteur basal, suggérant qu'il manque un co-facteur aux formes purifiées de TFIIA (Ma et al., 1993; Ozer et al., 1994; Ge et Roeder, 1994; Cortes et al., 1992). TFIIA favorise également la liaison de TFIID à l'ADN par un mécanisme anti-répressif en compétitionnant avec le domaine N-terminal de TAF1 (Kokubo et al., 1998). En effet, le domaine N-terminal de TAF1 inhibe TBP en se liant aux surfaces concave et convexe de ce facteur. D'autre part, les études montrent que TFIIA est déterminant pour la transcription basale à partir de promoteurs ne possédant pas de boîte TATA (Pugh, 2000). TFIIA semble aussi agir comme cofacteur de plusieurs activateurs transcriptionnels, notamment AP-1, Gal4-AH, Zta, VP16, CTF, NTF, Sp1 (Ma et al., 1993; Ozer et al., 1994; Yokomori et al., 1994; Sun et al., 1994; Lieberman et Berk, 1994). De plus, TFIIA peut stimuler l'effet de certains coactivateurs, tels PC4 et HMG2 (Ge et Roeder, 1994; Shykind et al., 1995). En somme, les chercheurs semblent s'accorder sur le fait que TFIIA se définit plutôt comme un coactivateur transcriptionnel que comme un FGT. Pour supporter cette hypothèse, des travaux ont révélé que TFIIA interagit, physiquement et génétiquement, avec un grand nombre de facteurs, incluant TFIID, des composants du complexe SAGA, ou encore des répresseurs, comme par exemple Mot1 et NC2 (Hahn, 2004).

Des travaux très récents ont révélé que le précurseur TFIIA $\alpha\beta$ est clivé par Taspase 1, une protéase initialement identifiée comme responsable de la protéolyse du proto-oncogène MLL (Zhou et al., 2006). Ces études montrent que le clivage de TFIIA ne semble pas être une étape nécessaire pour l'activation de ce facteur étant donné que le précurseur non-clivé est transcriptionnellement actif. En effet, le clivage de TFIIA agirait plutôt comme un mécanisme de dégradation permettant d'assurer un contrôle du niveau de TFIIA dans la cellule. Le clivage de TFIIA pourrait également être impliqué dans la régulation de la transcription étant donné que les formes non-clivée et clivée peuvent posséder des spécificités du promoteur différentes lors des étapes de développement et de différenciation cellulaire (Hoiby et al., 2007).

1.2.1.2.3 TFIIB

Le facteur TFIIB, formé d'une seule sous-unité d'environ 35 kDa, joue un rôle central dans l'assemblage du complexe de pré-initiation (revue par (Deng et Roberts, 2007)). Ainsi, TFIIB interagit directement avec TBP, stabilisant l'interaction TBP-ADN, et il participe au recrutement du complexe préformé TFIIF-ARN Pol II au promoteur. Comme TFIID, le facteur TFIIB peut lier directement l'ADN (Nikolov et al., 1995). En effet, TFIIB se fixe au niveau de séquences en amont et en aval de la boîte TATA appelées motifs BRE (*TFIIB recognition element*) (Lagrange et al., 1998; Deng et Roberts, 2005). TFIIB contient deux domaines fonctionnels : un module N-terminal possédant un motif ruban de zinc (*zinc ribbon*), attaché par une région flexible au domaine C-terminal, qui lui est composé d'hélices alpha (Bagby et al., 1995; Nikolov et al., 1995; Zhu et al., 1996; Chen et Hahn, 2003). La structure cristallographique d'un complexe formé de TFIIB lié à l'ARN Pol II a révélé que le motif ruban est localisé sur le domaine *dock* de l'enzyme, à proximité du canal de sortie de l'ARNm (Bushnell et al., 2004). La surface du motif ruban serait essentielle pour le recrutement de l'ARN Pol II au promoteur (Hahn et Roberts, 2000). La région flexible séparant les deux domaines de TFIIB, quant à elle, contient une courte séquence phylogénétiquement conservée, connue sous le nom de *B-finger*, laquelle est localisée au centre catalytique de l'ARN Pol II, jouant un rôle dans la détermination du site d'initiation de la transcription. Enfin, le domaine C-terminal en interagissant avec l'ARN Pol II et le complexe TBP-ADN guide l'ADN pour son désenroulement et oriente la transcription (Bushnell et al., 2004).

Le domaine N-terminal de TFIIB contient aussi une région très conservée et hautement chargée, appelée CDD (*charge clustered domain*). Les domaines N et C-terminaux peuvent former des interactions intramoléculaires, résultant dans une conformation dite fermée (Hawkes et al., 2000; Fairley et al., 2002; Zheng et al., 2004). Le CDD module l'affinité des interactions intramoléculaires entre les domaines N et C-terminaux de TFIIB. Ces interactions intramoléculaires de TFIIB semblent agir comme un mécanisme de contrôle transcriptionnel (Elsby et Roberts, 2004). Par exemple, le domaine d'activation de l'activateur VP16 peut défaire l'interaction intramoléculaire,

induisant ainsi un changement conformationnel dans TFIIB (Bagby et al., 1995). En somme, les changements conformationnels influencent la fonction de TFIIB dans la formation du complexe de pré-initiation et modulent sa réponse aux facteurs transcriptionnels (Elsby et Roberts, 2004).

Des travaux très récents révèlent que TFIIB peut être localisé aussi bien au niveau de la région du promoteur que de la région du terminateur transcriptionnel chez certains gènes de levure (Singh et Hampsey, 2007). Ces résultats suggèrent que TFIIB joue un rôle dans la formation de boucles géniques ou *gene looping* en juxtaposant les régions du promoteur et du terminateur. Les auteurs proposent un modèle selon lequel TFIIB lie l'ARN Pol II au terminateur, laquelle s'associe ensuite avec le promoteur basal, facilitant ainsi la ré-initiation de la transcription par la même molécule de l'enzyme.

1.2.1.2.4 TFIIF

TFIIF a initialement été identifié grâce à sa capacité d'interaction avec l'ARN Pol II, d'où le nom de RAP (*RNA polymerase II associated protein*) attribué à ses deux sous-unités (Sopta et al., 1985). Ce facteur est un hétérotetramère composé de deux grandes (RAP74) et deux petites (RAP30) sous-unités. RAP74 et RAP30 jouent divers rôles durant les différentes étapes du cycle transcriptionnel de l'ARN Pol II. Des études biochimiques ont révélé que TFIIF facilite l'entrée de l'enzyme au complexe de pré-initiation et que ce facteur est nécessaire pour l'association subséquente de TFIIE et TFIIH (Buratowski et al., 1991; Conaway et al., 1991; Flores et al., 1991; Maxon et al., 1994; Parvin et al., 1992; Chambers et al., 1995). Selon des données de microscopie électronique, TFIIF de levure interagit avec l'ARN Pol II près de l'élément *clamp* ainsi que des sous-unités Rpb4-Rpb7 de l'enzyme (Chung et al., 2003). Des expériences de photo-pontage protéine-ADN ont permis d'identifier les contacts établis par TFIIF et la matrice d'ADN. Ainsi, les deux sous-unités RAP30 et RAP74 interagissent avec l'ADN de part et d'autre de la boîte TATA (Kim et al., 1997). Ces résultats sont supportés par des images de microscopie électronique du complexe TBP-TFIIB-TFIIF-ARN Pol II-

ADN, qui suggèrent aussi un enroulement du promoteur autour du complexe de pré-initiation de l'ARN Pol II (Forget et al., 1997). Une fois l'assemblage du complexe de pré-initiation complété, la sous-unité RAP74 induit un changement important dans la topologie du promoteur (Robert et al., 1998). Ainsi, RAP74 entraîne la formation d'une courbure dans l'ADN, à proximité du site +1, ce qui permet un enroulement serré du promoteur autour de l'ARN Pol II. De plus, TFIIF serait impliqué dans le dégagement du promoteur suite à la formation du premier lien phosphodiester du transcrit naissant (Yan et al., 1999). D'autre part, TFIIF peut stimuler l'activité de la phosphatase Fcp1 du CTD de l'ARN Pol II (Chambers et al., 1995). Enfin, plusieurs études ont montré que le facteur TFIIF joue aussi un rôle dans l'étape d'élongation. Ce rôle de TFIIF sera développé dans une section suivante (1.2.1.3.1).

En accord avec les multiples rôles mécanistiques de TFIIF, des expériences biochimiques ont permis d'identifier différents domaines au niveau des sous-unités RAP30 et RAP74 de ce facteur, lui permettant d'interagir avec divers facteurs transcriptionnels et l'ADN. RAP74 humain comprend un domaine N-terminal globulaire qui lie RAP30, une région centrale hautement chargée, et un domaine C-terminal globulaire impliqué dans l'interaction avec l'ARN Pol II, TFIIB et Fcp1 (Kobor et al., 2000; Chambers et al., 1995; Wang et Burton, 1995; Fang et Burton, 1996). Quant à RAP30 humain, ce facteur est composé d'un domaine N-terminal responsable de lier RAP74 et TFIIB, un domaine central qui interagit avec l'ARN Pol II et un domaine C-terminal de liaison à l'ADN similaire à celui trouvé chez la famille σ^{70} de facteurs sigma bactériens (Tan et al., 1994; Izban et Luse, 1992; Groft et al., 1998; Garrett et al., 1992; McCracken et Greenblatt, 1991).

1.2.1.2.5 TFIIE

Le facteur TFIIE humain est un hétérodimère formé des sous-unités TFIIE α (TFIIE56) et TFIIE β (TFIIE34) (Ohkuma et al., 1990; Ohkuma et al., 1991; Sumimoto et al., 1991). Ces deux sous-unités possèdent plusieurs motifs structuraux. Une région

hautement conservée de TFIIE56 contient un motif *zing finger*, lequel est essentiel pour l'activité transcriptionnelle de TFIIE (Ohkuma et al., 1995; Kuldell et Buratowski, 1997; Maxon et Tjian, 1994). Le domaine central de TFIIE34 forme un motif hélice capable de lier l'ADN de façon non-spécifique (Okuda et al., 2000). Cette sous-unité de TFIIE lie également l'ADN simple-brin et double-brin (Okamoto et al., 1998; Okuda et al., 2000; Yokomori et al., 1998). D'autre part, les analyses biochimiques ont montré que TFIIE interagit avec plusieurs composants du complexe de pré-initiation. Ainsi, TFIIE interagit directement avec l'ARN Pol II (Bushnell et al., 1996). Plus précisément, des expériences de cristallographie de TFIIE en complexe avec l'ARN Pol II ont révélé que TFIIE est localisé près du centre catalytique de l'enzyme (Leuther et al., 1996). De plus, TFIIE lie le promoteur à environ 10 pb en amont du site +1 (Douziech et al., 2000). En effet, TFIIE contacte le promoteur aux alentours de la région de la bulle transcriptionnelle, sans altérer les interactions protéine-ADN entre l'ARN Pol II et les autres FGT (Kim et al., 2000).

TFIIE est impliqué dans les événements tardifs de l'assemblage du complexe de pré-initiation, incluant le recrutement de TFIIH et la régulation des activités de ce facteur (Ohkuma, 1997). Ainsi, TFIIE stimule l'activité kinase de TFIIH responsable de phosphoryler le domaine CTD de l'ARN Pol II (Ohkuma et al., 1995). De plus, TFIIE assiste la sous-unité hélicase XPB de TFIIH dans l'ouverture de la bulle de transcription à la position -10 (Ohkuma et Roeder, 1994). En effet, TFIIE et TFIIH sont nécessaires pour l'ouverture du promoteur dépendante d'ATP avant la formation du premier lien phosphodiester. Au moment de la transition entre les étapes d'initiation et d'élongation, TFIIE joue aussi un rôle dans le dégagement du promoteur en régulant les activités kinase et hélicase de TFIIH. Finalement, TFIIE coopère avec les facteurs TFIIF et TFIIH afin de supprimer les pauses à proximité du promoteur, facilitant ainsi la transition de l'ARN Pol II vers une élongation productive (Dvir et al., 2001).

1.2.1.2.6 TFIIH

Contrairement aux autres FGT, le facteur TFIIH, lequel joue un rôle non seulement dans la transcription mais également dans la réparation de l'ADN et le contrôle du cycle cellulaire, est composé de dix sous-unités divisées en deux sous-complexes : un module central (*core* TFIIH) et un module kinase, nommé CAK (*cyclin-activating kinase complex*) (revu par (Coin et Egly, 1998; Zurita et Merino, 2003; Kraemer, 2004)). Durant de nombreuses années, *core* TFIIH était connu pour être composé de six sous-unités : XPB, XPD, p62, p52, p44 et p34 (Takagi et al., 2003). Les protéines XPD (*xeroderma pigmentosum D*) et XPB (*xeroderma pigmentosum B*) sont deux ADN hélicases utilisant l'énergie de l'hydrolyse de l'ATP pour catalyser l'ouverture du promoteur autour du site +1 (Tirode et al., 1999). XPB est absolument requise pour la transcription alors que l'hélicase XPD, laquelle est dispensable, stimule la réaction transcriptionnelle. De plus, XPD est impliquée dans la liaison de CAK au sous-complexe *core* TFIIH (Chang et Kornberg, 2000; Schultz et al., 2000). La sous-unité p44 interagit avec XPD afin de stimuler son activité hélicase. Pour sa part, p52 interagit avec XPB, stimulant son activité ATPase (Coin et al., 2007). La sous-unité p62 est un composant structural du *core* TFIIH connu pour interagir avec divers activateurs transcriptionnels ainsi que l'endonucléase XPG impliquée dans la réparation de l'ADN (revu par (Zurita et Merino, 2003)). Quant au CAK, ce sous-complexe contient trois sous-unités : MAT1, Cycline H et CDK7. (Chen et al., 2003). La kinase CDK7, associée à la Cycline H, phosphoryle la sérine à la position 5 (Ser 5) du CTD de la sous-unité Rpb1 de l'ARN Pol II (Lu et al., 1992). De plus, ce sous-complexe joue un rôle important dans le contrôle du cycle cellulaire pendant la transition de la phase G2 à M (Larochelle et al., 1998). Par ailleurs, XPD semble réguler l'activité de CAK. Récemment, la dixième sous-unité de TFIIH, appelée TFB5, a été identifiée chez la levure et chez l'humain grâce à l'analyse protéomique quantitative de complexes d'initiation transcriptionnels (Giglia-Mari et al., 2004; Ranish et al., 2004). TFB5 est une petite protéine de 8 kDa. Ces études montrent que TFB5 co-précipite et co-purifie avec *core* TFIIH. En particulier, des travaux très récents révèlent que TFB5 interagit avec les sous-unités p52 et XPD (Coin et al., 2006; Zhou et al., 2007). Contrairement aux autres sous-unités de TFIIH, le facteur TFB5 n'est pas essentiel pour la viabilité cellulaire. En outre, TFB5 semble être spécifiquement impliqué dans la réparation de l'ADN, ne participant pas au mécanisme de transcription (Coin et al., 2006)

(Giglia-Mari et al., 2006). Ainsi, TFB5 serait capable de stimuler l'activité ATPase de XPB, importante pour l'ouverture de la bulle transcriptionnelle lors du processus de réparation des dommages à l'ADN (Coin et al., 2006).

TFIIH est le dernier composant à intégrer le complexe de pré-initiation transcriptionnel de l'ARN Pol II. Ce facteur contient des sous-unités possédant des activités hélicase et kinase essentielles pour l'initiation, l'élongation et le dégagement du promoteur lors de la transcription par l'ARN Pol II (Tirode et al., 1999). La structure à faible résolution déterminée par microscopie électronique du sous-complexe *core* de TFIIH a révélé que celui-ci a la forme d'un anneau avec les deux sous-unités hélicases XPD et XPB localisées de chaque côté d'une importante protubérance (Chang et Kornberg, 2000; Schultz et al., 2000). Le centre de l'anneau semble pouvoir accommoder l'ADN double-brin. XPB est une hélicase possédant une activité 3'-5' alors que XPD a une activité hélicase 5'-3' (Schaeffer et al., 1993; Schaeffer et al., 1994; Drapkin et al., 1994). Des études de photo-pontage protéine-ADN ont permis de localiser XPD et XPB dans la région -9 à +2 du promoteur, où ces deux hélicases participent à l'ouverture de la double hélice d'ADN nécessaire pour l'initiation transcriptionnelle (Douziech et al., 2000).

Lorsque certaines sous-unités de TFIIH sont mutées, elles sont responsables de maladies génétiques chez l'humain. Ainsi, des mutations au sein des hélicases XPB et XPD sont à l'origine de trois syndromes : le *xeroderma pigmentosum* (XP), le syndrome de Cockayne (CS) et la trichothiodystrophie (TTD) (Vermeulen et al., 1997). Ces maladies sévères sont caractérisées par une photosensibilité extrême, des problèmes de peau et des cheveux ultra-cassants chez l'humain (Itin et al., 2001). Ces mutations affectent le mécanisme de réparation des lésions de l'ADN par excision de nucléotides (NER, *nucleotide excision repair*), la transcription basale ainsi que l'activation transcriptionnelle par certains récepteurs nucléaires (Dubaele et al., 2003). L'activité hélicase de XPB est essentielle dans le processus de réparation de l'ADN ainsi que pour la transcription, alors que l'activité hélicase de XPD n'est requise que pour la réparation de l'ADN (Tirode et al., 1999). Pour ce qui est de TFB5, ce facteur est codé par le gène

TTD-A et des mutations au niveau de ce gène sont responsables du syndrome TTD (Giglia-Mari et al., 2004). Des cellules dépourvues de la protéine TTD-A sont connues pour être déficientes dans la réparation de l'ADN par NER ainsi que pour être sensibles aux irradiations aux UV (de Boer et Hoeijmakers, 2000) (Stefanini et al., 1993). En somme, ces résultats semblent indiquer que TFB5 joue un rôle important dans la réparation de l'ADN. En effet, deux équipes ont récemment montré que TFB5 participe directement au NER (Zhou et al., 2007; Ding et al., 2007).

TFIIH interagit avec une grande variété de facteurs durant la transcription par l'ARN Pol II, incluant des récepteurs nucléaires, des facteurs transcriptionnels spécifiques de tissus, des complexes de remodelage de la chromatine et le snARN U1, tel que vu précédemment (Zurita et Merino, 2003). Enfin, TFIIH semble participer aussi dans la transcription par l'ARN Pol I. Toutefois, son rôle exact dans ce processus de synthèse de l'ARNr reste encore à déterminer.

1.2.1.3 Les facteurs d'élongation

Les facteurs d'élongation modulent l'activité de l'ARN Pol II lors de l'étape d'élongation transcriptionnelle par différents mécanismes (Shilatifard et al., 2003; Sims, III et al., 2004; Saunders et al., 2006). Ces facteurs peuvent prévenir les pauses et les arrêts de l'ARN Pol II, modulant ainsi l'activité de l'enzyme, ou encore remodeler la chromatine afin de faciliter l'élongation du transcrit.

1.2.1.3.1 Les facteurs d'élongation modulant l'activité de l'ARN Pol II

Les pauses transcriptionnelles ont lieu lorsque l'ARN Pol II bloque l'ajout de rNTP au transcrit naissant pour un certain temps, sans jamais réussir à poursuivre l'élongation (Saunders et al., 2006). Contrairement aux arrêts transcriptionnels, les pauses sont réversibles et elles semblent constituer un mode naturel de régulation de la

transcription. Plusieurs facteurs suppriment les pauses, et par conséquent, ils modulent le taux d'élongation transcriptionnelle. Ces facteurs incluent le système P-TEFb/NELF/DSIF, TFIIF, les Elongins, la famille ELL et TFIIS (figure 8).

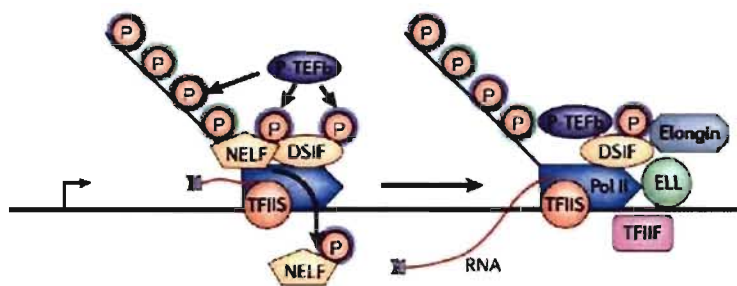


Figure 8 : Modulation de l'étape d'élongation de l'ARN Pol II par les facteurs P-TEFb, NELF, DSIF, TFIIF, Elongins, ELL et TFIIS. En phosphorylant les facteurs d'élongation négatifs DSIF et NELF ainsi que la sérine à la position 2 (Ser 2) du CTD de l'ARN Pol II, le facteur P-TEFb stimule une élongation processive de la transcription. TFIIS aide l'ARN Pol II à sortir de l'arrêt transcriptionnel en stimulant son activité de clivage de l'extrémité 3' du transcrit. Une fois phosphorylé, NELF se dissocie du complexe transcriptionnel alors que DSIF, TFIIS et P-TEFb poursuivent la transcription avec l'ARN Pol II. Les facteurs TFIIF, ELL et Elongins stimulent l'activité d'élongation de l'ARN Pol II. P, groupement phosphorylé. Figure tirée de (Saunders et al., 2006).

Price et ses collaborateurs ont été les premiers à constater que l'étape d'élongation de la transcription est hautement régulée (revu par (Peterlin et Price, 2006)). Ces chercheurs ont montré que l'élongation par l'ARN Pol II est sujette à une régulation négative et qu'une kinase du CTD de l'enzyme, spécifique de la sérine à la position 2 (Ser 2), est requise pour supprimer ce blocage de l'élongation. Cette kinase, CDK9, est une sous-unité du complexe P-TEFb, qui contient aussi une cycline (cycline T1 ou T2) (Marshall et Price, 1995; Fu et al., 1999; Peng et al., 1998). Cette dernière est absolument nécessaire pour l'activité du complexe P-TEFb (Peng et al., 1998). Plus tard, d'autres travaux ont permis d'identifier deux complexes impliqués dans l'induction des pauses de la transcription : les facteurs DSIF (*DRB sensitivity-inducing factor*) et NELF (*negative elongation factor*). DSIF, comme son nom l'indique, est un complexe capable d'induire la sensibilité au DRB, un inhibiteur de kinases (Price, 2000). Constitué par deux sous-unités, Spt5 (160 kDa) et Spt4 (14 kDa), DSIF interagit directement avec l'ARN Pol II

(Hartzog et al., 1998). Pour sa part, le complexe multi-protéique NELF, composé de cinq sous-unités, coopère avec le complexe DSIF afin de provoquer la pause de l'ARN Pol II (Yamaguchi et al., 1999). Une étude récente montre que DSIF et NELF agissent en inhibant la fonction du facteur d'élongation TFIIS {Palangat, 2005 7030 /id}. En phosphorylant DSIF et NELF, P-TEFb neutralise les effets négatifs de ces deux facteurs (Fujinaga et al., 2004). La kinase CDK9 de P-TEFb phosphoryle également la Ser 2 du CTD de l'ARN Pol II, déjà partiellement phosphorylé au niveau de la Ser 5 par le facteur TFIIH (Marshall et al., 1996). P-TEFb est localisé au niveau de gènes activement transcrits et sa déplétion engendre des défauts dans l'expression génique (Lis et al., 2000; Ni et al., 2004; Shim et al., 2002). De plus, plusieurs régulateurs transcriptionnels spécifiques des gènes interagissent avec P-TEFb (Lis et al., 2000). Dans les cellules humaines, la protéine Brd4 est responsable de recruter P-TEFb à la grande majorité des gènes via son interaction avec des histones acétylées (Jang et al., 2005; Yang et al., 2005). Comme discuté auparavant, l'activité de P-TEFb est hautement régulée par le snARN 7SK, en association avec les protéines HEXIM1/HEXIM2 (revu par (Peterlin et Price, 2006; Zhou et Yik, 2006)).

Les facteurs TFIIF, ELL (*eleven-nineteen lysine-rich in leukemia*) et Elongins ont tous des effets similaires sur l'élongation transcriptionnelle *in vitro*, et stimulent la transcription de l'enzyme par la suppression des pauses transitoires (Conaway et Conaway, 1999). Toutefois, ces facteurs peuvent agir à différentes étapes de l'élongation *in vivo*. ELL et Elongins ont été purifiées par leur capacité à stimuler le taux de synthèse de l'ARN Pol II *in vitro* (Saunders et al., 2006; Sims, III et al., 2004). TFIIF joue non seulement un rôle dans l'étape d'initiation mais aussi en élongation. Ainsi, les deux sous-unités de TFIIF stimulent la vitesse d'élongation et suppriment les pauses transitoires de l'ARN Pol II (Kephart et al., 1994; Lei et al., 1998; Tan et al., 1994; Izbán et Luse, 1992; Bengal et al., 1991). De plus, suite à son dégagement du promoteur, TFIIF se réassocie avec le complexe d'élongation transcriptionnel, et en particulier lorsque l'ARN Pol II est arrêtée aux sites de pauses (Zawel et al., 1995). Il semblerait que TFIIF stimule surtout la transcription de l'ARN Pol II au niveau du dégagement du promoteur, alors que les facteurs Elongins et ELL stimulent l'élongation le long de la région codante d'un gène

(Saunders et al., 2006). En support de ce modèle, les Elongins ne stimulent la synthèse de l'ARNm que lorsque le transcrit a plus de 8-9 nt, et la stimulation maximale de ce facteur requiert la dissociation de TFIIF du complexe transcriptionnel (Moreland et al., 1998). Par ailleurs, chez la Drosophile, Elongin et ELL co-localisent avec l'ARN Pol II hyperphosphorylée, la forme de l'enzyme trouvée en élongation (Gerber et al., 2005; Gerber et al., 2001).

Les arrêts transcriptionnels peuvent être définis comme une interruption irréversible de la synthèse d'ARNm, où l'ARN Pol II est incapable de reprendre l'étape d'élongation sans l'aide de facteurs auxiliaires. L'arrêt semble résulter d'un recul de l'enzyme sur la matrice d'ADN (*backtracking*) lorsque celle-ci possède un hybride ADN-ARN instable au sein de son centre catalytique (Nudler et al., 1997). Pour sortir de l'arrêt transcriptionnel et poursuivre la transcription, l'ARN Pol II doit cliver l'extrémité 3' du transcrit. Chez les eucaryotes, cette activité de clivage intrinsèque à l'enzyme est abondamment stimulée par le facteur TFIIS (revu par (Fish et Kane, 2002; Wind et Reines, 2000)). TFIIS est probablement le facteur d'élongation le mieux caractérisé. *In vitro*, TFIIS stimule l'ARN Pol II arrêtée à cliver le transcrit naissant (Izban et Luse, 1992). Ceci génère une nouvelle extrémité 3' dans l'ARNm qui pourra être utilisé par l'enzyme pour poursuivre la transcription. La récente structure cristallographique de TFIIS en complexe avec ARN Pol II par le groupe de Cramer a permis de dévoiler des détails importants quant à la fonction de TFIIS qui sont en accord avec les activités biochimiques connues de ce facteur (Kettenberger et al., 2003). Ainsi, ces chercheurs ont montré que le site catalytique de l'ARN Pol II subit des changements structuraux importants pendant la liaison de TFIIS à l'enzyme, où il entre par le pore jusqu'au site catalytique. À ce niveau, TFIIS insère deux acides aminés conservés entre les espèces, suggérant que ceux-ci participent au clivage du transcrit. D'autre part, plusieurs observations *in vivo* indiquent que TFIIS joue un rôle dans l'étape d'élongation. Par exemple, des expériences de CHIP ont localisé TFIIS au niveau des régions codantes d'un nombre de gènes dans des conditions de stress ou en présence de la drogue 6-azauracil, connue pour diminuer la concentration cellulaire en nucléotides (Exinger et Lacroute, 1992; Pokholok et al., 2002). Ainsi, la concentration de TFIIS en ChIP est modulée par le

stress et la drogue 6-azauracil. En support à la fonction de TFIIS en élongation, celui-ci interagit avec plusieurs facteurs d'élongation dont *SPT4*, *SPT5* et *SPT16* (Hartzog et al., 1998; Orphanides et al., 1999; Lindstrom et al., 2003). Finalement, des études rapportent l'existence de trois domaines indépendants dans TFIIS (Morin et al., 1996; Olmsted et al., 1998; Awrey et al., 1998). Le domaine I, à l'extrémité N-terminale, est formé de quatre hélices et cette région de TFIIS est connue pour interagir avec plusieurs facteurs d'initiation de la transcription (Ranish et al., 2003; Wery et al., 2004; Davie et Kane, 2000) (Malagon et al., 2004). Récemment, une étude a montré que le domaine N-terminal était requis pour la formation du complexe d'initiation (Kim et al., 2007). Ces données révèlent qu'en plus de son rôle en élongation, TFIIS participe dans l'étape d'initiation transcriptionnelle. Quant au domaine II, cette région forme trois hélices et elle est liée au domaine III via une courte séquence lien (*linker*) non-structurée (Olmsted et al., 1998). Finalement, le domaine III est composé de trois feuillets β antiparallèles formant un motif ruban de zinc (*zinc ribbon*). Le domaine II et le *linker* sont requis pour la liaison à l'ARN Pol II, alors que le domaine III est essentiel pour la stimulation du clivage du transcrit (Awrey et al., 1998).

1.2.1.3.2 Les facteurs d'élongation modifiant la structure de la chromatine

Bien que la chromatine représente une barrière considérable pour la transcription de l'ARN Pol II *in vitro*, ce processus est plutôt efficace *in vivo* (revu par (Kulaeva et al., 2007; Li et al., 2007)). En effet, les études révèlent que la structure de la chromatine des gènes activement transcrits est altérée par des facteurs impliqués dans le contrôle de l'élongation transcriptionnelle. Deux de ces facteurs, FACT et SPT6, sont capables de lier les histones directement (Bortvin et Winston, 1996) (Orphanides et al., 1999). De façon générale, ces deux facteurs d'élongation facilitent le passage de l'ARN Pol II à travers la chromatine en déstabilisant la structure du nucléosome ainsi qu'en réassemblant le nucléosome, suite au passage de l'enzyme (Belotserkovskaya et Reinberg, 2004). FACT a été le premier facteur à être identifié comme ayant la capacité de stimuler la transcription sur une matrice de chromatine (Orphanides et al., 1998). Ce

facteur présente une cinétique de recrutement sur les gènes similaire à celle de l'ARN Pol II et plusieurs évidences génétiques lient FACT à l'élongation *in vivo* (Saunders et al., 2003; Mason et Struhl, 2003; Orphanides et al., 1999). D'autre part, FACT agit comme un chaperon d'histones en déplaçant le dimère H2A/H2B du nucléosome (Belotserkovskaya et al., 2003). De plus, FACT est aussi requis pour le rétablissement de la structure chromatinienne, suite à la transcription *in vivo* (Schwabish et Struhl, 2004; Mason et Struhl, 2003; Kaplan et al., 2003). Pour ce qui est de SPT6, ce gène a été isolé grâce à sa capacité à rétablir la structure de la chromatine au niveau des promoteurs (Bortvin et Winston, 1996). Ces travaux ont révélé que Spt6 agit comme un chaperon d'histones jouant un rôle dans l'assemblage du nucléosome *in vitro*. Des analyses génétiques et biochimiques ont montré que Spt6 est impliqué dans l'élongation transcriptionnelle en présence de la chromatine (Hartzog et al., 1998). Spt6 interagit physiquement avec l'ARN Pol II, FACT et la sous-unité Spt5 du complexe DSIF décrit précédemment (Krogan et al., 2002). De plus, SPT6 stimule la vitesse de l'enzyme *in vitro* (Endoh et al., 2004).

Au moins deux autres complexes protéiques, PAF et Elongator, ont été purifiés initialement comme des composants associés à l'ARN Pol II engagée dans l'étape d'élongation (Belotserkovskaya et Reinberg, 2004). Le complexe PAF est un facteur d'élongation conservé chez les eucaryotes (revu par (Rosonina et Manley, 2005)). Avec l'assistance du complexe DSIF, PAF semble lier le CTD de l'ARN Pol II au niveau de la Ser 5 phosphorylée (Qiu et al., 2006). PAF joue un rôle important en modulant la liaison de plusieurs régulateurs de la chromatine au CTD phosphorylé. Chez la levure, PAF participe au recrutement du complexe contenant la méthyltransférase Set1 de l'histone H3 lysine 4 (H3-K4), appelé COMPASS, à l'ARN Pol II en élongation, facilitant ainsi la méthylation de la chromatine au sein de la région codante des gènes transcrits (Krogan et al., 2003a; Ng et al., 2003; Wood et al., 2003). Ces travaux semblent suggérer que PAF régule directement l'ubiquitinylation de l'histone H2B ainsi que la méthylation de H3-K4. Pour ce qui est de l'Elongator, ce complexe contient une activité HAT (Otero et al., 1999; Wittschleben et al., 1999; Hawkes et al., 2002). Les données concernant le rôle de ce facteur dans l'élongation transcriptionnelle sont ambiguës : si certaines études

montrent que ce facteur est impliqué en transcription et dans l'acétylation des histones (Svejstrup, 2007), d'autres suggèrent qu'il n'interagit pas avec l'ARN Pol II et qu'il ne semble pas faciliter la transcription sur une matrice de chromatine (Krogan et al., 2002; Kim et al., 2002).

1.2.1.3.3 Rôle du facteur d'élongation P-TEFb dans la transcription du VIH-1

Les études montrent que le facteur P-TEFb est non seulement nécessaire pour la synthèse de long transcrits *in vitro* mais également pour la transcription du virus de l'immunodéficience acquise humaine 1 (VIH-1) (figure 9) (Barboric et Peterlin, 2005; Price, 2000). Par ailleurs, la sous-unité cycline T de P-TEFb a été identifiée en association avec le transactivateur Tat du VIH-1 (Wei et al., 1998). De plus, deux protéines impliquées dans l'épissage, Tat-SF1 et CA150, s'associent avec le complexe formé par P-TEFb et Tat, stimulant la transcription du VIH-1 *in vivo* et *in vitro* (Zhou et Sharp, 1996; Sune et al., 1997; Zhou et al., 1998; Li et Green, 1998). Tout récemment, les travaux de l'équipe de K. Jones ont révélé que le facteur d'épissage SKIP (*c-Ski-interacting protein*) s'associe avec P-TEFb et stimule l'élongation transcriptionnelle induite par Tat du VIH-1 (Bres et al., 2005). Il est à souligner, en outre, que le promoteur du VIH-1 est grandement utilisé afin de comprendre le rôle de P-TEFb dans l'élongation.

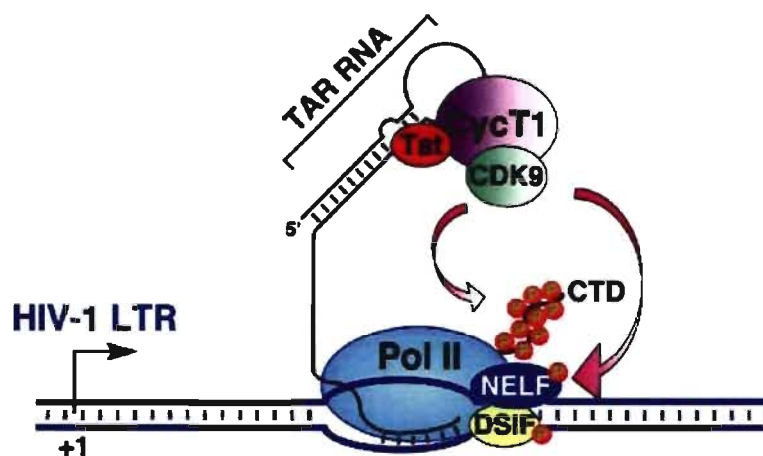


Figure 9 : P-TEFb est nécessaire pour la transcription du VIH-1. Suite à l'initiation de la transcription à partir du promoteur (LTR, *long terminal repeat*) du VIH-1, l'ARN Pol II est bloquée par l'action concertée des facteurs d'élongation négatifs DSIF et NELF. Afin de permettre à l'enzyme de poursuivre l'élongation, la protéine Tat du VIH-1 lie le facteur de l'hôte P-TEFb et le recrute sur l'ARN Pol II, formant un complexe stable avec l'ARN TAR constitué par un élément structuré en épingle à cheveux et localisée à l'extrémité 5' du transcrit viral. Par la suite, P-TEFb phosphoryle le CTD de l'ARN Pol II ainsi que DSIF et NELF, neutralisant les effets négatifs de ces deux facteurs et stimulant l'élongation transcriptionnelle. Figure tirée de (Zhou et Yik, 2006).

1.2.1.4 Les facteurs de terminaison

Les facteurs de terminaison interviennent dans le relâchement du transcrit préalablement clivé, ainsi que dans le relâchement de l'ARN Pol II de la matrice d'ADN, permettant à l'enzyme d'initier un nouveau cycle transcriptionnel (Proudfoot, 2004; Buratowski, 2005; Rosonina et al., 2006).

La terminaison de la transcription par l'ARN Pol II débute généralement par la reconnaissance, de manière co-transcriptionnelle, de séquences localisées à l'extrémité 3' du transcrit naissant (Lykke-Andersen et Jensen, 2007). Pour ce qui est de la synthèse de l'ARNm, une séquence de polyadénosines, appelée le signal de poly(A), constitue un signal de terminaison essentiel qui est reconnu par un large complexe multi-protéique de clivage/polyadénylation. La portion d'ARN en aval, générée suite au clivage du transcrit, est dégradée par les exonucléases Xrn2 chez les mammifères et Rat1 chez la levure (Luo et al., 2006; Kim et al., 2004b; Gromak et al., 2006). Ces deux exonucléases sont nécessaires afin d'obtenir une terminaison efficace. En plus du clivage au site de poly(A), certains transcrits chez les eucaryotes supérieurs sont initialement clivés en aval du site de poly(A) (Dye et Proudfoot, 2001; West et al., 2006). Le meilleur exemple de cette activité de clivage, nommée CoTC (*co-transcriptional cleavage*), est celui du gène de la β -globine (Dye et Proudfoot, 2001). La dégradation de l'ARN CoTC de ce gène par Xrn2 est aussi corrélée avec l'efficacité de terminaison de la transcription (West et al., 2004). D'autre part, plusieurs autres protéines jouant des rôles importants dans la maturation de l'extrémité 3' du transcrit sont également requises pour la terminaison de la transcription par l'ARN Pol II (Birse et al., 1998). Un de ces facteurs, Pcf1 1, semble être un joueur clé

dans ce processus. Les études montrent que Pcf11 est non seulement capable de promouvoir le relâchement de l'ARN Pol II de la matrice d'ADN *in vitro* mais il est également requis pour la terminaison *in vivo* (Birse et al., 1998; Sadowski et al., 2003; Zhang et al., 2005c; West et Proudfoot, 2008).

1.2.2 Le cycle transcriptionnel

Le cycle transcriptionnel peut être divisé en plusieurs étapes (figure 10) (revu par (Sims, III et al., 2004; Svejstrup, 2004)). Afin de transcrire un gène, l'ARN Pol II doit d'abord être recrutée au promoteur basal et s'assembler avec les FGT, pour former le complexe de pré-initiation. La double hélice d'ADN est ensuite ouverte par un processus nécessitant la présence d'ATP entre les positions -9 et +2, de part et d'autre du site d'initiation de la transcription (site +1), formant ainsi la bulle transcriptionnelle. Ceci permet à l'ARN Pol II d'initier la synthèse du transcrit en catalysant l'appariement des ribonucléotides libres au brin matrice d'ADN. Le CTD de l'ARN Pol II est par la suite phosphorylé et l'enzyme dégage le promoteur, avant d'entrer dans l'étape d'élongation. Le dégagement du promoteur est une étape dépendante du taux de transcription et l'ARN Pol II peut subir un cycle de transcriptions avortées, où des courts transcrits (entre 2 et 10 nucléotides) sont synthétisés et puis relâchés. Lors de l'étape d'élongation, les FGT, à l'exception du facteur TFIIF, sont relâchés et l'ARN Pol II peut s'associer aux facteurs d'élongation. L'étape finale du cycle de transcription est la terminaison. Lors de cette étape, le transcrit est clivé et polyadénylé, avant d'être transporté au cytoplasme pour sa traduction en protéine. Finalement, l'ARN Pol II est relâchée de la matrice d'ADN et le CTD de l'enzyme devient par la suite déphosphorylé, ce qui permet à l'ARN Pol II d'entreprendre un nouveau cycle transcriptionnel (figure 10).

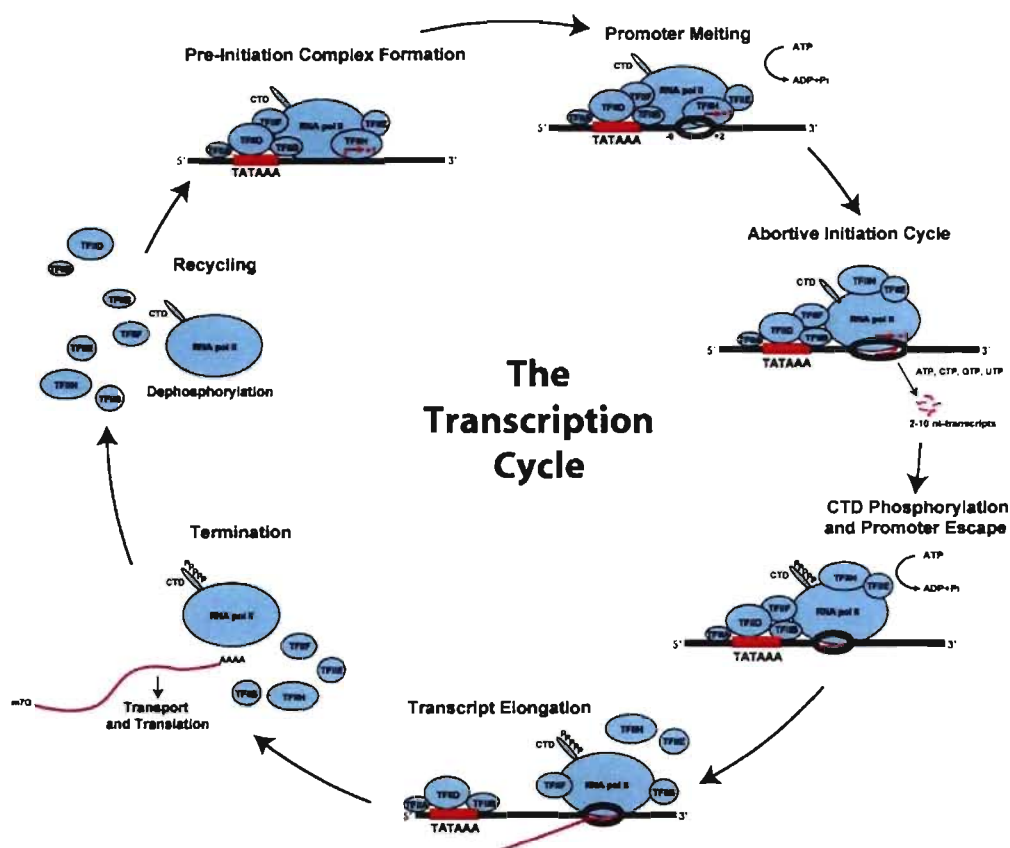


Figure 10 : Le cycle transcriptionnel de l'ARN Pol II. L'ARN Pol II et les facteurs généraux de transcription TFIIA, TFIIB, TFIID, TFIIE, TFIIIF et TFIIF s'assemblent d'abord sur le promoteur pour former le complexe de pré-initiation. En présence d'ATP, la double hélice d'ADN est ouverte, permettant à l'ARN Pol II d'initier la synthèse du transcrit. L'ARN Pol II subit ensuite un cycle de transcriptions avortées. L'ARN Pol II dégage le promoteur et entre alors dans un mode d'élongation productive. Finalement, le transcrit est clivé et polyadénylé, avant d'être transporté au cytoplasme pour sa traduction. La transcription se termine par le relâchement de l'ARN Pol II de la matrice d'ADN, permettant à l'enzyme d'être recyclée et d'entreprendre un nouveau cycle transcriptionnel. Figure modifiée de (Svejstrup, 2004).

1.2.2.1 Le promoteur basal

Au sein d'un gène, la région entourant le site d'initiation de la transcription (site +1) est souvent désignée sous le nom de promoteur basal (Butler et Kadonaga, 2002). Le promoteur basal, d'une longueur approximative de 40-50 nucléotides, a été défini biochimiquement comme une séquence d'ADN dirigeant l'assemblage du complexe de

pré-initiation, ainsi que la direction de la transcription (Juven-Gershon et al., 2006). Plusieurs éléments sur l'ADN contribuent à l'activité du promoteur basal et les propriétés spécifiques de chaque promoteur sont dictées par la présence ou l'absence de ces motifs. Les différents éléments du promoteur basal connus à ce jour incluent la boîte TATA, l'initiateur (Inr), le DPE (*downstream core promoter element*), le DCE (*downstream core element*), le MTE (*motif ten element*) et le BRE (*TFIIB recognition element*) (revu par (Juven-Gershon et al., 2006; Smale et Kadonaga, 2003)).

La boîte TATA, le site de liaison pour la sous-unité TBP du complexe multi-protéique TFIID, a été le premier élément du promoteur basal à être identifié (Breathnach et Chambon, 1981). Chez la plupart des eucaryotes supérieures, la boîte TATA ayant la séquence consensus TATAAA est localisée entre -25 et -30 par rapport au site +1 (Patikoglou et al., 1999). Le motif Inr, quant à lui, a été découvert lors de la comparaison de séquences contenant un site d'initiation de la transcription (Corden et al., 1980; Smale et Baltimore, 1989). En effet, ces études montrent que l'élément Inr chevauche le site +1. Dans les cellules de mammifères, la séquence consensus du motif Inr est Py-Py(C)-A₊₁-N-T/A-Py-Py (Py, pyrimidines C ou T) (Javahery et al., 1994; Lo et Smale, 1996). L'élément DPE a été retrouvé dans plusieurs promoteurs ne contenant pas de boîte TATA (Burke et Kadonaga, 1996; Burke et Kadonaga, 1997). Ce motif est localisé précisément de +28 à +32 et sa séquence consensus est la suivante : A/G₊₂₈-G-A/T-C/T-G/A/C (Butler et Kadonaga, 2002). Pour sa part, l'élément DCE a été identifié lors d'analyses du gène de la β -globine (Lewis et al., 2000; Lee et al., 2005). Présent dans les promoteurs contenant une boîte TATA, le DCE est localisé entre +10 et +45 et il est constitué de trois sous-éléments. L'élément MTE a été trouvé grâce à des études informatiques et fonctionnelles chez la drosophile (Ohler et al., 2002; Lim et al., 2004). Ce motif est localisé de +18 à +27. Enfin, deux éléments BRE ont été identifiés : BRE^u et BRE^d, situés respectivement en amont et en aval de la boîte TATA (Lagrange et al., 1998; Deng et Roberts, 2005). Leur séquence consensus est G/C-G/C-G/A-C-G-C-C. À ce jour, seuls la boîte TATA et l'élément Inr semblent pouvoir diriger l'initiation de la transcription de l'ARN Pol II indépendamment des autres éléments du promoteur basal (Gross et Oelgeschlager, 2006). Alors que l'élément BRE est spécifiquement reconnu par le facteur

TFIIB, tous les autres éléments du promoteur sont des sites interagissant avec le complexe TFIID (Juven-Gershon et al., 2006). Ainsi, les sous-unités TAF1 et TAF2 contactent le motif Inr, TAF6 et TAF9 lient l'élément DPE, alors que TAF1 contacte le DCE. Ces observations suggèrent que le facteur TFIID joue un rôle clé dans la reconnaissance du promoteur basal.

Des études récentes révèlent qu'il existe une grande diversité dans la structure et la fonction des promoteurs basaux (Muller et al., 2007). Ainsi, des nouvelles technologies expérimentales ont permis de découvrir et de caractériser, à l'échelle du génome, les divers promoteurs basaux, révélant que la majorité des gènes de mammifères ne se conforment pas au modèle simple, selon lequel une boîte TATA dirige la transcription à partir d'un nucléotide bien défini (Sandelin et al., 2007). En effet, chez les eucaryotes, il est maintenant évident que l'assemblage du complexe transcriptionnel sur la boîte TATA constitue l'exception plutôt que la règle, étant donné que seulement une fraction (10-20%) des promoteurs de mammifères contient une boîte TATA fonctionnelle (Cooper et al., 2006; Gershenzon et Ioshikhes, 2005). De plus, les promoteurs de mammifères comportant une boîte TATA sont souvent associés à des gènes spécifiques de tissus ou à un contexte en particulier (Schug et al., 2005). Enfin, les analyses informatiques montrent que la majorité des gènes possèdent de multiples promoteurs, lesquels contiennent plusieurs sites d'initiation de la transcription (Sandelin et al., 2007). La présence de ces nombreux paramètres des promoteurs rend compte de la réelle complexité et diversité du transcriptome et du protéome chez les mammifères. Par ailleurs, des conclusions similaires ont été rapportées, grâce à des analyses à l'échelle du génome, pour les promoteurs de *Drosophila melanogaster* et de *Arabidopsis thaliana* (Ohler et al., 2002; Ohler, 2006; Molina et Grotewold, 2005).

1.2.2.2 L'assemblage du complexe de pré-initiation au promoteur

La reconnaissance du promoteur basal par la machinerie de transcription est essentielle pour le positionnement et l'assemblage appropriés de l'ARN Pol II et des

facteurs généraux (Hahn, 2004). L'assemblage du complexe de pré-initiation au niveau du promoteur peut être considéré comme un processus en plusieurs étapes séquentielles (Svejstrup, 2004; Woychik et Hampsey, 2002). Le premier composant à reconnaître et à se fixer sur la matrice d'ADN est le complexe TFIID (via ses sous-unités TBP et TAF). Par la suite, le facteur TFIIA vient s'ajouter au complexe TFIID-promoteur, suivi de TFIIB, ce qui permet l'ancrage du complexe de pré-initiation ainsi que l'orientation de la transcription vers le site +1. Le complexe promoteur-TFIID-TFIIA-TFIIB recrute ensuite l'ARN Pol II et le facteur TFIIF. La liaison de TFIIF-ARN Pol II non seulement stabilise le complexe promoteur-TFIID-TFIIA-TFIIB mais elle est également obligatoire pour l'entrée des facteurs TFIIE et TFIIH dans le complexe de pré-initiation. Dans le but de former un complexe actif de transcription, plusieurs étapes de contrôle ont lieu lors de l'assemblage des FGT au promoteur (Svejstrup, 2004). Par exemple, un contrôle viserait la reconnaissance du complexe TFIID-TFIIA-TFIIB assemblé sur la boîte TATA par le complexe TFIIF-ARN Pol II. Ainsi, les études révèlent que TFIIF réduit l'affinité de l'ARN Pol II pour l'ADN non-spécifique *in vitro*, suggérant que ce facteur assurerait à l'enzyme de ne former des complexes productifs que sur la boîte TATA (Finkelstein et al., 1992). D'autre part, des études ont aussi suggéré un autre modèle pour l'assemblage du complexe de pré-initiation : le recrutement d'un complexe pré-assemblé nommé holoenzyme (Ossipow et al., 1995; Koh et al., 1998; Maldonado et al., 1996). Ainsi, selon ce modèle, l'holoenzyme constituée de tous les composants du complexe de pré-initiation et incluant aussi le Médiateur, serait prête à s'assembler et à initier la transcription dès qu'elle est recrutée au promoteur. Toutefois, ce modèle est contesté car le recrutement concerté des facteurs de transcription ne serait important que si la transcription était un événement en une seule étape, ce qui n'est pas le cas (Svejstrup, 2004). En effet, plusieurs études ont montré que les FGT tels TFIID, TFIIA et TFIIB ainsi que le complexe Médiateur restent au promoteur lorsque l'ARN Pol II continue l'étape d'élongation (Zawel et al., 1995; Pokholok et al., 2002; Yudkovsky et al., 2000; Svejstrup et al., 1997), permettant l'entrée rapide de nouvelles polymérases pour la ré-initiation de la transcription.

Tel que vu précédemment, la phosphorylation du CTD joue un rôle crucial en régulant les activités de l'ARN Pol II (Phatnani et Greenleaf, 2006). Le CTD contient multiples sites de phosphorylation mais la phosphorylation a lieu de façon prédominante sur la Ser 5 de l'heptapeptide répété (Dahmus, 1996). Typiquement, l'ARN Pol II participe dans la formation du complexe de pré-initiation sous sa forme non-phosphorylée, connue sous le nom d'ARN Pol IIa (Laybourn et Dahmus, 1989; Lu et al., 1991). Le CTD est nécessaire pour la formation stable du complexe de pré-initiation en interagissant avec le complexe Médiateur (Myers et Kornberg, 2000). En effet, des expériences de ChIP chez la levure révèlent que, *in vivo*, la liaison stable de TBP à la boîte TATA nécessite la fonction du complexe Médiateur (Li et al., 1999b; Kuras et Struhl, 1999). Ce dernier faciliterait la communication entre l'ARN Pol II et les régulateurs transcriptionnels liés au promoteur (Hartzog et Tamkun, 2007). Par ailleurs, les données actuelles de ChIP *in vivo* chez la levure ainsi que dans les cellules de mammifères suggèrent que l'assemblage du complexe de pré-initiation transcriptionnel est un processus hautement dynamique (Svejstrup, 2004).

1.2.2.3 L'initiation de la synthèse de l'ARNm

L'étape d'initiation de la transcription est divisée en deux phases : la première consiste en la formation d'un complexe ouvert requis pour le démarrage de la transcription et la deuxième le dégagement du promoteur.

Deux modèles différents ont été suggérés pour expliquer la catalyse de l'ouverture du promoteur par l'hélicase XPB de TFIIH. La première étude a montré que XPB contacte l'ADN en amont et en aval du site +1, engendrant un enroulement serré de l'ADN autour de l'ARN Pol II. L'hydrolyse de l'ATP induit ensuite une torsion dans l'ADN près du site +1, qui pourrait résulter dans la séparation des deux brins d'ADN afin de former la bulle transcriptionnelle (Douziech et al., 2000). Dans la deuxième étude, XPB n'a pu être photo-pontée qu'en aval du site d'initiation de la transcription (Kim et al., 2000). Ces données ont permis aux auteurs de suggérer que l'hélicase XPB catalyse,

de façon dépendante d'ATP, une rotation de l'ADN en aval à partir d'un site de rotation fixe en amont, permettant ainsi de déenrouler l'ADN au site +1.

Suite à l'ouverture de la bulle transcriptionnelle, le dégagement de l'ARN Pol II du complexe de pré-initiation coïncide avec la phosphorylation du CTD au sein de la Ser 5 par CDK7, la sous-unité kinase de TFIIH (McCracken et al., 1997; Komarnitsky et al., 2000; Schroeder et al., 2000). Le CTD phosphorylé facilite le dégagement du promoteur en se dissociant du Médiateur et en recrutant ensuite la machinerie de coiffage pour la réaction d'ajout de la coiffe en 5' du transcrit (Ares, Jr. et Proudfoot, 2005). Ceci permettra aux transcrits d'être coiffés tôt dans le cycle de transcription. La phosphorylation de la Ser 5 est aussi associée à des marques spécifiques du nucléosome à l'extrémité 5' du gène, telle la trimétylation de la lysine 4 de l'histone H3 (Hampsey et Reinberg, 2003). En somme, cette étape est importante car elle permet à l'enzyme non seulement de dégager le promoteur et de relâcher certains FGT, qui demeurent au niveau de la région promotrice, mais également de débiter l'élongation de la transcription (Kobor et Greenblatt, 2002; Pokholok et al., 2002).

1.2.2.4 L'élongation de l'ARNm

Comme discuté dans les sections précédentes, l'étape d'élongation transcriptionnelle est contrôlée par de multiples facteurs dont le rôle est d'augmenter la vitesse catalytique de l'enzyme. Au stade initial, l'ARN Pol II est réprimée par des facteurs d'élongation négatifs. DSIF et NELF coopèrent pour inhiber l'élongation de l'ARN Pol II en augmentant le temps que l'enzyme passe aux sites de pause au promoteur-proximal (Wada et al., 1998; Yamaguchi et al., 1999). Cette répression peut être enlevée par le facteur d'élongation positif P-TEFb (Price, 2000). Sa sous-unité kinase CDK9 phosphoryle en effet plusieurs sous-unités des complexes DSIF et NELF (Yamada et al., 2006). Ainsi, P-TEFb stimule le débarquement de NELF du complexe d'élongation, permettant à l'ARN Pol II de poursuivre la transcription, et il convertit DSIF en un régulateur positif de l'élongation (Saunders et al., 2006). Par ailleurs, une étude montre

que l'enzyme de coiffage peut également contrecarrer les effets négatifs des facteurs DSIF et NELF (Mandal et al., 2004). De plus, et tel que vu précédemment, CDK9 phosphoryle la Ser 2 du CTD, ce qui engendre la formation d'un complexe d'élongation productif (Ni et al., 2004). En contraste avec la phosphorylation de la Ser 5 par CDK7, la phosphorylation de la Ser 2 persiste le long de la région codante et l'extrémité 3' du gène. D'autre part, chez la levure les histones méthyltransférases Set1 et Set2 sont recrutées au niveau des complexes d'élongation de manière dépendante du statut de phosphorylation du CTD (Hampsey et Reinberg, 2003). Enfin, le CTD phosphorylé sur la Ser 2 recrute la méthyltransférase responsable de la méthylation de la lysine 36 de l'histone H3 ainsi que des enzymes de remodelage de la chromatine requis pour l'élongation chez la levure (Sims, III et al., 2004).

1.2.2.5 La terminaison de la transcription

La dernière étape de la transcription, la terminaison, implique l'arrêt de la synthèse d'ARN, le relâchement du transcrit et la dissociation de l'ARN Pol II de la matrice d'ADN. Comme mentionné auparavant, la terminaison de la transcription requiert un signal de poly(A), de multiples protéines et, chez les eucaryotes supérieurs, des séquences en aval additionnelles (Proudfoot et al., 2002). Deux modèles ont été proposés afin d'expliquer le mécanisme de terminaison de la transcription par l'ARN Pol II. Le premier, le modèle allostérique, prédit que la transcription du signal de poly(A) induit un changement de structure du complexe d'élongation de l'ARN Pol II menant à la terminaison (Logan et al., 1987). Le second, le modèle « torpille » (*torpedo*), stipule qu'une exonucléase 5'→3' dégrade le transcrit de l'ARN Pol II associé au site poly (A) de clivage, engendrant la terminaison de la transcription lorsque l'exonucléase atteint le complexe d'élongation (Proudfoot, 1989; Connelly et Manley, 1988). Il est à souligner que diverses observations suggèrent que la terminaison de la transcription par l'ARN Pol II réunit des aspects de ces deux modèles (Rosonina et al., 2006).

Une fois l'ARN Pol II relâchée de l'ADN, l'enzyme nécessite d'être recyclée pour le prochain tour de transcription (Cho et al., 1999). Ainsi, son CTD devient déphosphorylé mais le mécanisme moléculaire impliqué dans la déphosphorylation du CTD n'a pas encore été bien défini. Fcp1, une phosphatase hautement conservée chez les eucaryotes, est responsable d'enlever les groupements phosphates du CTD (Kobor et al., 1999; Archambault et al., 1997). Connue pour lier l'ARN Pol II, l'activité de Fcp1 est stimulée par le facteur TFIIF et réprimée par le facteur TFIIB (Kobor et al., 2000). La phosphatase Fcp1 de mammifères est capable de déphosphoryler l'ARN Pol II phosphorylée aux positions Ser 2 et Ser 5 du CTD (Lin et al., 2002). Toutefois, chez la levure *Schizosaccharomyces pombe*, Fcp1 déphosphoryle préférentiellement la Ser 2 phosphorylée (Hausmann et Shuman, 2002). Récemment, une famille de phosphatases du CTD de faible poids moléculaire, appelées SCP (*small CTD phosphatases*) et possédant une préférence pour la Ser 5 phosphorylée, a été identifiée dans les cellules humaines (Yeo et al., 2003). Les SCP contiennent un domaine phosphatase similaire à celui de Fcp1. Enfin, les travaux de Hampsey *et al.* ont permis de montrer qu'un composant du complexe CPF de clivage et de polyadénylation, Ssu72, est une phosphatase du CTD ayant une spécificité pour la Ser 5 phosphorylée (Krishnamurthy et al., 2004).

1.3 Couplage entre la transcription et la maturation de l'ARNm

Chez les eucaryotes, la maturation de l'ARNm est grandement facilitée grâce au couplage de cette étape à la transcription par l'ARN Pol II (figure 11) (revu par (Neugebauer, 2002; Proudfoot, 2004; Kornblihtt et al., 2004; Maniatis et Reed, 2002; Calvo et Manley, 2003)). La maturation co-transcriptionnelle implique l'ajout d'une structure coiffée à l'extrémité 5' de l'ARNm, l'épissage des séquences non codantes (introns) et la synthèse d'une queue de poly(A) à l'extrémité 3' (Proudfoot et al., 2002). Le processus de maturation de l'ARNm n'est optimal que lorsqu'il est couplé à la transcription par l'ARN Pol II. En effet, l'ARN Pol II est la seule polymérase possédant ce domaine particulier appelé CTD à l'extrémité C-terminale de sa plus grande sous-unité

Rpb1, servant de plateforme pour l'arrivée des facteurs de maturation de l'ARNm (Bentley, 2005; Phatnani et Greenleaf, 2006). Le CTD est un domaine essentiel pour la viabilité cellulaire et il joue un rôle clé dans la régulation de l'expression génique. Ainsi, le CTD est important non seulement pour la transcription mais également pour le couplage des divers événements co-transcriptionnels de maturation de l'ARNm, incluant l'ajout de la coiffe, l'épissage et la polyadénylation (Hirose et Manley, 2000; Maniatis et Reed, 2002; Sims, III et al., 2004; Meinhart et al., 2005). Les différentes étapes de maturation de l'ARNm ainsi que le rôle joué par le CTD de l'ARN Pol II dans le couplage entre la transcription et la maturation seront discutés dans la section qui suit.

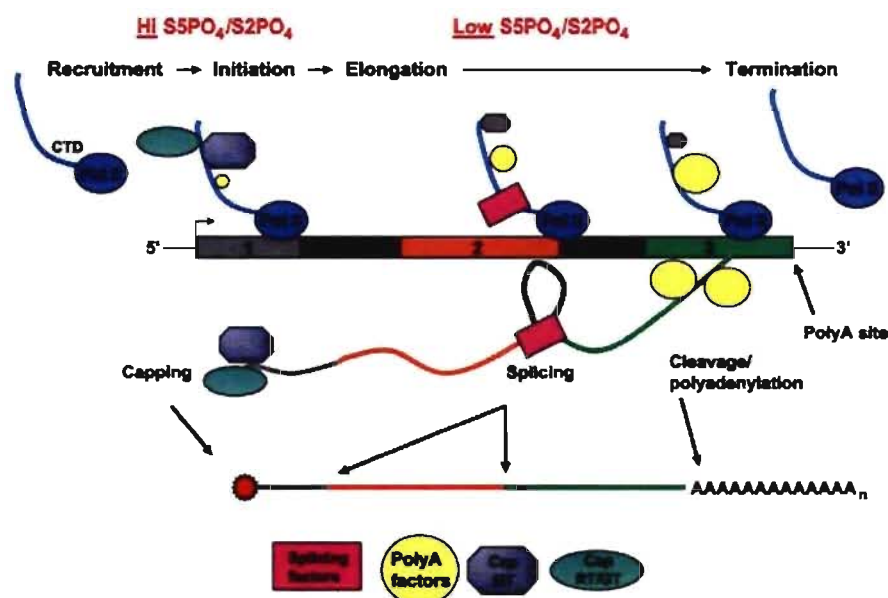


Figure 11 : Représentation schématique du couplage entre la transcription et la maturation de l'ARNm. Les facteurs impliqués dans la maturation de l'ARNm interagissent avec la machinerie de l'ARN Pol II via le domaine carboxy-terminal (CTD) de la plus grande sous-unité de l'enzyme, Rpb1. La taille des symboles représentant les divers facteurs de maturation de l'ARNm est proportionnelle au niveau de détection de ces facteurs lors d'expériences d'immunoprécipitation de la chromatine. Les enzymes de coiffage (RT, GT et MT) ainsi que les facteurs modifiant l'extrémité 3' (poly A) sont recrutés aux extrémités 5' des gènes. Pendant que l'ARN Pol II traverse le gène, des facteurs d'épissage s'associent au complexe transcriptionnel. Le degré de phosphorylation des résidus Ser 2 (S2PO₄) et Ser 5 (S5PO₄) du CTD est illustré en rouge. Le nombre d'exons est indiqué par des boîtes colorées et les introns par des boîtes noires. L'étoile en rouge représente la structure coiffée. Figure tirée de (Zorio et Bentley, 2004).

1.3.1 Les étapes de la maturation de l'ARNm

1.3.1.1 L'ajout d'une structure coiffe en 5'

La maturation de l'ARNm débute par l'ajout d'une structure coiffe à l'extrémité 5' des transcrits nouvellement synthétisés (Neugebauer, 2002; Shatkin et Manley, 2000). Trois activités enzymatiques sont requises pour la formation de cette structure coiffe (revu par (Gu et Lima, 2005). Ainsi, le triphosphate à l'extrémité 5' du premier nucléotide du transcrit est hydrolysé en un diphosphate par une ARN triphosphatase (RT); le diphosphate est ensuite ajouté d'une coiffe guanosine mono phosphate par une ARN guanylyltransférase (GT); et enfin, la base guanine est méthylée par l'ARN méthyltransférase (MT) sur la guanine en position N7, formant la structure coiffe ^{m7}GpppN (Shuman, 2001). La coiffe joue un rôle fondamental dans le métabolisme de l'ARNm, en le protégeant de la dégradation exonucléolytique, en participant à son transport vers le cytoplasme, et en facilitant l'initiation efficace de la synthèse protéique (Cougot et al., 2004).

Les enzymes impliquées dans la catalyse de la coiffe sont les premiers facteurs de maturation de l'ARNm recrutés au CTD de l'ARN Pol II lors du cycle transcriptionnel (figure 11). Ainsi, suite à l'étape d'initiation, les complexes d'élongation négatifs DSIF et NELF induisent une pause de la transcription, permettant à l'ARN Pol II phosphorylée à la position Ser 5 du CTD ainsi qu'à la sous-unité Spt5 de DSIF de recruter la machinerie de coiffage au complexe transcriptionnel (Lindstrom et al., 2003; Pei et Shuman, 2002; Pei et al., 2003; Wen et Shatkin, 1999). L'extrémité 5' du transcrit naissant (les 20 premiers nucléotides) est décorée d'une coiffe aussitôt que celui-ci laisse le pore de sortie au sein de l'ARN Pol II (Chiu et al., 2002; Mandal et al., 2004). L'ARN Pol II phosphorylée sur la Ser 5 non seulement recrute les enzymes de coiffage mais elle module également de façon allostérique leur activité (Pei et al., 2001). Par ailleurs, les travaux de Mandal *et al.* (2004) démontrent que l'enzyme de coiffage humain interagit avec le complexe NELF et participe à la modulation de l'arrêt transcriptionnel causé par

ce facteur, suggérant que la machinerie catalysant l'ajout de la coiffe joue aussi un rôle dans le contrôle de l'élongation transcriptionnelle (Mandal et al., 2004).

1.3.1.2 L'épissage alternatif

La grande majorité des gènes eucaryotes sont interrompus par des séquences non-codantes (introns) qui doivent être excisées de manière précise des ARN pré-messagers nouvellement synthétisés afin de générer des molécules d'ARNm matures et fonctionnelles (revu par (Black, 2003)). Le processus d'excision des introns et de ligation des séquences codantes adjacentes (exons), connu sous le nom d'épissage, est catalysé par une succession de deux réactions de trans-estérification au sein d'un complexe macromoléculaire connu sous le nom de splicéosome (Jurica et Moore, 2003). La plupart des introns possède des séquences consensus à leurs extrémités 5' et 3' qui sont reconnues par des composants du splicéosome et nécessaires pour la formation de ce dernier. L'assemblage du splicéosome au niveau du transcrit naissant est un processus ordonné impliquant cinq petites ribonucléoprotéines nucléaires snRNP (U1, U2, U4, U5 et U6) ainsi que de nombreux facteurs accessoires, incluant les protéines de la famille SR (*serine/arginine-rich*) (Will et Luhrmann, 2001; Patel et Steitz, 2003). La catalyse de la réaction d'épissage comprend une série d'interactions ARN-ARN, ARN-protéine et protéine-protéine, aboutissant en l'excision intronique et la ligation exonique. Les snARN U2, U4, U5 et U6 jouent un rôle central dans la réaction d'épissage. Brièvement, le site consensus donneur d'épissage situé en 5' de l'intron est reconnu par U1 alors que le snARN U2 se fixe au niveau du point d'embranchement. Un complexe formé par des protéines et les snARN U4/U5/U6 se lie ensuite au complexe formé par U2. Par la suite, le snARN U6 prend la place du snARN U1 au niveau du site donneur d'épissage et forme des interactions avec le snARN U2. La catalyse des deux réactions de trans-estérification caractéristiques de la réaction d'épissage peut alors avoir lieu (Black, 2003).

Dans le cas de l'épissage alternatif, un seul ARN pré-messager peut générer deux ARNm matures ou plus, dépendamment de la combinaison des exons épissés. Ce

mécanisme d'épissage alternatif semble jouer un rôle fondamental dans la régulation génique chez les eucaryotes supérieurs (Black, 2003; Matlin et al., 2005; Blencowe, 2006). Grâce à ce mécanisme, plusieurs protéines différentes peuvent être générées à partir d'un même transcrit d'ARN, engendrant une grande diversité tant au niveau protéomique que fonctionnel. Des analyses récentes au moyen de puces à ADN au niveau des chromosomes 21 et 22 révèlent que près de 80% des gènes humains subissent l'épissage alternatif (Kampa et al., 2004).

L'épissage alternatif est intimement lié à la transcription par l'ARN Pol II (revu par (Proudfoot et al., 2002; Kornblihtt et al., 2004)). Plusieurs études ont révélé que la transcription module l'épissage et que, réciproquement, l'épissage module la transcription (Das et al., 2006; Hicks et al., 2006; Kwek et al., 2002; Fong et Zhou, 2001). Le type de promoteur, la vitesse d'élongation, les activateurs transcriptionnels, ou encore, un facteur de remodelage de la chromatine, semblent affecter l'épissage alternatif *in vivo* (Batsche et al., 2006; de la et al., 2003; Listerman et al., 2006; Nogues et al., 2003; Lacadié et al., 2006). D'autre part, l'excision des introns au sein des gènes réduit le taux de transcription alors que les sites d'épissage à proximité du promoteur semblent stimuler la réaction transcriptionnelle (Furger et al., 2002). De plus, des travaux récents ont impliqué une grande variété de facteurs d'épissage et de protéines liant l'ARN dans les différents événements transcriptionnels. Tel que vu précédemment, le snARN U1, un composant important du splicéosome, a été trouvé associé au complexe TFIIF (Kwek et al., 2002). Cette étude a aussi montré que l'étape de ré-initiation modulée par TFIIF est stimulée par un site d'épissage 5' à proximité du promoteur spécifiquement reconnu par le snARN U1. En outre, un groupe de protéines SR est recruté au sein des gènes via une liaison directe avec le CTD phosphorylé sur la Ser 2 (Zorio et Bentley, 2004). En particulier, la protéine U2AF65 s'associe directement aux transcrits naissants, pouvant induire une diminution de la pause par l'ARN Pol II *in vitro* (Ujvari et Luse, 2004). Par ailleurs, une étude récente montre que les protéines SR sont impliquées dans le couplage entre la transcription et l'épissage (Das et al., 2007). Finalement, le facteur d'élongation P-TEFb semble interagir avec des composants de la machinerie d'épissage, tels les snRNP (Fong et Zhou, 2001).

1.3.1.3 Le clivage et la polyadénylation en 3'

Les facteurs de clivage et de polyadénylation sont impliqués dans la formation de l'extrémité 3' du transcrit. Chez les mammifères, cinq protéines s'assemblent pour moduler, d'abord, le clivage endonucléolytique de l'extrémité 3' du transcrit naissant, et ensuite, l'ajout d'un segment d'adénosines, appelé queue de poly(A) (revu par (Proudfoot, 2004)). Ces facteurs sont la poly(A) polymérase (PAP), CPSF (*cleavage and polyadenylation specificity factor*), CstF (*cleavage stimulation factor*) CFIm (*cleavage factor I*) et CFIIIm (*cleavage factor II*). La composition en sous-unités des facteurs multiprotéiques de clivage CPSF et CstF est bien connue ainsi que leurs sites de contact sur l'ARN. Ces facteurs lient le signal de poly(A) constitué par la séquence hautement conservée AAUAAA et la séquence riche en GU en aval de l'élément DES (*downstream sequence element*) (Colgan et Manley, 1997; Zhao et al., 1999). Des facteurs homologues chez *S. cerevisiae* ont également été caractérisés et ils ont été montrés pour interagir avec des éléments de séquences constituant le signal de poly(A) de levure. Chez cet organisme, deux facteurs de clivage, appelés CFIA et CPF, ont été purifiés et ils partagent des sous-unités avec respectivement les facteurs CstF et CPSF de mammifère (Proudfoot et O'Sullivan, 2002). CFIA contient quatre sous-unités, dont deux homologues des sous-unités CstF77 et CstF64 de mammifère. Toutefois, CFIA possède également les facteurs additionnels Pcf11p et Clp1p (de et al., 2000). Par ailleurs, les complexes CFIA et CPF sont également impliqués dans la maturation de l'extrémité 3' d'autres classes de transcrits générés par l'ARN Pol II, notamment les snARN et les snoARN (Fatica et al., 2000; Morlando et al., 2002). Pour ce qui est de la polyadénylation, les études montrent que le clivage est couplé à la synthèse de la queue de poly(A) par la PAP, laquelle a également besoin des facteurs accessoires CPSF et PABPN1 (*poly A binding protein*) (Kerwitz et al., 2003). Chez les eucaryotes, la queue de poly(A) agit en tant que stabilisateur de l'ARNm. En effet, ce segment de poly(A) joue plusieurs rôles, incluant le transport de l'ARNm vers le cytoplasme, la stabilisation du transcrit en procurant une protection de l'extrémité 3' contre une dégradation par des nucléases et une augmentation de l'efficacité de l'initiation de la traduction (Dreyfus et Regnier, 2002).

La première évidence de communication entre la maturation de l'ARNm et la transcription a été la découverte que le site de poly(A) était nécessaire pour la transcription par l'ARN Pol II (Logan et al., 1987; Whitelaw et Proudfoot, 1986; Zaret et Sherman, 1982). Le couplage entre la formation de l'extrémité 3' et la terminaison assure que l'ARN Pol II ne soit relâchée de la matrice d'ADN qu'une fois le transcrit complètement synthétisé. De plus, la vitesse d'élongation transcriptionnelle semble pouvoir influencer le choix du site de poly(A) (Cui et Denis, 2003). D'autre part, le CTD interagit directement avec plusieurs facteurs impliqués dans la maturation de l'extrémité 3' de l'ARNm (Zorio et Bentley, 2004). Toutefois, il n'est pas clair comment ces contacts contribuent aux processus de clivage et de polyadénylation des transcrits par l'ARN Pol II. Plusieurs études chez la *Drosophile* et la levure suggèrent que la phosphorylation du CTD à la position Ser 2 est importante pour la liaison des facteurs poly(A) ainsi que pour l'efficacité de clivage (Ni et al., 2004; Licatalosi et al., 2002; Ahn et al., 2004). Ainsi, deux facteurs impliqués dans la maturation de l'extrémité 3', Pcf11 et Rtt103, lient préférentiellement le CTD phosphorylé sur la Ser 2 via leurs domaines de liaison au CTD (Licatalosi et al., 2002; Kim et al., 2004b). De plus, l'inactivation de la kinase CTK1 impliquée dans la phosphorylation de la Ser 2 du CTD chez la levure (Jones et al., 2004) inhibe le recrutement des facteurs de maturation de l'extrémité 3' de l'ARNm (Ahn et al., 2004).

Le recrutement des facteurs de clivage et de polyadénylation a lieu non seulement à l'extrémité 3' mais également au promoteur ainsi qu'au niveau de la région codante du gène (Bentley, 2005). Des expériences de ChIP chez la levure ont révélé que ces facteurs sont recrutés sur le transcrit progressivement, avec une faible mais significative quantité à l'extrémité 5', laquelle augmente vers l'extrémité 3' du gène (Licatalosi et al., 2002; Kim et al., 2004a; Calvo et Manley, 2005). Cette distribution 5'-3' des facteurs de clivage/polyadénylation, prédominante vers l'extrémité 3', est similaire à celle des facteurs d'exportation de l'ARNm mais opposée à la distribution observée pour les enzymes de coiffage, lesquelles sont surtout présentes au niveau de l'extrémité 5' du gène (Bentley, 2005).

1.3.2 Le rôle du CTD de l'ARN Pol II

Tous les événements co-transcriptionnels mentionnés précédemment nécessitent un grand nombre de facteurs protéiques, lesquels sont recrutés de façon coordonnée par le CTD de l'ARN Pol II (Buratowski, 2003). Ce dernier doit posséder une structure flexible et dynamique. La flexibilité structurale du CTD a été révélée par les études cristallographiques de l'ARN Pol II de levure, selon lesquelles le CTD est directement adjacent au sillon de sortie de l'ARNm par une région de liaison (*linker*) longue et mobile (Cramer et al., 2001). Pour sa part, la nature dynamique du CTD est supportée par des données structurales récentes grâce à des peptides du CTD en complexe avec des protéines, incluant l'enzyme de coiffage de levure Cgt1, l'enzyme prolyl isomérase humaine Pin1, et le facteur de maturation de l'extrémité 3' de l'ARNm Pcf11 (Meinhart et al., 2005). La structure du CTD sous sa forme non-phosphorylé semble relativement compacte (Meredith et al., 1996). Par contre, la phosphorylation semble résulter en une structure du CTD étendue, laquelle est plus sensible à la digestion par des protéases (Laybourn et Dahmus, 1989; Zhang et Corden, 1991). Ces résultats suggèrent que la phosphorylation du CTD affecte non seulement son état conformationnel, mais qu'elle génère également des interfaces de reconnaissance distinctes permettant la liaison de différents partenaires protéiques.

Le couplage entre la transcription et la maturation de l'ARNm semble être régulé par la phosphorylation du CTD (Proudfoot et al., 2002; Phatnani et Greenleaf, 2006). Par exemple, un peptide du CTD contenant la Ser 5 phosphorylée est capable de stimuler une des enzymes de coiffage, l'ARN guanylyltransférase (Ho et Shuman, 1999). D'autre part, des études *in vitro* ont montré que plusieurs facteurs impliqués dans la maturation de l'extrémité 3' du transcrit peuvent s'associer avec le CTD phosphorylé sur la Ser 2 (Barilla et al., 2001; Licatalosi et al., 2002). Enfin, des facteurs de clivage et de polyadénylation sont stimulés par les formes hypophosphorylée et hyperphosphorylée de l'ARN Pol II (Hirose et Manley, 1998; Proudfoot et O'Sullivan, 2002).

1.4 Objectifs de recherche

Le premier objectif de mon travail de doctorat était de purifier des complexes protéiques solubles interagissant avec la machinerie de l'ARN Pol II dans les cellules humaines afin d'établir des réseaux d'interactions protéine-protéine impliqués dans la transcription par l'ARN Pol II. À l'aide de la technique de double chromatographie d'affinité couplée à la MS, nous avons d'abord pu purifier un complexe comprenant l'ARN Pol II, les FGT TFIIB et TFIIF ainsi que Fcp1, la phosphatase de l'ARN Pol II. Par la suite, trente-deux protéines étiquetées ont permis de générer un réseau de 820 interactions protéiques validées bioinformatiquement. Tous les facteurs étiquetés que nous avons isolés et testés sont actifs dans des réactions de transcription *in vitro* et ils ont été localisés le long de gènes humains par la technique de CHIP *in vivo*, indiquant que des complexes protéiques fonctionnels transcriptionnels ont été purifiés.

Mon second objectif était d'identifier et de caractériser de nouveaux régulateurs de la machinerie transcriptionnelle de l'ARN Pol II. Par l'approche de double chromatographie d'affinité, nous avons purifié plusieurs polypeptides à fonction inconnue, dont trois associés à l'ARN Pol II, que nous avons nommés *RNA polymerase II-associated protein* (RPAP) 1, 2 et 3. Nos résultats montrent que les protéines RPAPs sont physiquement et fonctionnellement associées à l'ARN Pol II, formant une interface entre l'enzyme et des protéines chaperons et d'échafaudage. Les rôles de BCDIN3 et LARP7, des polypeptides à fonction inconnue trouvés en association avec le facteur d'élongation P-TEFb et le snARN 7SK, ont aussi été étudiés. Nous avons montré que BCDIN3 est l'enzyme de coiffage de 7SK. BCDIN3 a été nommée *methylphosphate capping enzyme* (MEPCE). De plus, nous avons pu déterminer que BCDIN3 et LARP7 jouent un rôle dans la stabilité de 7SK dans les cellules humaines, contribuant ainsi au maintien de P-TEFb dans un état transcriptionnel inactif.

CHAPITRE 2

RPAP1, UNE NOUVELLE PROTÉINE ASSOCIÉE À L'ARN POLYMÉRASE II PURIFIÉE À PARTIR DE SOUS-UNITÉS RECOMBINANTES SAUVAGES ET MUTANTES

2.1 Préambule

Dans l'article présenté au chapitre 2 pour lequel je suis co-auteure principale, j'ai mis au point la méthode décrite à la figure 1A permettant la purification de complexes protéiques par double chromatographie d'affinité, en collaboration avec Mahel Zeghouf. J'ai montré que notre système d'expression inductible permettait d'exprimer des protéines portant des étiquettes TAP à des niveaux physiologiques dans les cellules de mammifères (figure 1B). À l'aide de la lignée Rpb11-TAP, j'ai purifié pour la première fois un complexe comprenant l'ARN Pol II, les facteurs TFIIB et TFIIF, la phosphatase Fcp1 et un polypeptide à fonction inconnue, que nous avons nommé *RNA polymerase II-associated protein 1* (RPAP1) (figure 2A). J'ai réalisé le clonage et l'établissement des lignées stables TFIIB-TAP et RPAP1-TAP et j'ai confirmé la présence des différents composants dans le complexe étudié (figure 2B), conjointement avec Dania Baali et Marie-France-Langelier. J'ai montré que deux molécules de RAP30 étaient présentes dans le complexe purifié (figure 2C). J'ai aussi montré que l'ARN Pol II purifiée par la méthode TAP était sous la forme hypophosphorylée (figure 2D) et qu'elle était active en transcription *in vitro* (figure 5A). Finalement, j'ai déterminé les partenaires du nouveau polypeptide RPAP1-TAP et j'ai effectué les analyses bioinformatiques pour les protéines présentant des homologies avec RPAP1 chez *S. cerevisiae* et *D. melanogaster*, en collaboration avec Dominique Bergeron (figure 3A). J'ai rédigé l'article avec Marie-France Langelier et le Dr Benoît Coulombe.

Marie-France Langelier a effectué les expériences de caractérisation biochimique des mutants de l'ARN Pol II présentées aux figures 6 et 7 ainsi que l'expérience décrite à

la figure 5B. Mahel Zeghouf, stagiaire postdoctorale chez le Dr Jack Greenblatt, a réalisé le clonage et l'établissement des lignées cellulaires stables Rpb11-TAP et RAP30-TAP. L'identification des protéines par MS a été effectuée par Veronica Canadien, Bryan Beattie et Dawn Richards (figure 2B). La figure 3B a été réalisée par l'équipe du Dr Timothy Hughes (Sanie Mnaimneh, Armaity Davierwala et Jeff Pootoolal) et la figure 4C par Mark Chandy, étudiant chez le Dr Jerry Workman. Diane Forget a effectué l'essai de retard sur gel (figure 4A) et Marilena Cojocararu les expériences d'immunoprécipitation de la chromatine (figures 4B et 8).

2.2 Article

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 using recombinant wild type and mutated polymerase subunits. **Mol. Cell. Biol.**, 24(16):7043-7058.

* Ces deux auteures ont contribué de façon équivalente à cet article et sont considérées comme co-premières auteures.

RPAP1, a Novel Human RNA Polymerase II-Associated Protein Affinity Purified with Recombinant Wild-Type and Mutated Polymerase Subunits

Célia Jeronimo,^{1†} Marie-France Langelier,^{1†} Mahel Zeghouf,² Marilena Cojocaru,¹ Dominique Bergeron,¹ Dania Baali,¹ Diane Forget,¹ Sanie Mnaimneh,² Armaity P. Davierwala,² Jeff Pootoolal,² Mark Chandy,^{3,4} Veronica Canadien,⁵ Bryan K. Beattie,⁵ Dawn P. Richards,⁵ Jerry L. Workman,³ Timothy R. Hughes,² Jack Greenblatt,² and Benoit Coulombe^{1*}

Laboratory of Gene Transcription, Institut de Recherches Cliniques de Montréal, Montréal, Québec H2W 1R7,¹ Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L6,² and Affinium Pharmaceuticals, Toronto, Ontario M5J 1V6,⁵ Canada; Stowers Institute for Medical Research, Kansas City, Missouri 64110³; and Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033⁴

Received 24 December 2003/Returned for modification 9 February 2004/Accepted 14 May 2004

We have programmed human cells to express physiological levels of recombinant RNA polymerase II (RNAPII) subunits carrying tandem affinity purification (TAP) tags. Double-affinity chromatography allowed for the simple and efficient isolation of a complex containing all 12 RNAPII subunits, the general transcription factors TFIIB and TFIIF, the RNAPII phosphatase Fcp1, and a novel 153-kDa polypeptide of unknown function that we named RNAPII-associated protein 1 (RPAP1). The TAP-tagged RNAPII complex is functionally active both *in vitro* and *in vivo*. A role for RPAP1 in RNAPII transcription was established by shutting off the synthesis of Ydr527wp, a *Saccharomyces cerevisiae* protein homologous to RPAP1, and demonstrating that changes in global gene expression were similar to those caused by the loss of the yeast RNAPII subunit Rpb11. We also used TAP-tagged Rpb2 with mutations in fork loop 1 and switch 3, two structural elements located strategically within the active center, to start addressing the roles of these elements in the interaction of the enzyme with the template DNA during the transcription reaction.

RNA polymerase II (RNAPII) is the multisubunit enzyme that synthesizes all mRNA precursors in eukaryotes. RNAPII is highly conserved among species, and in humans, RNAPII consists of 12 subunits, named Rpb1 to Rpb12 (16, 88). The two largest subunits, Rpb1 (220 kDa) and Rpb2 (140 kDa), form the enzyme's catalytic center and are homologous to the β' and β subunits of bacterial RNAP, respectively. Five subunits, Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12, are also found in RNAPI and RNAPIII. Rpb3 and Rpb11 are homologous to the α_2 homodimer involved in bacterial RNAP assembly. Rpb9 was attributed a role in elongation through its action at DNA arrest sites (3). In *Saccharomyces cerevisiae*, Rpb4 and Rpb7 form a subcomplex that can dissociate from the enzyme upon changes in environmental conditions. Under active growth conditions, most yeast RNAPII molecules do not contain the Rpb4-Rpb7 dimer, which primarily associates during the stationary phase or following stress (11). Functional studies of human RNAPII have been limited due to the lack of appropriate systems for purifying variant forms of the human enzyme.

The availability of crystal structures of both yeast (17, 18, 26, 30) and bacterial (8, 57, 58, 84, 89) RNAPs has been invaluable

for understanding many of the molecular features of the transcription reaction. For example, the structure of elongating RNAPII has revealed the positioning of the RNA-DNA duplex during the transcription reaction (30). The structures available support a model in which the DNA enters the enzyme through a channel formed by a pair of "jaws" before accessing a deep cleft, at the bottom of which is buried the Mg^{2+} -ion-containing active site; the DNA then turns by about 90° along a wall where the upstream end exits the enzyme (30). Many loops and helices either directly contact or closely approach the RNA-DNA duplex, thus suggesting putative functions for these structural elements in the transcription reaction. Again, the lack of an efficient method for purifying variant forms of RNAPII with mutations within specific structural elements has precluded any detailed structure-function analysis of human RNAPII.

The activity of RNAPII at the various stages of the transcription reaction is regulated by several proteins, including the general transcription factors (14, 32, 65). TATA-binding protein (TBP), the TATA box-binding subunit of TFIID, recognizes the TATA box of promoters, and its binding induces a bend of approximately 80° in the DNA (45, 46). TFIIB associates with the TBP-promoter complex and participates in RNAPII recruitment (54). TFIIF, composed of subunits RAP74 and RAP30, binds directly to RNAPII and also helps to recruit the enzyme to the promoter (15, 80). TFIIF can induce further bending and wrapping of the promoter DNA around RNAPII during preinitiation complex formation (72).

* Corresponding author. Mailing address: Laboratory of Gene Transcription, Institut de Recherches Cliniques de Montréal, 110 avenue des Pins Ouest, Montréal, Québec, Canada H2W 1R7. Phone: (514) 987-5662. Fax: (514) 987-5663. E-mail: [REDACTED]

† C.J. and M.-F.L. contributed equally to this work.

Both TFIIE and TFIIH were shown to participate in promoter melting at the transcription initiation site (38, 66). TFIIE, which is also composed of two subunits, TFIIE56/α and TFIIE34/β, stimulates the helicase and kinase activities of TFIIH (53, 62, 63). The helicase activity of TFIIH is required to fully open the promoter DNA between nucleotides (nt) -9 and +2 (37, 39, 44, 83), and its kinase activity phosphorylates the carboxy-terminal domain (CTD) of Rpb1 (21, 53, 75-77). The phosphorylation of the CTD on Ser5 by TFIIH has been shown to be required for the transition from initiation to early elongation and allows for the recruitment of the capping enzyme (35, 73). In yeast, the Ctk1 kinase phosphorylates the CTD on Ser2 at a later stage of the elongation phase (10), leading to the recruitment of mRNA processing factors (78). The CTD phosphatase Fcp1 is able to dephosphorylate the Rpb1 CTD on Ser2 in order to recycle RNAPII for reinitiation (10, 48, 49).

Unlike the case for all of the general transcription factors and the prokaryotic RNAP, it has been impossible to date to purify functionally active forms of mammalian RNAPII reconstituted from its recombinant subunits. We now report the purification of a functionally active human RNAPII by the use of doubly tagged subunits expressed in mammalian cells. Using the TAP procedure (71), we isolated a human RNAPII-containing complex which comprises the 12 subunits of core RNAPII, TFIIB, TFIIF, Fcp1, and RPAP1, a novel polypeptide of unknown function. The tagged RNAPII complex is functionally active both *in vitro* and *in vivo*. We capitalized on this purification procedure to generate two variant forms of human RNAPII by using TAP-tagged Rpb2. The first mutant has a deletion in fork loop 1, a structure that is expected to participate in maintaining the open state of the transcription bubble in elongating RNAPII (18, 30) or during RNA-DNA strand separation (85). The second mutant carries a triple alanine substitution in switch 3, a loop that directly contacts the DNA template within the DNA-RNA hybrid. Our results allowed us to detail the roles of fork loop 1 and switch 3 in transcriptional mechanisms both *in vitro* and *in vivo*.

MATERIALS AND METHODS

Protein factors and antibodies. Recombinant yeast TBP (42), human TFIIB (31), RAP74 and RAP30 (22), TFIIE56/α and TFIIE34/β (64, 67, 81), highly purified TFIIH (20, 27), and calf thymus RNAPII (36) were purified as previously described. Antibodies raised against Rpb1 (N-20) were obtained from Santa Cruz Biotechnology.

TAP of human RNAPII-containing complex. The cDNAs encoding human Rpb2 (kindly provided by Marc Vigneron, IGBMC, Strasbourg, France) (1), Rpb4 (Invitrogen; accession number BE883306), Rpb7 (Invitrogen; accession number BE256210), Rpb11 (Invitrogen; accession number AA814184), TFIIB (31), and RAP30 (22) were cloned into the mammalian expression vector pMZ1 (61) in such a way that each polypeptide carries a tandem affinity purification (TAP) tag at its C terminus (71). Stable cell lines carrying these plasmids in EcR-293 cells (Invitrogen) were produced by transfection, using the calcium phosphate method. Induction for 24 h with 1 to 3 μM ponasterone A, an ecdysone analog, was used to express the TAP-tagged subunits of RNAPII, TFIIB, and RAP30 at near-physiological levels. The TAP-tagged complexes were then purified as previously described (71). The TAP eluates were concentrated and stored at -80°C until use.

Protein identification. The TAP eluates were run in sodium dodecyl sulfate (SDS) gels and silver stained. The protein bands were excised from the gels, reduced, alkylated, and subjected to in-gel tryptic digestion as previously described (50). The resulting tryptic peptides were purified and identified by matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (50).

DNA microarray analysis of Tet-promoter mutants. Tet-promoter alleles were constructed as previously described (86). Mutant and isogenic wild-type cultures were grown in parallel in SC medium with 10 μg of doxycycline/ml for a total of 22 h. Dilutions were made at 16 h to ensure that mutant and wild-type optical densities matched as closely as possible to 10⁷ cells/ml at harvest. The cells were pelleted by a 2-min centrifugation in a room-temperature table-top centrifuge and then frozen immediately in liquid nitrogen. Total RNAs were prepared by hot acidic phenol-chloroform extraction followed by ethanol precipitation. Poly(A)⁺ RNA purification and cDNA labeling were performed as described previously (55). Microarrays were composed of 70-mer oligonucleotides, each specific to a different yeast gene (Operon Technologies), which were spotted onto poly-L-lysine-coated slides. Spotting buffer, blocking, hybridization, and washing were according to a procedure described by Hegde and collaborators (34). Each array was normalized by using grid-by-grid lowess smoothing (87). All measurements were taken in fluor-reversed pairs [i.e., each time a mutant was analyzed, it was hybridized to two arrays, and data were combined by averaging the normalized log(ratio)].

ChIP. Chromatin immunoprecipitation (ChIP) assays were performed as previously described (60, 82), with the following minor modifications. Approximately 2 × 10⁸ cells were cross-linked by the use of formaldehyde, and chromatin fragments of an average size of 0.3 to 0.5 kbp were prepared by four sonication treatments (Fisher Sonic Dismembrator; 20-s burst at 50% maximum power). For the immunoprecipitation step, immunoglobulin G (IgG)-Sepharose beads (Amersham Biosciences) were first blocked with a 5-mg/ml solution of bovine serum albumin in phosphate-buffered saline and then incubated with the sample for 1 h. After reversal of the cross-links, PCR analyses were performed with both the immunoprecipitated and input DNAs. Primer sets designed for amplifying both the promoter and the 5' untranslated region of the CCNA1 gene were used.

For the experiments shown in Fig. 8, relative occupancy was evaluated by quantitative PCR (Q-PCR) as previously described (5), with minor modifications. The quantification of the target DNA in the immunoprecipitated sample was carried out by generating a standard curve with a fivefold dilution series of the nonimmunoprecipitated sample (input DNA) for each couple of primer sets and DNA. The apparent immunoprecipitation efficiency was calculated by dividing the amount of the DNA fragments spanning the promoter or the coding region of the specified genes by the amount of the DNA fragment from the control region in the immunoprecipitated sample. The following cell lines were analyzed: Rpb2 wt, Rpb2 fork1 Δ458-459, Rpb2 sw3-1078, and Rpb2 fork1 Δ453-465. Q-PCR was performed by using the Mx4000 Multiplex Quantitative PCR system (Stratagene). The PCR protocol was applied according to the manufacturer's specifications (Qiagen). Primers were designed for both the promoter and transcribed regions of the GNB2L1 (positions -200 to +55 and +998 to +1165, respectively) and FTL (positions -157 to +33 and +955 to +1176, respectively) genes and also for an internal sequence of the expressed sequence tag located at chr17:49926700-49927999 as a control. Oligonucleotide sequences are available upon request. All amplified DNA fragments were unique, as determined by a BLAST search on the April 2003 assembly version of the UCSC Genome Browser (<http://www.genome.ucsc.edu>), by a melting curve analysis, and by agarose gel electrophoresis (data not shown).

Gel mobility shift assays. Gel mobility shift assays were performed as previously described (52). Complexes were assembled by using highly purified TBP (80 ng), TFIIB (60 ng), RAP30 (320 ng), RAP74 (640 ng), TFIIE34 (160 ng), TFIIE56 (320 ng), and calf thymus RNAPII or the TAP-tagged human RNAPII complex (the amounts of RNAPII used in the various experiments are detailed in the figure legends).

Histone tail peptide binding assay. Histone tail peptide binding reactions were performed to determine whether the complex binds specifically to acetylated histone tails. For each reaction, 10 ng of biotinylated H3 peptides was bound to streptavidin-conjugated magnetic beads at 25°C for 30 min. The beads were then incubated with 1 nM RNAPII complex (Rpb11 TAP purified or calf thymus) for 1 h at 30°C. The beads were pelleted with a magnetic particle concentrator (MPC-S; Dynal) and the supernatant was removed. After being washed twice, the beads were resuspended in SDS-polyacrylamide gel electrophoresis (SDS-PAGE) lysis buffer and subjected to SDS-PAGE analysis. The complex was detected by immunoblotting with an anti-Rpb1 antibody (N-20; Santa Cruz).

Transcription assay. Transcription reactions were performed as described previously (6). TBP (30 ng), TFIIB (30 ng), RAP30 (30 ng), RAP74 (65 ng), TFIIE34 (40 ng), TFIIE56 (60 ng), TFIIH (4 ng), and calf thymus RNAPII or the TAP-tagged human RNAPII complex (the amounts of RNAPII used for the various experiments are detailed in the figure legends) were incubated with 500 ng of linearized DNA template containing the adenovirus major late (AdML) promoter from nt -50 to +10 fused to a G-less cassette. Under these conditions, a 391-nt runoff transcript was produced.

Abortive initiation assay. Abortive initiation assays were performed as described previously (52), with the following modifications. The DNA template (12 ng), either premelted in the -9 to $+2$ region or fully double-stranded, was incubated with TBP (60 ng), TFIIB (30 ng), RAP30 (30 ng), RAP74 (65 ng), TFIIE34 (40 ng), TFIIE56 (60 ng), and the TAP-tagged human RNAPII complex (the amounts of RNAPII used for the various experiments are detailed in the figure legends) in a 20- μ l reaction mixture containing 750 μ M ATP, 750 μ M CTP, 10 μ M UTP, 2.5 μ Ci of [α - 32 P]UTP, 12.5 mM $MgCl_2$, 3 mM EGTA, and 0.84 U of RNase inhibitor/ml. The transcripts were analyzed in a 23% polyacrylamide denaturing gel containing 7 M urea.

Elongation assay. The template was prepared by annealing two complementary oligonucleotides of random sequences, one of which carried an extension of 15 CMP as described previously (66). Typically, the template (9.3 ng) was incubated with 88 ng of TAP-tagged human RNAPII Rpb2 wt or 440 ng of TAP-tagged human RNAPII Rpb2 fork1 (Δ 458-459)-TAP, and the reactions were performed as described previously (66). The transcripts were analyzed in 18% polyacrylamide-9% urea denaturing gels.

Nucleotide sequence accession number. The RPAP1 name and sequence were registered in the HUGO database under accession number NP_056355.

RESULTS

TAP of human RNAPII-containing complex. We have adapted the TAP procedure (69, 71), originally developed to isolate protein complexes from yeast, for the purification of human RNAPII and general transcription factors. Native conditions were maintained during the TAP procedure through the use of buffers with a low salt concentration (0.1 M Na^+) and a neutral pH. As shown in Fig. 1A, which summarizes the TAP procedure, full-length cDNAs encoding polypeptides of interest were cloned in an ecdysone-inducible mammalian expression vector (61). The vector was engineered to encode polypeptides carrying the following at the C terminus: a TAP tag consisting of two copies of the IgG-binding domain of *Staphylococcus aureus* protein A, a cleavage site for the tobacco etch virus protease, and a calmodulin-binding peptide (CBP) (71). The tagged polypeptides were expressed in EcR-293 cells at near-physiological levels after a 24-h induction with the ecdysone analog ponasterone A. The concentration of ponasterone A required to obtain near-physiological levels of expression of the tagged polypeptides was determined in a pilot experiment in which we compared the levels of TAP-tagged polypeptides with the endogenous polypeptide by Western blotting (Fig. 1B). Whole-cell extracts were prepared from the ponasterone A-induced cell lines and were used in the double-affinity chromatography procedure, which consisted of successive steps with IgG-Sepharose and calmodulin beads (71). The eluates were frozen and analyzed by different assays.

We first expressed the Rpb11 subunit of RNAPII in EcR-293 cells and used the lysate of ponasterone A-induced cells for double-affinity purification. The eluate was subjected to SDS-PAGE and the bands were identified by mass spectrometry. A total of 17 polypeptides that are not present in mock-induced cells and that represent major components of the eluate were identified (Fig. 2A). The other major bands that can be visualized on the gel have all been identified and correspond to proteins that bind nonspecifically to our affinity columns (data not shown). The complex contained the 12 subunits of RNAPII, TFIIB, both subunits of TFIIF (RAP30 and RAP74), Fcp1, the TFIIF-associated CTD phosphatase, and a novel human polypeptide that we named RPAP1.

To validate the existence of a unique 17-subunit complex containing RNAPII, we tagged other subunits of the complex,

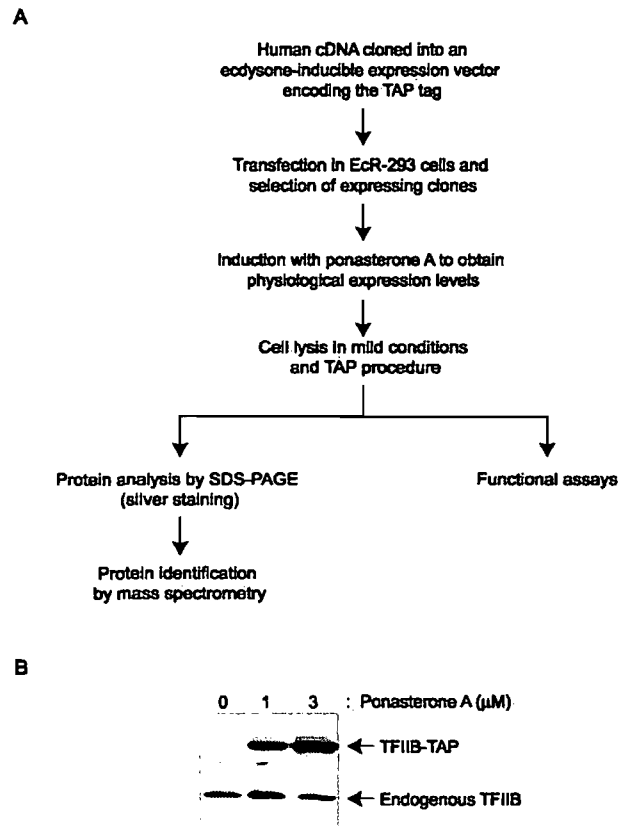
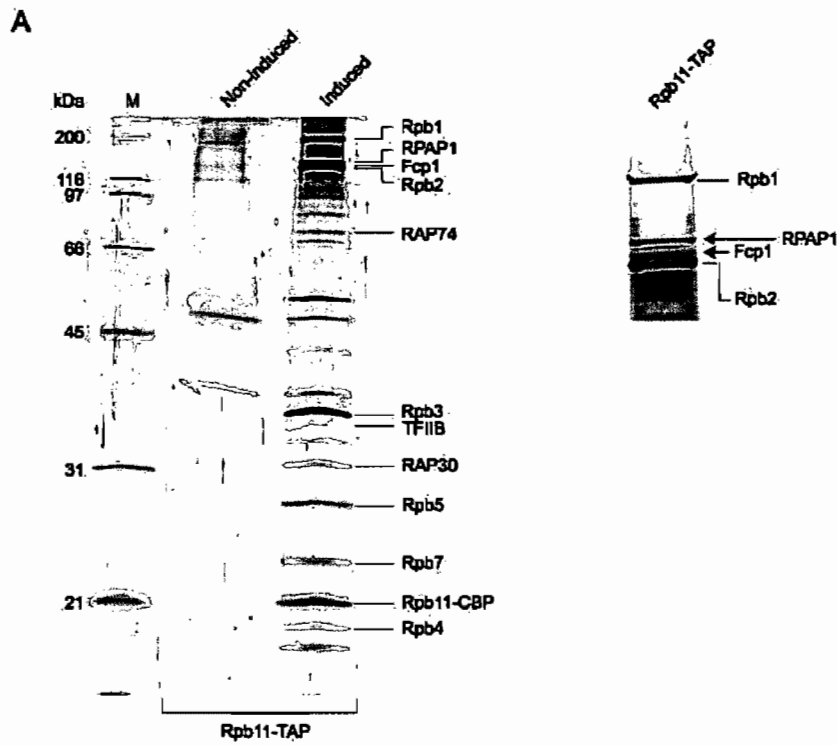


FIG. 1. TAP of human transcription factors. (A) Overview of the TAP procedure. (B) Data from a pilot experiment comparing the expression levels of a TAP-tagged polypeptide and its endogenous counterpart upon induction with ponasterone A. TFIIB-TAP expression in EcR-293 cells after induction with different concentrations of ponasterone A (0, 1, and 3 μ M) for 24 h was compared to that of endogenous TFIIB by Western blotting with an antibody raised against TFIIB.

including Rpb2, Rpb4, Rpb7, TFIIB, and RAP30, expressed them in stably transfected EcR-293 cells, and used them as baits in the TAP procedure. The eluates were analyzed by silver staining of SDS-PAGE gels, Western blotting, and/or mass spectrometry (Fig. 2B). The results indicated that all 17 polypeptides are bona fide components of the human RNAPII complex. Notably, the eluate prepared with TAP-tagged RAP30 contained both tagged RAP30 with its CBP residual domain and nontagged RAP30 (Fig. 2C). The presence of RAP30 lacking the CBP domain was confirmed by Western blotting (data not shown). These results indicate that two molecules of RAP30 are present within the complex, suggesting that TFIIF is associated with RNAPII as an $\alpha_2\beta_2$ heterotetramer (see Discussion). Finally, our Western blot analysis with an antibody specifically directed against the N-terminal part of Rpb1 (N-20) revealed that the tagged RNAPII complex contains the hypophosphorylated form of RNAPII (IIA form) (Fig. 2D).

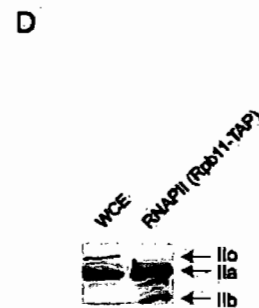
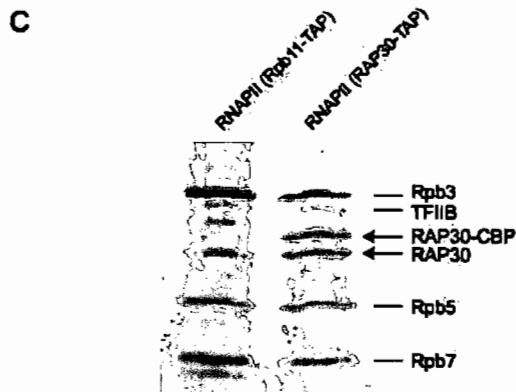
RPAP1 is structurally and functionally related to human RNAPII. RPAP1 is a previously identified human gene product (DKFZP727M111 protein) of unknown function. BLAST se-



B

Cell lines expressing TAP-tagged polypeptides

Polypeptides	Rpb11	Rpb2	Rpb4	Rpb7	TFIIB	RAP30
Rpb1	SS/MS/WB	SS/WB	SS	SS	SS/MS	SS
RPAP1	SS/MS	SS	SS	SS	SS	SS
Fcp1	SS/MS	SS	SS	SS	SS	SS
Rpb2	SS/MS	SS	SS/MS	SS	SS	SS
RAP74	SS/MS/WB	SS	SS/WB	SS/WB	WB	SS
Rpb3	SS/MS	SS/MS	SS	SS	SS	SS
TFIIB	SS/WB	SS	SS/WB	SS/WB	SS/MS/WB	SS
RAP30	SS/MS/WB	SS/WB	SS/WB	SS/WB	SS/WB	SS/WB
Rpb5	SS/MS	SS	SS	SS	SS	SS
Rpb7	SS	SS	SS/MS	SS/MS	SS	SS
Rpb4	SS	SS	SS/MS	SS/MS	SS	SS
Rpb6,8-12	SS	SS	SS	SS	SS	SS



quence searches against publicly available databases revealed highly significant alignments over the full length of the RPAP1 sequence with proteins of unknown function from *Mus musculus* (accession no. BAC65787), *Rattus norvegicus* (accession no. XP_230480), and to a lesser extent, *Drosophila melanogaster* (accession no. NP_648573 and CG32104-PB). In addition, these proteins show significant similarities to the *Saccharomyces cerevisiae* Ydr527wp protein (accession no. NP_010816), which is required for cell viability and was shown to bind the Rpb10 subunit of RNAPII in a two-hybrid screen (28, 43). The positions of the different alignments are shown in Fig. 3A. Notably, no particular motif or domain could be predicted with high confidence on the RPAP1 sequence. However, in view of the participation of RPAP1 in the RNAPII complex, an IMPALA analysis against the BLOCK database (74) revealed a very weak similarity ($E = 0.23$) between the bromodomain-containing protein 4 (Brd4) and the RPAP1 region spanning amino acids 196 to 312 (see below). We also performed a sequence-structure comparison analysis by using the structure-guided alignment program FUGUE against the HOMSTRAD database (56). The result suggests that RPAP1 belongs to the ARM repeat protein superfamily (12). Indeed, the RPAP1 region spanning amino acids 508 to 960 shows an alignment with ARM with a Z score of 4.08 (corresponding to 95% confidence). Interestingly, both Ydr527wp and the RPAP1 *D. melanogaster* homologue were also predicted, as reported in the *Saccharomyces* Genome and FlyBase databases, respectively, to belong to the ARM repeat superfamily (<http://genome-www4.stanford.edu/cgi-bin/SGD/locus.pl?locus=YDR527W> and <http://flybase.bio.indiana.edu/bin/fbidq.html?FBgn0052104>).

We created a Tet-promoter allele of the yeast *YDR527w* gene, turned off gene expression by adding doxycycline to the growth medium for 24 h, and compared gene expression in this mutant to expression patterns from 121 other Tet-promoter mutants affecting genes implicated in diverse biological processes (S. Mnaimneh, A. P. Davierwala, J. Haynes, J. Moffat, W. T. Peng, W. Zhang, X. Yang, J. Pootoolal, G. Chua, A. Lopez, M. Trochesset, D. Morse, N. J. Krogan, S. L. Hiley, Z. Li, Q. Morris, J. Grigull, N. Mitsakakis, C. J. Roberts, J. F. Greenblatt, C. Boone, C. A. Kaiser, B. J. Andrews, and T. R. Hughes, submitted for publication). Among all of these, the effect of the *tet-YDR527w* allele correlated most strongly with that of *tet-RPB11* (Fig. 3B). A clustering analysis did not reveal any specific group of functionally related transcripts that were up- or down-regulated uniquely in these two mutants, supporting the idea that the expression pattern observed is a conse-

quence of an overall perturbation in transcription rather than a disruption of any individual pathway.

To further define the interaction between RPAP1 and RNAPII, we cloned the human RPAP1 cDNA and derived a human cell line that can be induced to express TAP-tagged RPAP1. An analysis of the TAP-tagged RPAP1 eluate revealed the presence of two RNAPII subunits, Rpb2 and Rpb3 (Fig. 3A, right panel). These results further demonstrate the physical interaction between RPAP1 and RNAPII. Recent TAP-tagging experiments in yeast indicated that Ydr527wp associates with Rpb2, Rpb3, and Rpb11 (33). Because Ydr527wp and RPAP1 are both physically and functionally related, the 50-kDa yeast protein encoded by the *YDR527W* gene was named RBA50 (the name was registered in the *Saccharomyces* Genome Database).

The TAP-tagged RNAPII complex can be recruited to promoters both in vitro and in vivo and can bind to acetylated histone H3. In order to determine whether the TAP-tagged RNAPII complex is functional, we first verified its ability to enter a preinitiation complex in vitro. The electrophoretic mobility shift assay (EMSA) results in Fig. 4A show that the human RNAPII complex can form a preinitiation complex in the presence of TBP, TFIIB, TFIIF, and TFIIE on a radiolabeled probe carrying the AdML promoter. Compared to an equivalent preinitiation complex assembled with highly purified calf thymus RNAPII, the TAP-tagged RNAPII complex had a slightly lower mobility. Because our calf thymus RNAPII contains neither Fcp1 nor RPAP1 (data not shown), our results indicate that the preinitiation complex assembled with the TAP-tagged human RNAPII complex includes at least one of these two polypeptides (Fcp1 and/or RPAP1).

ChIP experiments were used to determine whether the tagged RNAPII complex can also be recruited to promoters in vivo. As shown in Fig. 4B, DNA fragments containing the *CCNA1* (cyclin A) promoter (positions -50 to $+50$) were enriched more efficiently than DNA fragments containing an upstream region (positions -1000 to -900) of the same gene when IgG-Sepharose beads were used to bind the protein A component of the TAP tag. Our results demonstrate that TAP-tagged RNAPII is specifically recruited to the promoter of actively transcribed genes in vivo.

As mentioned above, BLAST searches against publicly available databases revealed weak similarities between the amino acid sequence of RPAP1 and a variety of proteins and domains, including Brd4. Because Brd4 was recently shown to specifically interact with histones H3 and H4 through acety-

FIG. 2. Purification of TAP-tagged human RNAPII complex. (A) Whole-cell extracts prepared from induced or noninduced EcR-293 cells programmed to express TAP-tagged Rpb11 (Rpb11-TAP) were purified by the TAP procedure, and the eluates were analyzed by SDS-PAGE. Gel slices containing the most abundant polypeptides were excised, digested with trypsin, and analyzed by peptide mass fingerprinting using MALDI-TOF analysis. The positions of core RNAPII subunits, including Rpb11 carrying the residual CBP domain (Rpb11-CBP), TFIIB, RAP74, RAP30, Fcp1, and RPAP1, a human gene product (DKFZP727M111 protein) of unknown function, are indicated. (B) TAP tagging of Rpb2, Rpb4, Rpb7, Rpb11, TFIIB, and RAP30 allowed for purification of the same 17-polypeptide RNAPII complex. The components of the complex were identified according to their molecular weights in SDS gels stained with silver (SS), their peptide mass fingerprints by MALDI-TOF analysis (MS), and their immunoreactivities in Western blots (WB). (C) SDS gel showing that both tagged (RAP30-CBP) and nontagged RAP30 are present in the RNAPII complex purified with TAP-tagged RAP30. Only nontagged RAP30 is found in the complex purified with TAP-tagged Rpb11. (D) The TAP-tagged RNAPII complex contains a hypophosphorylated CTD. The N-20 antibody, which specifically recognizes the N terminus of Rpb1, was used to detect both the hypophosphorylated (IIa) and hyperphosphorylated (IIo) forms of RNAPII in the whole-cell extract (WCE) and the Rpb11-TAP eluate. A truncated form of Rpb1 lacking the CTD (IIb) was also detected.

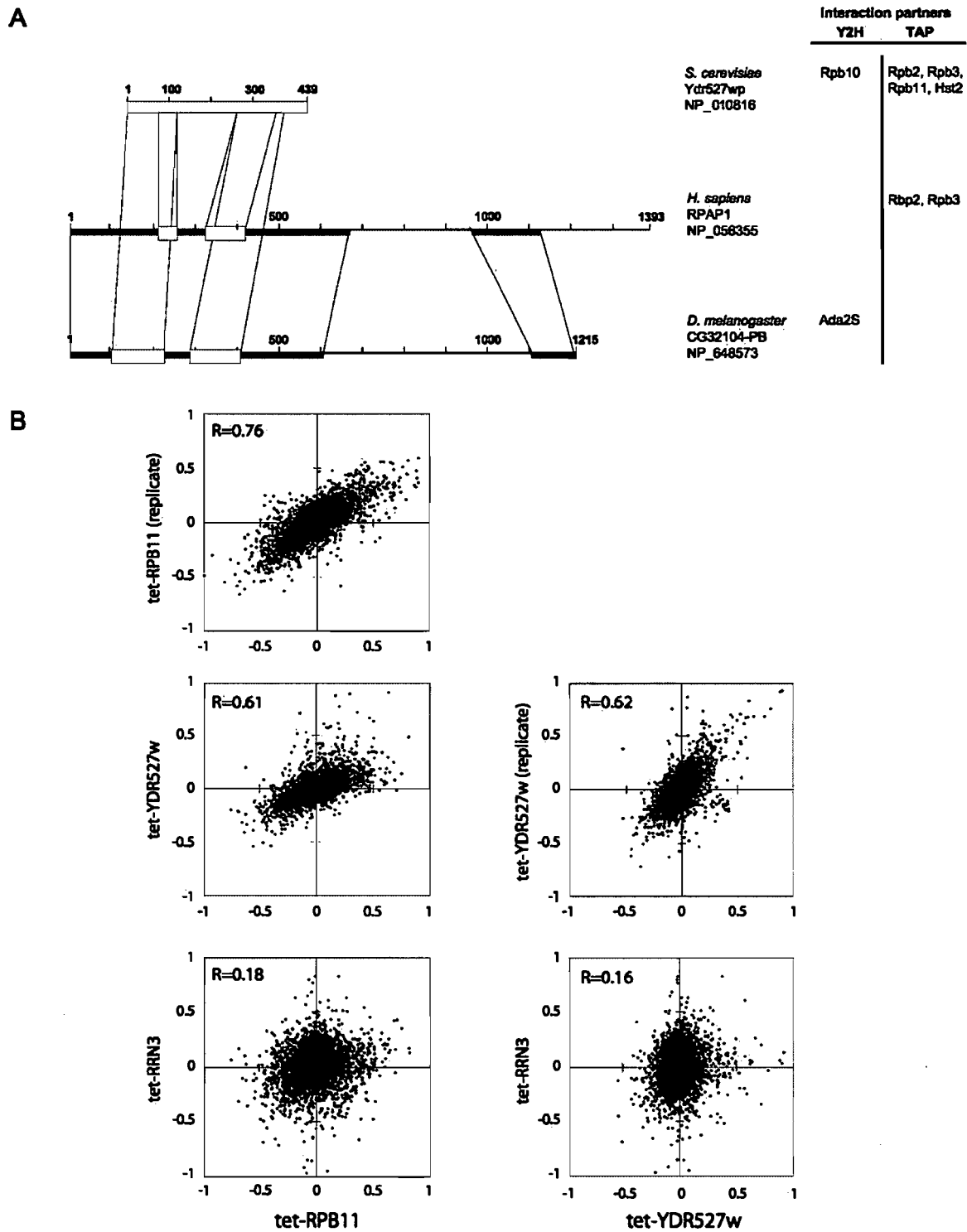


FIG. 3. (A) Regions of homology between *Homo sapiens* RPAP1, *D. melanogaster* CG32104-PB, and *S. cerevisiae* Ydr527wp/RBA50, as determined by BLAST analyses. Boxes represent regions with significant alignments. Black boxes, E values ranging from 8×10^{-37} to 1×10^{-10} ; open boxes, E values ranging from 4×10^{-5} to 0.1. The *M. musculus* and *R. norvegicus* orthologues, with 80% identity with the *H. sapiens* RPAP1 over the full length, were not presented. Reported protein-protein interactions of RPAP1 (this article) and Ydr527wp/RBA50 (33) with transcription factors are indicated on the right. (B) Correlations among mRNA abundance profiles for Tet-promoter mutants of *RPB11*, *YDR527w*, and *RRN3*, measured by using oligonucleotide microarrays. Each mutant was compared to an isogenic control without the *tetO7* promoter; log(ratios) are plotted. Tet-*RPB11* and Tet-*YDR527w* were assayed twice, with the replicate cultures being grown, extracted, and assayed on different dates. The correlation of Tet-*RPB11* and Tet-*YDR527w* to a Tet-promoter mutant of *RRN3*, a gene involved in polymerase I transcription, is shown for contrast.

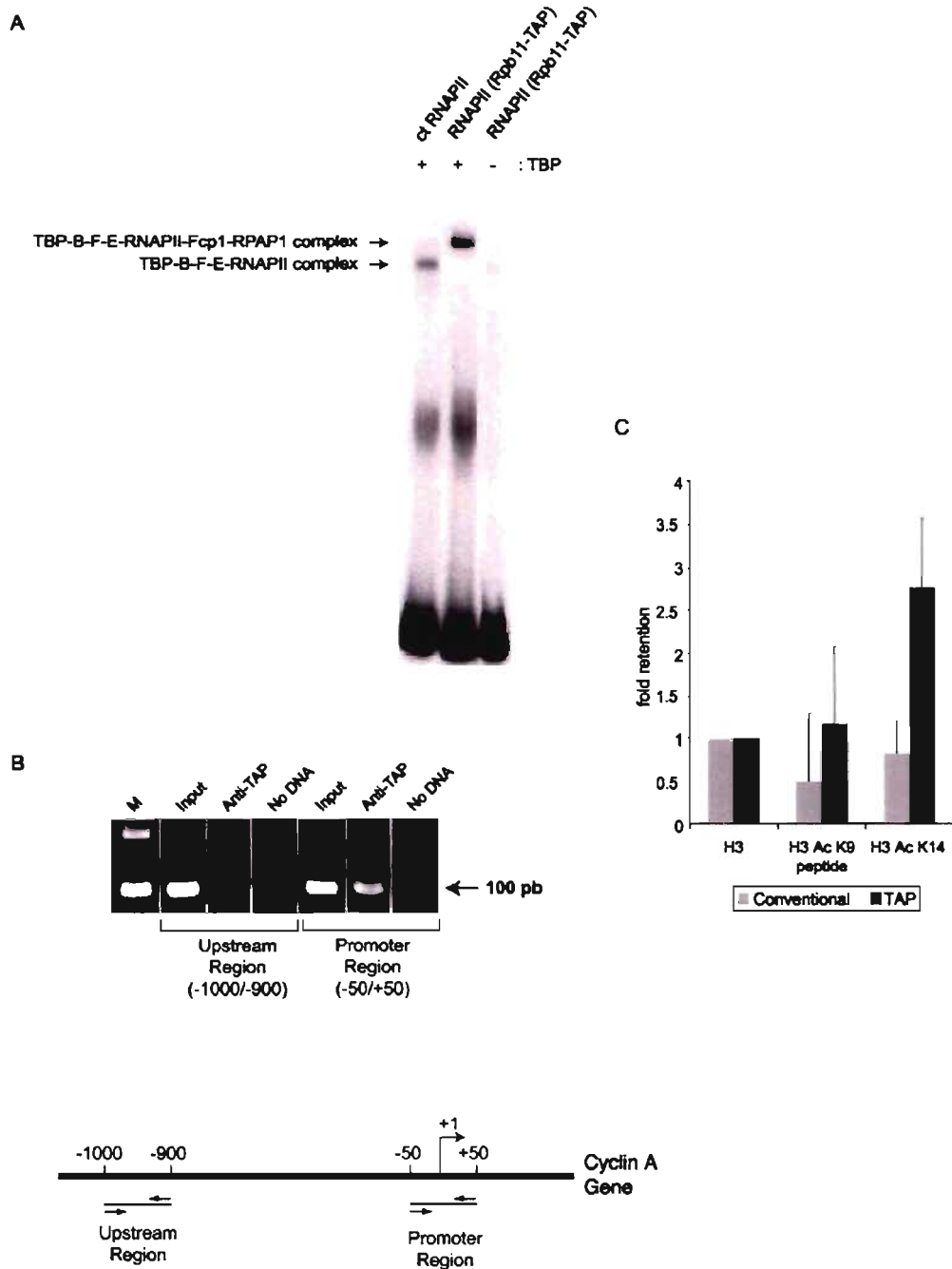


FIG. 4. TAP-tagged human RNAPII assembles on promoter DNA both in vitro and in vivo and binds to acetylated histones. (A) EMSA performed with 150 ng of purified calf thymus (ct) RNAPII or TAP-tagged human RNAPII in the presence of TBP, TFIIB, TFIIF, and TFIIIE. A control reaction assembled in the absence of TBP was included. A radiolabeled DNA fragment containing the AdML promoter was used as a probe. (B) ChIP experiments showing that TAP-tagged RNAPII is specifically recruited to the cyclin A promoter in vivo. PCR amplification using sets of primers specific to chromosomal regions either encompassing (-50 to +50) or located upstream of (-1000 to -900) the transcription start site was used on DNA fragments enriched with IgG beads (anti-TAP). The input lanes correspond to DNA that was not subjected to immunoprecipitation. A PCR control that lacks DNA (no DNA) was included in each case. (C) Peptide binding experiment showing that TAP-tagged RNAPII is retained on acetylated histones. The results of three independent experiments were analyzed with ImageQuant software, corrected for the background, and shown relative to unmodified H3. Conventionally purified RNAPII was used as a negative control. The TAP-tagged RNAPII-bound histone tails acetylated at lysine 14 more efficiently than at lysine 9.

lated lysine residues (19), we tested the ability of the complex to bind to acetylated histones. As shown in Fig. 4C, the TAP-tagged RNAPII complex specifically bound to acetylated histone H3 *in vitro*, while conventionally purified RNAPII was not retained on acetylated histone H3. TAP-tagged RNAPII bound histone tails acetylated at lysine 14 more efficiently than those acetylated at lysine 9. This result suggests that RPAP1 may recruit RNAPII to chromatin through its interaction with acetylated histones.

The TAP-tagged RNAPII complex supports accurately initiated transcription *in vitro*. TAP-tagged RNAPII was tested for the ability to initiate transcription *in vitro* by use of a transcription system reconstituted with purified general initiation factors. Purified calf thymus RNAPII or purified TAP-tagged human RNAPII was incubated with a DNA template containing the AdML promoter in the presence of purified TBP, TFIIB, TFIIF, TFIIE, and TFIIH. The formation of a 391-nt transcript indicated that the tagged RNAPII complex is competent for promoter-selective transcription *in vitro* (Fig. 5A).

When the general initiation factors were individually omitted from the transcription reaction (Fig. 5B), transcription was abolished except in the case of recombinant TFIIB or TFIIF. These data fully support our conclusion that TFIIB and TFIIF are components of the tagged RNAPII complex and establish that they are functionally active within this complex.

Purification and functional analysis of a human RNAPII variant with a site-directed mutation in fork loop 1. As part of an effort to understand the fine molecular details of the transcription reaction, we used the TAP procedure to generate and isolate RNAPII variants with site-directed mutations in structural elements predicted to have roles in the transcription reaction according to the crystal structure of the yeast enzyme. Because the Rpb2 element known as fork loop 1 (Fig. 6A) is highly conserved from yeast to humans, absent from bacterial RNAP, and proposed to be involved in maintaining the transcription bubble (17, 30) or in RNA-DNA strand separation (85), we first chose to modify this element. In yeast RNAPII, fork loop 1 amino acids Lys471 and Arg476 appear to contact the RNA phosphates at positions -5 , -6 , and -7 in the hybrid region. Deletion of the whole 13-amino-acid loop totally abolished the ability of Rpb2 to assemble with the other subunits of RNAPII (data not shown), probably because the deletion destabilized Rpb2, which then became misfolded. A smaller deletion of two amino acids located near the center of the loop ($\Delta 458-459$), which corresponded to amino acids Lys471 and Ala472 in the yeast enzyme, was successfully used to purify the 17-subunit RNAPII complex (Fig. 6B). This mutant was named Rpb2 fork1 $\Delta 458-459$. The silver-stained SDS gel in Fig. 6B shows that the relative abundance of some components of the tagged RNAPII complex varied between the wild type and the mutant, indicating that the mutation may have slightly affected complex assembly. By quantifying the relative intensity of each subunit, we established that the most important difference concerned Rpb1, which was at the very most fivefold less abundant in the mutant than in the wild-type complex.

In order to elucidate the function of fork loop 1 in transcription, we compared Rpb2 fork1 $\Delta 458-459$ with wild-type RNAPII in several different assays. The assays shown in Fig. 6C to E used fivefold larger amounts of the mutant complex than of

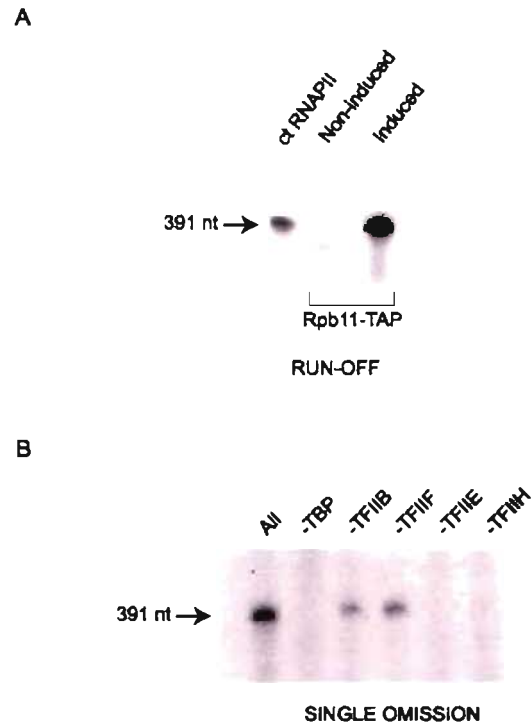


FIG. 5. TAP-tagged human RNAPII complex can initiate transcription *in vitro*. (A) *In vitro* transcription reactions were reconstituted by using 165 ng of purified calf thymus (ct) RNAPII or TAP-tagged human RNAPII in the presence of TBP, TFIIB, TFIIF, TFIIE, and TFIIH. A control reaction performed with an eluate from noninduced cells was included. The linearized DNA template carries the AdML promoter and directs the synthesis of a 391-nt transcript. (B) Single omission assay in which each general transcription factor was omitted from the reconstituted system (all) described for panel A.

the wild type to ensure that the amount of the mutant complex was not underestimated (see above). As shown in Fig. 6C, Rpb2 fork1 $\Delta 458-459$ was defective in runoff transcription assays when we used the AdML promoter in the presence of all of the general initiation factors.

In order to delineate with more precision the molecular defect that impairs the transcriptional activity of Rpb2 fork1 $\Delta 458-459$ *in vitro*, we tested its activity in a promoter-independent assay. As described previously (29, 79), RNAPII can transcribe from a template carrying a 3' C extension (C-tailed template) in the absence of the general initiation factors. The template used for Fig. 6D carried a 15-C extension on the 3' strand. The transcribed strand was devoid of dCMP except at positions $+25$ to $+27$ ($+1$ corresponds to the single-stranded-double-stranded junction). In the absence of CTP, wild-type RNAPII produced a transcript of 35 nt. In the presence of CTP, the major transcript was 60 nt in length, indicating that the polymerase starts transcription 10 nt upstream of the double-stranded junction. In contrast to the RNAPII complex carrying wild-type Rpb2, Rpb2 fork1 $\Delta 458-459$ was inactive in the C-tailed template elongation assay.

Figure 6E shows that Rpb2 fork1 $\Delta 458-459$ was also defective in an abortive initiation assay carried out with a short double-stranded oligonucleotide carrying the AdML promoter

in the presence of the general initiation factors, but in the absence of GTP. A G nucleotide strategically located at +11 in the template blocked transcription at position +10 (Fig. 6E). When we used a fully double-stranded template (0/0), Rpb2 fork1 Δ 458-459 did not produce abortive transcripts. We rationalized that if Rpb2 fork1 Δ 458-459 was defective in promoter melting, a premelted template would counteract the defect in the mutant. The premelted template (-9/+2), which carried an unpaired DNA region from -9 to +2 that mimicked the fully melted promoter, supported the synthesis of abortive transcripts, although with a poor efficiency compared to that of the wild type. This finding indicated that a mutation in fork loop 1 does not simply affect promoter melting and its maintenance, but rather affects the interaction of RNAPII with the melted promoter region.

We finally assessed the ability of Rpb2 fork1 Δ 458-459 to assemble into a preinitiation complex on promoter-containing DNA. EMSA experiments showed that Rpb2 fork1 Δ 458-459 can enter a preinitiation complex with the general initiation factors as efficiently as wild-type RNAPII (Fig. 6F).

Human RNAPII variant with a site-directed mutation in switch 3. The structure named switch 3 is the only loop of Rpb2 located near the DNA-RNA hybrid that is ordered in the crystal structure of elongating yeast RNAPII (30). Switch 3 makes direct contact with the template DNA near positions -5 (arginine 1122) and -2 (arginine 1129) (Fig. 6A). Since R1129 is juxtaposed with one of the Rpb2 domains that mediate the interaction with Rpb1, we reasoned that disruption of this amino acid would likely affect the assembly of RNAPII and decided to turn our attention to R1122. We produced a mutant in which amino acids R1122, S1123, and R1124 were replaced with alanine residues. In human Rpb2, these amino acids correspond to R1078, S1079, and R1080. This mutant was named Rpb2 sw3-1078.

Figure 7A shows that TAP-tagged Rpb2 with the sw3-1078 mutation is as efficient as the wild type at forming the 17-subunit RNAPII complex. However, EMSA experiments (Fig. 7B) indicated that purified Rpb2 sw3-1078 is not efficient at entering a preinitiation complex on promoter DNA in the presence of the general initiation factors. Increasing the concentration of Rpb2 sw3-1078 by a factor of 6 compared to the wild type was necessary in order to achieve efficient formation of a preinitiation complex. As expected on the basis of the EMSA results, Rpb2 sw3-1078 did not support transcription *in vitro* (Fig. 7C). Increasing the concentration of Rpb2 sw3-1078 by a factor of 6 compared to the wild type led to the production of runoff transcripts, but the transcription efficiency remained slightly lower than that of the wild type, indicating that the mutation in switch 3 not only affects the ability of Rpb2 sw3-1078 to assemble into a preinitiation complex, but also, to a lesser extent, affects its capacity to support transcription.

Mutants Rpb2 fork1 Δ 458-459 and Rpb2 sw3-1078 are defective *in vivo*. To assess the defects of Rpb2 fork1 Δ 458-459 and Rpb2 sw3-1078 *in vivo*, we used ChIP experiments to evaluate their association with both the promoter and the transcription unit of two genes that are actively transcribed in human cells, the guanine nucleotide binding protein beta polypeptide 2-like 1 (GNB2L1) gene and the ferritin light polypeptide (FTL) gene. As shown in Fig. 8, Rpb2 fork1 Δ 458-459 is recruited to both promoters with an efficiency similar to

that of a wild-type TAP-tagged RNAPII complex, but it is not found in association with the transcription units. In contrast, Rpb2 sw3-1078 is not enriched in the promoter regions or the transcribed regions. The mutant Rpb2 fork1 Δ 453-465, which carries a complete deletion of the 13-amino-acid loop and does not assemble into a bona fide 17-subunit RNAPII complex, was used as a negative control that was enriched at neither the promoter nor the transcription unit of either the GNB2L1 or FTL gene. These *in vivo* results fully support our *in vitro* studies indicating that fork loop 1 is essential for transcription but is not involved in preinitiation complex assembly, whereas switch 3 is required for efficient preinitiation complex formation.

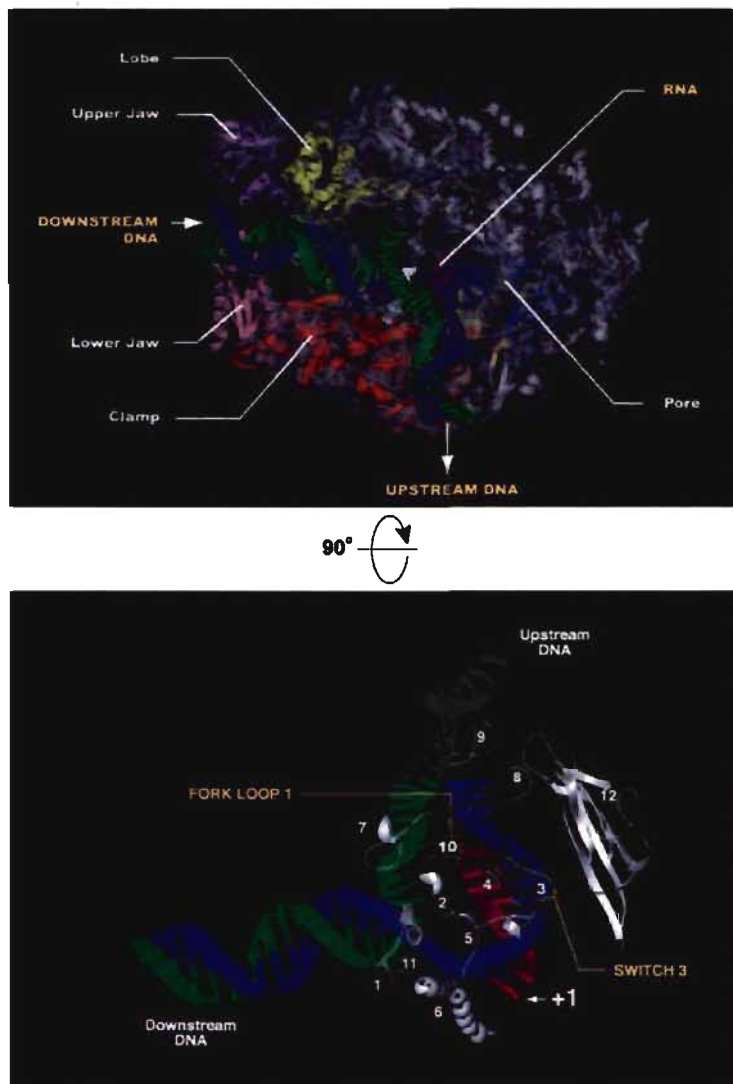
DISCUSSION

Simple, efficient purification of human RNAPII under native conditions. Classical approaches for the purification of mammalian RNAPII and its general transcription factors have typically used many different chromatography steps, including ion-exchange chromatography. Eluting these columns with buffers containing high salt concentrations most certainly affected the polypeptide composition of the purified complexes. For example, regulatory subunits, which are often less tightly associated with the enzyme they regulate, may have been lost during the purification procedure. Seraphin and colleagues (71) developed the TAP procedure to allow for the purification of protein complexes under native conditions. Two affinity chromatography steps are used. Elution from an IgG column is performed by a proteolytic treatment with the tobacco etch virus protease while elution from a calmodulin column uses EGTA as a chelating agent. For this paper, we adapted the TAP procedure for the purification of protein complexes from human cells, using an ecdysone-inducible expression system that allows for the expression of the TAP-tagged polypeptides at near-physiological levels. The expression levels were modulated by adjusting the concentration of the inducer, ponasterone A (Fig. 1B). By preventing overexpression of the tagged polypeptides, we limited the formation of nonspecific interactions.

We report here the use of the TAP procedure with human cells for the easy and efficient purification of a complex containing the 12 subunits of RNAPII, both subunits of TFIIF, TFIIB, Fcp1, and RPAP1, a novel polypeptide of unknown function. This 17-subunit complex was purified through the alternate TAP tagging of six of its subunits, demonstrating that all of the components are part of a unique complex. Because (i) the TAP-tagged polypeptides are expressed at near-physiological levels in stably transfected human cells, (ii) protein extraction is performed under gentle conditions, and (iii) the affinity purification steps are achieved under native conditions, our results indicate that this TAP-tagged RNAPII complex is the most abundant soluble, stable form of RNAPII in human cells. If a mediator complex (70) is associated with RNAPII in human cell extracts, either its concentration is much lower than that of RNAPII or its association with RNAPII is too weak to survive double affinity purification.

The presence of the TAP tag at the C termini of subunits of the RNAPII-containing complex does not alter its activity. The TAP-tagged RNAPII complex is fully active *in vitro*, as it can

A



B

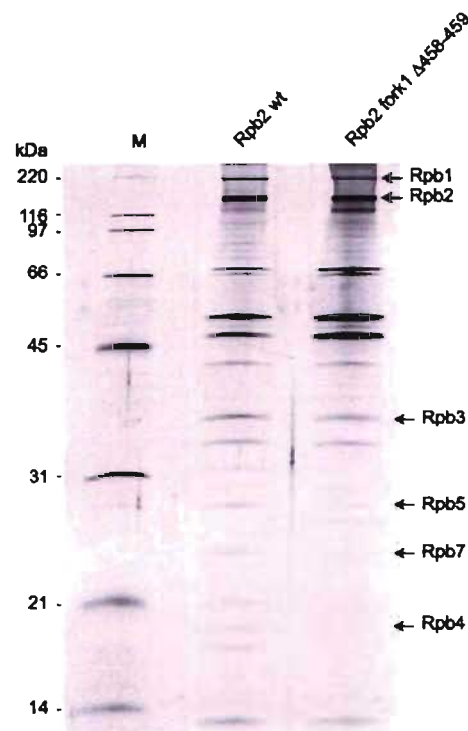


FIG. 6. Purification and functional analysis of TAP-tagged human RNAPII with a mutation in fork loop 1. (A) (Top) Model showing elongating yeast RNAPII (30). The template DNA strand (blue) from -3 to $+10$ and the 9-nt RNA (pink) were placed according to the crystal structure. The remainder of the template strand and the coding strand (green) were modeled in Cinema 4D, and their exact positions are speculative. (Bottom) Simplified view of yeast RNAPII catalytic center. Different domains of Rpb1 and Rpb2 located near the DNA-RNA hybrid are shown. Numbers: 1 to 5, switches; 6, bridge helix; 7, rudder; 8, lid; 9, zipper; 10, fork loop 1; 11, fork loop 2; 12, wall. The parts of the lid (amino acids 250 to 258) and the rudder (amino acids 315 to 320) that are absent in the PDB files corresponding to elongating RNAPII were reconstructed based on the crystal structure of free yeast RNAPII (18). The missing parts of fork loop 1 (amino acids 468 to 476) and fork loop 2 (amino acids 503 to 508) were modeled in Cinema 4D, and their exact positions are speculative. (B) Silver-stained SDS gel of wild type TAP-tagged RNAPII (Rpb2 wt) and a mutant of human RNAPII carrying a two-amino-acid deletion in the Rpb2 fork loop 1 (Rpb2 fork1 $\Delta 458-459$) domain. The positions of some RNAPII subunits and molecular size markers are indicated. (C) In vitro transcription reactions (runoff) contained either 24 ng of wild-type RNAPII (Rpb2 wt) or 120 ng of Rpb2 fork1 $\Delta 458-459$ in the presence of TBP, TFIIB, TFIIF, TFIIIE, and TFIIH. (D) Elongation assays were performed with a C-tailed template carrying a 15-C extension in the absence of general transcription factors. The coding strand lacks CMP except in the $+25$ -to- $+27$ region. The positions of transcripts produced by RNAPII in the presence (+CTP; 60 nt) or absence (-CTP; 35 nt) of CTP are indicated. The reactions contained 88 ng of Rpb2 wt and 440 ng of Rpb2 fork1 $\Delta 458-459$. (E) Abortive initiation assays were performed with 100 ng of Rpb2 wt and 500 ng of Rpb2 fork1 $\Delta 458-459$ in the presence of TBP, TFIIB, TFIIF, and TFIIIE on closed (0/0) or premelted ($-9/+2$) templates in the absence of GTP. The templates are schematized and the 3- to 10-nt abortive transcripts are indicated. (F) EMSAs were performed with the AdML promoter in the presence of TBP, TFIIB, TFIIF, and TFIIIE. Two hundred nanograms of Rpb2 wt and 500 ng of Rpb2 fork1 $\Delta 58-459$ were used in the reactions.

form a preinitiation complex with the general initiation factors on the AdML promoter, accurately initiate transcription, and elongate the transcripts (Fig. 4 and 5). In addition, both the TFIIB and TFIIF components of the complex are functionally

active in transcription reactions. Notably, TAP-tagged RNAPII can be recruited to transcriptionally active promoters in vivo, an indication that the purified human RNAPII complex is biologically active.

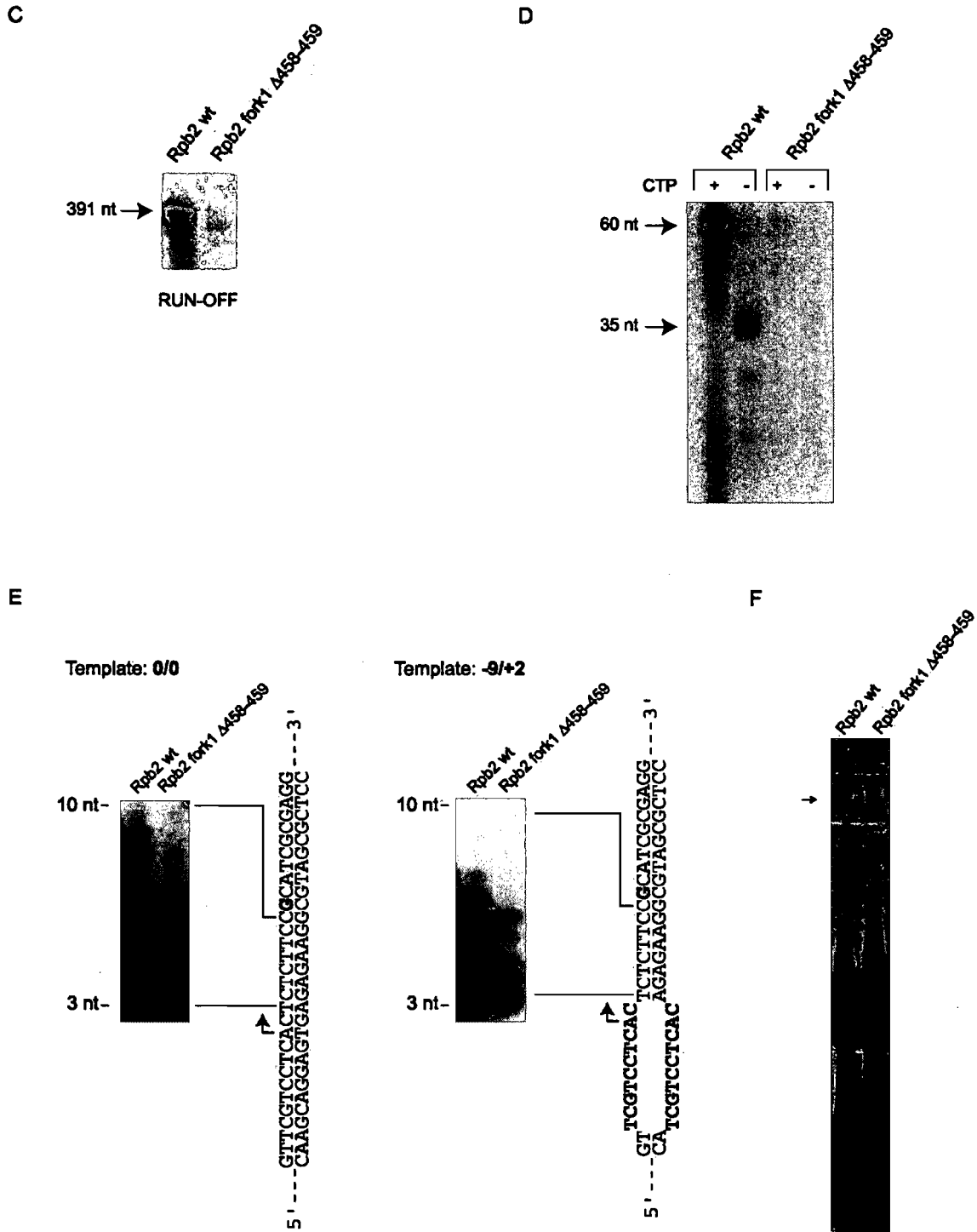


FIG. 6—Continued.

Five different regulatory subunits are stably associated with soluble RNAPII in human cells. The soluble form of RNAPII is expected to be the form of the enzyme that is recruited to promoters upon activation signals. In support of this hypothesis, the CTD of Rpb1 in the TAP-tagged RNAPII complex is in the hypophosphorylated state, the form of the enzyme that

is recruited to promoter DNA. A possible function of Fcp1 in this complex could be to maintain the CTD in this unphosphorylated state, as previously proposed (47). Our EMSAs suggested that Fcp1 can enter the preinitiation complex with the RNAPII complex. The entry of Fcp1 with RNAPII at the promoter is supported by the finding by Buratowski and col-

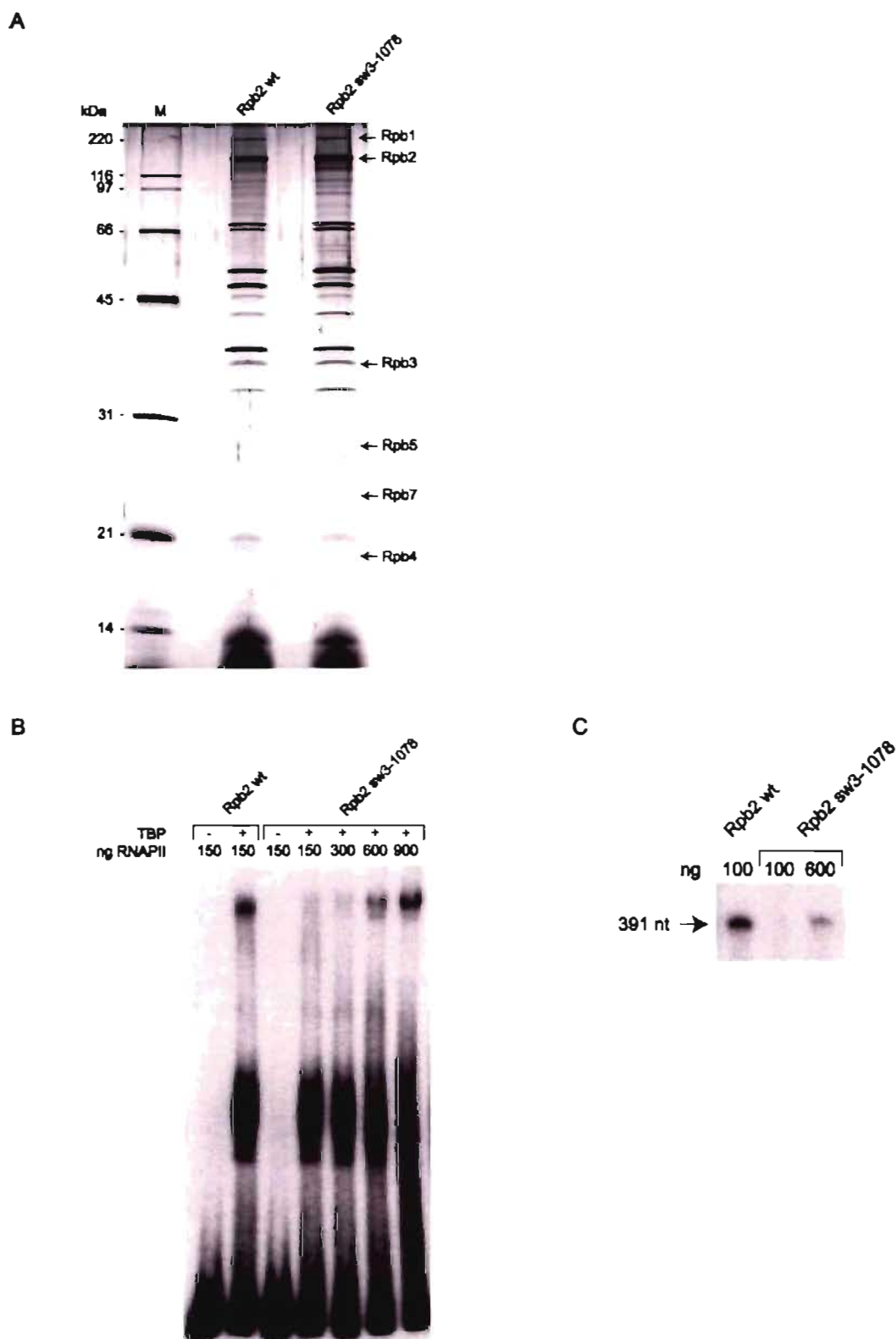


FIG. 7. Purification and functional analysis of TAP-tagged human RNAPII with a mutation in switch 3. (A) Silver-stained SDS gel of wild-type TAP-tagged RNAPII (Rpb2 wt) and a mutant of human RNAPII carrying a triple alanine substitution in the Rpb2 switch 3 (Rpb2 sw3-1078) domain. (B) EMSAs were performed with the AdML promoter with TFIIB, TFIIF, and TFIIE in either the presence or the absence of TBP. The amounts used for the wild type and the mutant are indicated. (C) In vitro transcription reactions (runoff) contained different amounts of Rpb2 wt and Rpb2 sw3-1078 in the presence of TBP, TFIIB, TFIIF, TFIIE, and TFIIH.

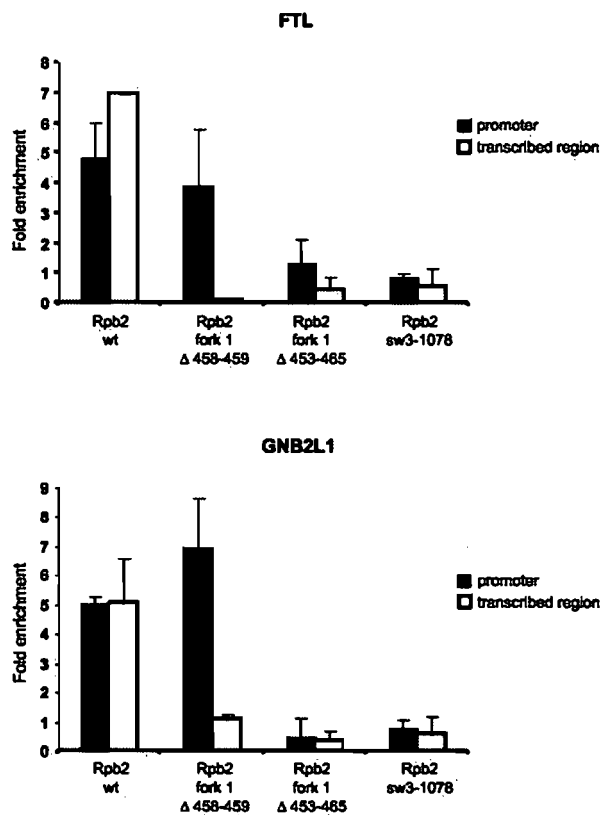


FIG. 8. Association of mutant forms of the RNAPII complex with chromatin in vivo. The fold enrichment over the control region of both the promoter and the transcribed region of the FTL and GNB2L1 genes after the immunoprecipitation of chromatin fragments was quantified by Q-PCR. Enrichment was normalized to an intergenic region of chromosome 17 (see Materials and Methods). Data points represent the means of four experiments, including two independent ChIP assays.

leagues that the CTD phosphatase colocalizes with RNAPII at active promoters in yeast (10). Therefore, the role of Fcp1 in maintaining RNAPII in the hypophosphorylated state may be required until RNAPII becomes processive and clears the promoter.

Both TFIIB and TFIIF have previously been shown to regulate the recruitment of RNAPII to promoter DNA (15, 54, 80). TFIIB binds to TBP and forms a TBP-TFIIB-promoter complex (4, 59). TFIIB binds directly to a DNA motif named the TFIIB recognition element, a core element found in several RNAPII promoters (51). TFIIF binds to both RNAPII and TFIIB and is also involved in RNAPII recruitment to core promoters (15, 54, 80). In addition, TFIIF was shown to regulate the contacts of RNAPII with promoter DNA before transcription initiation (20, 25, 72). TFIIF plays a central role in promoter bending and wrapping against the mobile clamp of RNAPII, a process that is essential for transcription initiation (24). The function of TFIIF in promoter wrapping and transcriptional initiation requires a homomeric interaction region of RAP74, called HIR-1, that maintains TFIIF as a $\alpha_2\beta_2$ heterotetramer (72). Our results showing that two molecules of RAP30 are present within the tagged RNAPII complex sup-

port the notion that TFIIF works as a $\alpha_2\beta_2$ heterotetramer in vivo (13, 23).

Notably, both TFIIF and TFIIB can also bind to Fcp1 in a competitive manner and with opposite effects, with TFIIF stimulating and TFIIB inhibiting the phosphatase activity (9). Fcp1 may not efficiently dephosphorylate RNAPII at Ser2 of the CTD in the region downstream of the promoter (10) because TFIIF appears not to be present in this complex (50, 68). Clearly, TFIIB, TFIIF, and Fcp1 participate in an important network of regulatory interactions during RNAPII transcription.

The TAP-tagged RNAPII complex also contains at least one polypeptide that can specifically bind to acetylated histones in vitro. One putative candidate is RPAP1. Whether or not RPAP1 contains a bromodomain remains unclear because the definitive identification of a bromodomain requires structural data. An essential yeast polypeptide with regions of homology with RPAP1 was previously shown to interact with the Rpb10 subunit of RNAPII (28, 43) and also affected global gene expression in the same way as Rpb11 in our Tet-promoter mutant experiments (Fig. 3). These findings support the notion that the association of RPAP1 with RNAPII is physiologically relevant, although its exact role remains to be elucidated.

Structure-guided alignments indicated that RPAP1 could belong to the ARM repeat protein superfamily (12). ARM repeat proteins contain tandem copies of a degenerate 40-amino-acid sequence motif that forms a conserved three-dimensional structure. A single ARM repeat consists of three α helices. Such repeats are found in a variety of proteins, including the mammalian armadillo homologs β -catenin and importin α (40). β -catenin is mainly detected in adherens junctions where it links the membrane-anchored cadherin to actin filament. Cytoplasmic β -catenin is degraded by the ubiquitin-proteasome system. However, when Wnt signaling is activated, cytoplasmic β -catenin is stabilized. The increased level of β -catenin leads to its nuclear translocation, where it affects transcriptional activation. Although it was reported to be a cytoplasmic protein (41), Ydr527wp/RBA50 is found in complex with three RNAPII subunits and appears to affect transcription. Similar to β -catenin, Ydr527wp/RBA50 may have a dual cellular location and may be translocated to the nucleus.

Proteins that were previously shown to interact with RNAPII, such as the mediator and various elongation factors, were not found to be major components of our TAP-tagged RNAPII eluate. This is likely because these RNAPII-interacting proteins are associated with RNAPII in the insoluble fraction and, consequently, are mainly lost during the purification procedure. However, these complexes may represent minor forms of RNAPII complexes in our TAP-tagged eluates.

Structure-function analysis of the human RNAPII active center. The crystal structure of eukaryotic RNAPII was used to predict the function of many structural elements found in the active center. For example, some loops and helices that either directly contact or closely approach the DNA-RNA duplex in the active site of the elongating enzyme were suggested to regulate the formation and maintenance of both the transcription bubble and the RNA-DNA hybrid. In order to begin assessing the functions of these structural elements, we expressed TAP-tagged Rpb2 mutants carrying site-directed mutations and purified the mutant forms of RNAPII. Our data

indicate that the assembly and/or stability of the RNAPII complex is affected by mutations in the active site. For example, the full deletion of fork loop 1 led to defects in the assembly of the 17-subunit RNAPII complex. However, a shorter mutant with a deletion of only two amino acids from Rpb2 was found to assemble correctly in solution *in vivo* (Fig. 6) and allowed us to analyze the function of this element during the transcription reaction. The Rpb2 fork1 $\Delta 458-459$ mutant was found to enter a preinitiation complex with the general initiation factors with an efficiency similar to that of the wild-type complex, but it was incapable of transcription initiation and elongation *in vitro*. When we used a premelted template in which the transcription bubble was fully open between positions -9 and $+2$ in our initiation assay, Rpb2 fork1 $\Delta 458-459$ was able to initiate transcript formation, but not very efficiently. These findings indicate that there may be a defect at the level of the interaction between fork loop 1 and the melted DNA at very early stages of the transcription reaction. For example, the two-amino-acid-shorter fork loop 1 in the RNAPII mutant may be unable to contact one of the DNA strands in the transcription bubble, leading to incorrect positioning of the DNA template in the active site and impairing the ability of the enzyme to initiate transcription efficiently. Since this mutant supported transcription initiation much less efficiently even when a fully open template was used, our results do not support a direct role for fork loop 1 in promoter melting.

Given the recent crystal structure of elongating yeast RNAPII showing an interaction between Lys 471 (Lys 458 in human RNAPII) and the RNA around positions -5 to -7 (85), an alternative possibility to explain the defect of Rpb2 fork 1 $\Delta 458-459$ is that during transcription initiation, the mobile fork loop 1 makes crucial contacts with the first arriving ribonucleotides, making it essential for phosphodiester bond formation. According to this speculative view, fork loop 1 would keep contact with the RNA as transcription progresses and be displaced until it reaches the position seen in the crystal structure, where it would participate in RNA-DNA strand separation.

The five switches located at the bottom of the cleft of RNAPII were first proposed to be involved in coupling the closure of the mobile clamp to the presence of a DNA-RNA hybrid (18, 30). This hypothesis was based on the fact that most of the switches are disordered in the crystal structure of the 10-subunit RNAPII complex in solution (e.g., lacking the Rpb4-Rpb7 dimer), whereas they are ordered in the elongating RNAPII structure. The ordering of the switches coincided with the movement of the clamp, which was in an open state in the structure of the free enzyme to permit the entry of DNA into the cleft and in a closed state in the structure of the RNAPII-DNA-RNA ternary complex to ensure the stability of the elongating complex. Since some switches, including switch 3, directly contact the DNA in the DNA-RNA hybrid region, they were proposed to be involved in sensing the presence of nascent RNA and coupling it to the closure of the clamp on the DNA. However, the recent crystal structures of the 12-subunit RNAPII complex (2, 7) revealed that the clamp is in a closed state in the presence of the Rpb4-Rpb7 dimer. These results led the authors to propose that the DNA is first loaded onto the clamp far from the active site in a way that is similar to that of the bacterial RNA polymerase holoenzyme-promoter DNA

complex (57) and that it reaches the cleft only after DNA melting has occurred. In this situation, switch 3 could only contact DNA after the opening of the promoter and entry of the DNA into the cleft.

Our results indicate that switch 3 is important for the assembly of RNAPII into a preinitiation complex on promoter DNA. Mutating amino acids 1078 to 1080 clearly reduced the ability of human RNAPII to form a preinitiation complex in the presence of the general transcription factors both *in vitro* and *in vivo*. *In vitro*, sixfold larger amounts of Rpb2 sw3-1078 were necessary to form a complex in gel shift experiments. Our results suggest that switch 3 is necessary for the accurate interaction of the clamp with the promoter DNA before opening of the transcription bubble and that this is the primary defect in our switch 3 mutant. Because the structure of the complete polymerase shows that a duplex DNA cannot bind switch 3 before melting, it is likely that the effect of the switch 3 mutation is either allosteric or mediated by interactions with a general initiation factor such as TFIIB or TFIIF. The importance of the interaction of switch 3 with the template DNA at positions -2 and -5 after promoter melting might be reflected by the fact that the transcription reactions carried out in the presence of sixfold larger amounts of the sw3-1078 mutant did not yield as many transcripts as did the wild type. This may also suggest that the preinitiation complexes formed with the sw3-1078 mutant are not fully functional, presumably because the interactions of switch 3 with the DNA template in the DNA-RNA hybrid region are impaired.

ACKNOWLEDGMENTS

We are grateful to the members of our laboratory for helpful discussions. We thank Diane Bourque for artwork and Julie Edwards for critical reading of the manuscript. We also thank Takahiro Nagase from the Kazusa DNA Research Institute for kindly providing the cDNA encoding RPAP1.

This work was supported by grants from the Canadian Institutes for Health Research (to B.C.), Genome Canada (to B.C., J.G., and T.R.H.), Genome Québec (to B.C.), the Ontario Genomics Institute (to J.G. and T.R.H.), and the Natural Sciences and Engineering Research Council of Canada (to T.R.H.). C.J. holds a studentship from the Canadian Institutes for Health Research and M.F.L. is supported by the Natural Sciences and Engineering Research Council of Canada and the Fonds Québécois de la Recherche sur la Nature et les Technologies. A.P.D. is supported by a C. H. Best Postdoctoral Fellowship. B.C. is a senior scholar from the Fonds de la Recherche en Santé du Québec.

REFERENCES

1. Acker, J., M. Wintzerith, M. Vigneron, and C. Keding. 1992. Primary structure of the second largest subunit of human RNA polymerase II (or B). *J. Mol. Biol.* **226**:1295-1299.
2. Armache, K. J., H. Kettenberger, and P. Cramer. 2003. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **100**:6964-6968.
3. Awrey, D. E., R. G. Weibaecker, S. A. Hemming, S. M. Orlicky, C. M. Kane, and A. M. Edwards. 1997. Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J. Biol. Chem.* **272**:14747-14754.
4. Bagby, S., S. Kim, E. Maldonado, K. I. Tong, D. Reinberg, and M. Ikura. 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.
5. Bryant, G. O., and M. Ptashne. 2003. Independent recruitment *in vivo* by Gal4 of two complexes required for transcription. *Mol. Cell* **11**:1301-1309.
6. Burton, Z. F., L. G. Ortolan, and J. Greenblatt. 1986. Proteins that bind to RNA polymerase II are required for accurate initiation of transcription at the adenovirus 2 major late promoter. *EMBO J.* **5**:2923-2930.
7. Bushnell, D. A., and R. D. Kornberg. 2003. Complete, 12-subunit RNA

- polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc. Natl. Acad. Sci. USA* **100**:6969–6973.
8. Campbell, E. A., O. Muzzin, M. Chlenov, J. L. Sun, C. A. Olson, O. Weinman, M. L. Trester-Zedlitz, and S. A. Darst. 2002. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell* **9**:527–539.
 9. Chambers, R. S., B. Q. Wang, Z. F. Burton, and M. E. Dahmus. 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.
 10. Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt, and S. Buratowski. 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev.* **15**:3319–3329.
 11. Choder, M., and R. A. Young. 1993. A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. *Mol. Cell. Biol.* **13**:6984–6991.
 12. Coates, J. C. 2003. Armadillo repeat proteins: beyond the animal kingdom. *Trends Cell Biol.* **13**:463–471.
 13. Conaway, J. W., and R. C. Conaway. 1989. A multisubunit transcription factor essential for accurate initiation by RNA polymerase II. *J. Biol. Chem.* **264**:2357–2362.
 14. Conaway, J. W., A. Shilatfard, A. Dvir, and R. C. Conaway. 2000. Control of elongation by RNA polymerase II. *Trends Biochem. Sci.* **25**:375–380.
 15. Conaway, R. C., K. P. Garrett, J. P. Hanley, and J. W. Conaway. 1991. Mechanism of promoter selection by RNA polymerase II: mammalian transcription factors alpha and beta gamma promote entry of polymerase into the preinitiation complex. *Proc. Natl. Acad. Sci. USA* **88**:6205–6209.
 16. Coulombe, B., and Z. F. Burton. 1999. DNA bending and wrapping around RNA polymerase: a “revolutionary” model describing transcriptional mechanisms. *Microbiol. Mol. Biol. Rev.* **63**:457–478.
 17. Cramer, P., D. A. Bushnell, J. Fu, A. L. Gnatt, B. Maier-Davis, N. E. Thompson, R. R. Burgess, A. M. Edwards, P. R. David, and R. D. Kornberg. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**:640–649.
 18. Cramer, P., D. A. Bushnell, and R. D. Kornberg. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**:1863–1876.
 19. Dey, A., F. Chitsaz, A. Abbasi, T. Misteli, and K. Ozato. 2003. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proc. Natl. Acad. Sci. USA* **100**:8758–8763.
 20. Douzich, M., F. Coin, J. M. Chipoulet, Y. Arai, Y. Ohkuma, J. M. Egly, and B. Coulombe. 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.
 21. Feaver, W. J., O. Gileadi, Y. Li, and R. D. Kornberg. 1991. CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* **67**:1223–1230.
 22. Finkelstein, A., C. F. Kostrub, J. Li, D. P. Chavez, B. Q. Wang, S. M. Fang, J. Greenblatt, and Z. F. Burton. 1992. A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II. *Nature* **355**:464–467.
 23. Flores, O., I. Ha, and D. Reinberg. 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.
 24. Forget, D., M. F. Langelier, C. Therien, V. Trinh, and B. Coulombe. 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.
 25. Forget, D., F. Robert, G. Grondin, Z. F. Burton, J. Greenblatt, and B. Coulombe. 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. USA* **94**:7150–7155.
 26. Fu, J., A. L. Gnatt, D. A. Bushnell, G. J. Jensen, N. E. Thompson, R. R. Burgess, P. R. David, and R. D. Kornberg. 1999. Yeast RNA polymerase II at 5 Å resolution. *Cell* **98**:799–810.
 27. Gerard, M., L. Fischer, V. Moncollin, J. M. Chipoulet, P. Chambon, and J. M. Egly. 1991. Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. *J. Biol. Chem.* **266**:20940–20945.
 28. Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A. P. Arkin, A. Astromoff, M. El Bakkoury, R. Bangham, R. Benito, S. Brachet, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K. D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kotter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Y. Wang, T. R. Ward, J. Wilhelm, E. A. Winzler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis, and M. Johnston. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**:387–391.
 29. Gnatt, A., J. Fu, and R. D. Kornberg. 1997. Formation and crystallization of yeast RNA polymerase II elongation complexes. *J. Biol. Chem.* **272**:30799–30805.
 30. Gnatt, A. L., P. Cramer, J. Fu, D. A. Bushnell, and R. D. Kornberg. 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**:1876–1882.
 31. Ha, I., W. S. Lane, and D. Reinberg. 1991. Cloning of a human gene encoding the general transcription initiation factor IIB. *Nature* **352**:689–695.
 32. Hampsey, M. 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* **62**:465–503.
 33. Hazbun, T. R., L. Malmstrom, S. Anderson, B. J. Graczyk, B. Fox, M. Riffle, B. A. Sundin, J. D. Aranda, W. H. McDonald, C. H. Chiu, B. E. Snysman, P. Bradley, E. G. Muller, S. Fields, D. Baker, J. R. Yates III, and T. N. Davis. 2003. Assigning function to yeast proteins by integration of technologies. *Mol. Cell* **12**:1353–1365.
 34. Hegde, P., R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. E. Hughes, E. Snesrud, N. Lee, and J. Quackenbush. 2000. A concise guide to cDNA microarray analysis. *BioTechniques* **29**:548–556.
 35. Ho, C. K., and S. Shuman. 1999. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol. Cell* **3**:405–411.
 36. Hodo, H. G., III, and S. P. Blattl. 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.
 37. Holstege, F. C., U. Fiedler, and H. T. Timmers. 1997. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.* **16**:7468–7480.
 38. Holstege, F. C., D. Tantin, M. Carey, P. C. van der Vliet, and H. T. Timmers. 1995. The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO J.* **14**:810–819.
 39. Holstege, F. C., P. C. van der Vliet, and H. T. Timmers. 1996. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* **15**:1666–1677.
 40. Huber, A. H., W. J. Nelson, and W. I. Weis. 1997. Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* **90**:871–882.
 41. Huh, W. K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman, and E. K. O’Shea. 2003. Global analysis of protein localization in budding yeast. *Nature* **425**:686–691.
 42. Ingles, C. J., M. Shales, W. D. Cress, S. J. Triezenberg, and J. Greenblatt. 1991. Reduced binding of TFIID to transcriptionally compromised mutants of VP16. *Nature* **351**:588–590.
 43. Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* **98**:4569–4574.
 44. Jiang, Y., S. J. Triezenberg, and J. D. Gralla. 1994. Defective transcriptional activation by diverse VP16 mutants associated with a common inability to form open promoter complexes. *J. Biol. Chem.* **269**:5505–5508.
 45. Kim, J. L., D. B. Nikolov, and S. K. Burley. 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**:520–527.
 46. Kim, Y., J. H. Geiger, S. Hahn, and P. B. Sigler. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**:512–520.
 47. Kimura, M., H. Suzuki, and A. Ishihama. 2002. Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (Pol II) complex in *Schizosaccharomyces pombe* involves direct interaction between Fcp1 and the Rpb4 subunit of Pol II. *Mol. Cell. Biol.* **22**:1577–1588.
 48. Kobor, M. S., J. Archambault, W. Lester, F. C. Holstege, O. Gileadi, D. B. Jansma, E. G. Jennings, F. Kouyoumdjian, A. R. Davidson, R. A. Young, and J. Greenblatt. 1999. An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in *S. cerevisiae*. *Mol. Cell* **4**:55–62.
 49. Kobor, M. S., L. D. Simon, J. Omichinski, G. Zhong, J. Archambault, and J. Greenblatt. 2000. A motif shared by TFIIF and TFIIB mediates their interaction with the RNA polymerase II carboxy-terminal domain phosphatase Fcp1p in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **20**:7438–7449.
 50. Krogan, N. J., M. Kim, S. H. Ahn, G. Zhong, M. S. Kobor, G. Cagney, A. Emili, A. Shilatfard, S. Buratowski, and J. F. Greenblatt. 2002. RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell. Biol.* **22**:6979–6992.
 51. Lagrange, T., A. N. Kapanidis, H. Tang, D. Reinberg, and R. H. Eubright. 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **12**:34–44.
 52. Langelier, M. F., D. Forget, A. Rojas, Y. Porlier, Z. F. Burton, and B. Coulombe. 2001. Structural and functional interactions of transcription factor (TF) IIA with TFIIE and TFIIF in transcription initiation by RNA polymerase II. *J. Biol. Chem.* **276**:38652–38657.
 53. Lu, H., L. Zawel, L. Fisher, J. M. Egly, and D. Reinberg. 1992. Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* **358**:641–645.
 54. Maldonado, E., I. Ha, P. Cortes, L. Weis, and D. Reinberg. 1990. Factors involved in specific transcription by mammalian RNA polymerase II: role of

- transcription factors IIA, IID, and IIB during formation of a transcription-competent complex. *Mol. Cell. Biol.* **10**:6335–6347.
55. Marton, M. J., J. L. DeRisi, H. A. Bennett, V. R. Iyer, M. R. Meyer, C. J. Roberts, R. Stoughton, J. Burchard, D. Slade, H. Dai, D. E. Bassett, Jr., L. H. Hartwell, P. O. Brown, and S. H. Friend. 1998. Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat. Med.* **4**:1293–1301.
 56. Mizuguchi, K., C. M. Deane, T. L. Blundell, and J. P. Overington. 1998. HOMSTRAD: a database of protein structure alignments for homologous families. *Protein Sci.* **7**:2469–2471.
 57. Murakami, K. S., S. Masuda, E. A. Campbell, O. Muzzin, and S. A. Darst. 2002. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**:1285–1290.
 58. Murakami, K. S., S. Masuda, and S. A. Darst. 2002. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**:1280–1284.
 59. Nikolov, D. B., H. Chen, E. D. Halay, A. A. Usheva, K. Hisatake, D. K. Lee, R. G. Roeder, and S. K. Burley. 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**:119–128.
 60. Nissen, R. M., and K. R. Yamamoto. 2000. The glucocorticoid receptor inhibits NFκB by interfering with serine-2 phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **14**:2314–2329.
 61. No, D., T. P. Yao, and R. M. Evans. 1996. Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. USA* **93**:3346–3351.
 62. Ohkuma, Y., S. Hashimoto, C. K. Wang, M. Horikoshi, and R. G. Roeder. 1995. Analysis of the role of TFIIE in basal transcription and TFIIF-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-α. *Mol. Cell. Biol.* **15**:4856–4866.
 63. Ohkuma, Y., and R. G. Roeder. 1994. Regulation of TFIIF ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature* **368**:160–163.
 64. Ohkuma, Y., H. Sumimoto, A. Hoffmann, S. Shimasaki, M. Horikoshi, and R. G. Roeder. 1991. Structural motifs and potential sigma homologies in the large subunit of human general transcription factor TFIIE. *Nature* **354**:398–401.
 65. Orphanides, G., T. Lagrange, and D. Reinberg. 1996. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**:2657–2683.
 66. Pan, G., and J. Greenblatt. 1994. Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. *J. Biol. Chem.* **269**:30101–30104.
 67. Peterson, M. G., J. Inostroza, M. E. Maxon, O. Flores, A. Admon, D. Reinberg, and R. Tjian. 1991. Structure and functional properties of human general transcription factor IIE. *Nature* **354**:369–373.
 68. Pokholok, D. K., N. M. Hannett, and R. A. Young. 2002. Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol. Cell* **9**:799–809.
 69. Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, and B. Seraphin. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**:218–229.
 70. Rachez, C., and L. P. Freedman. 2001. Mediator complexes and transcription. *Curr. Opin. Cell Biol.* **13**:274–280.
 71. Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**:1030–1032.
 72. Robert, F., M. Douziech, D. Forget, J. M. Egly, J. Greenblatt, Z. F. Burton, and B. Coulombe. 1998. Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. *Mol. Cell* **2**:341–351.
 73. Rodriguez, C. R., E. J. Cho, M. C. Keogh, C. L. Moore, A. L. Greenleaf, and S. Buratowski. 2000. Kin28, the TFIIF-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol. Cell. Biol.* **20**:104–112.
 74. Schaffer, A. A., Y. I. Wolf, C. P. Ponting, E. V. Koonin, L. Aravind, and S. F. Altschul. 1999. IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. *Bioinformatics* **15**:1000–1011.
 75. Serizawa, H., R. C. Conaway, and J. W. Conaway. 1992. A carboxyl-terminal-domain kinase associated with RNA polymerase II transcription factor delta from rat liver. *Proc. Natl. Acad. Sci. USA* **89**:7476–7480.
 76. Serizawa, H., T. P. Makela, J. W. Conaway, R. C. Conaway, R. A. Weinberg, and R. A. Young. 1995. Association of Cdk-activating kinase subunits with transcription factor TFIIF. *Nature* **374**:280–282.
 77. Shiekhattar, R., F. Mermelstein, R. P. Fisher, R. Drapkin, B. Dynlacht, H. C. Wessling, D. O. Morgan, and D. Reinberg. 1995. Cdk-activating kinase complex is a component of human transcription factor TFIIF. *Nature* **374**:283–287.
 78. Skaar, D. A., and A. L. Greenleaf. 2002. The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Ptilp. *Mol. Cell* **10**:1429–1439.
 79. Sluder, A. E., D. H. Price, and A. L. Greenleaf. 1988. Elongation by *Drosophila* RNA polymerase II. Transcription of 3'-extended DNA templates. *J. Biol. Chem.* **263**:9917–9925.
 80. Sopta, M., R. W. Carthew, and J. Greenblatt. 1985. Isolation of three proteins that bind to mammalian RNA polymerase II. *J. Biol. Chem.* **260**:10353–10360.
 81. Sumimoto, H., Y. Ohkuma, E. Sinn, H. Kato, S. Shimasaki, M. Horikoshi, and R. G. Roeder. 1991. Conserved sequence motifs in the small subunit of human general transcription factor TFIIE. *Nature* **354**:401–404.
 82. Takahashi, Y., J. B. Rayman, and B. D. Dynlacht. 2000. Analysis of promoter binding by the E2F and pRB families in vivo: distinct E2F proteins mediate activation and repression. *Genes Dev.* **14**:804–816.
 83. Tirode, F., D. Busso, F. Coin, and J. M. Egly. 1999. Reconstitution of the transcription factor TFIIF: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol. Cell* **3**:87–95.
 84. Vassilyev, D. G., S. Sekine, O. Laptenko, J. Lee, M. N. Vassilyeva, S. Borukhov, and S. Yokoyama. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**:712–719.
 85. Westover, K. D., D. A. Bushnell, and R. D. Kornberg. 2004. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* **303**:1014–1016.
 86. Wu, L. F., T. R. Hughes, A. P. Davierwala, M. D. Robinson, R. Stoughton, and S. J. Altschul. 2002. Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. *Nat. Genet.* **31**:255–265.
 87. Yang, Y. H., and T. Speed. 2002. Design issues for cDNA microarray experiments. *Nat. Rev. Genet.* **3**:579–588.
 88. Young, R. A. 1991. RNA polymerase II. *Annu. Rev. Biochem.* **60**:689–715.
 89. Zhang, G., E. A. Campbell, L. Minakhin, C. Richter, K. Severinov, and S. A. Darst. 1999. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**:811–824.

CHAPITRE 3

L'ANALYSE SYSTÉMATIQUE DU RÉSEAU D'INTERACTIONS PROTÉIQUES DE LA MACHINERIE DE TRANSCRIPTION RÉVÈLE L'IDENTITÉ DE L'ENZYME DE COIFFAGE DU snARN 7SK

3.1 Préambule

Pour ce manuscrit, j'ai réalisé la purification des complexes protéiques pour 16 des 32 polypeptides étiquettes TAP selon la méthode schématisée à la figure 1A, ainsi que pour trois autres facteurs (DDX5, HNRPU et XAB1) en collaboration avec Cynthia Thérien (figures 1C et 1D). J'ai participé intellectuellement à la conception de l'outil informatique (IR score) utilisé dans cette étude, qui a été réalisé par Mathieu Blanchette (figure 1B). J'ai effectué les analyses bioinformatiques permettant la validation des interactions protéiques à l'aide de la base de données *Proteus* développée par Christian Poitras et Dominique Bergeron. J'ai généré la carte d'interactions des complexes protéiques présentée à la figure 2, en collaboration avec Mathieu Blanchette. J'ai réalisé le réseau d'interactions, l'expérience de filtration sur gel et les analyses bioinformatiques pour le groupe de protéines à fonction inconnue XAB1-RPAPs (figures 3A, 3B et 3C). J'ai effectué tous les essais de caractérisation de BCDIN3 (figures 4, 5, 6 et 7), à l'exception de l'expérience présentée à la figure 4C, montrant ainsi pour la première fois que cette méthyltransférase jusqu'ici à fonction inconnue est responsable de l'ajout de la coiffe en 5' du petit ARN nucléaire (snARN) 7SK, ce qui lui confère de la stabilité dans les cellules. J'ai rédigé l'article avec le Dr Benoit Coulombe.

Diane Forget et Annie Bouchard ont purifié 13 des 32 complexes protéiques étudiés (figures 1C et 1D). Diane Forget a aussi réalisé l'analyse bioinformatique des complexes TAF10, TAF13 et TRFP. Les Drs Guy G. Poirier et Sylvie Bourassa ont

effectué les analyses de MS. Les Drs Jack Greenblatt et Benoit Chabot ont fourni certains cDNAs pour les facteurs étudiés. La figure 3D a été réalisée par Gordon Chua, stagiaire postdoctoral chez le Dr Timothy Hughes et la figure 4C par Qintong Li, étudiant chez le Dr David Price.

3.2 Article

Jerónimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J., Chabot, B., Poirier, G.G., Hughes, T.R., Blanchette, M., Price, D.H. and Coulombe, B. (2007) Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. **Mol. Cell.**, 27(2):262-274.

Systematic Analysis of the Protein Interaction Network for the Human Transcription Machinery Reveals the Identity of the 7SK Capping Enzyme

Célia Jeronimo,¹ Diane Forget,¹ Annie Bouchard,¹ Qintong Li,² Gordon Chua,³ Christian Poitras,¹ Cynthia Thérien,¹ Dominique Bergeron,¹ Sylvie Bourassa,⁴ Jack Greenblatt,³ Benoit Chabot,⁵ Guy G. Poirier,⁴ Timothy R. Hughes,³ Mathieu Blanchette,⁶ David H. Price,² and Benoit Coulombe^{1,*}

¹Laboratory of Gene Transcription and Proteomics Discovery Platform, Institut de recherches cliniques de Montréal, 110 avenue des Pins Ouest, Montréal, QC H2W 1R7, Canada

²Biochemistry Department, University of Iowa, Iowa City, IA 52242-1109, USA

³Banting and Best Department of Medical Research, University of Toronto, Toronto, ON M5G 1L6, Canada

⁴Centre hospitalier universitaire de Québec, Université Laval, Québec, QC G1V 4G2, Canada

⁵Département de microbiologie et infectiologie, Université de Sherbrooke, Sherbrooke, QC J1H 5N4, Canada

⁶McGill Centre for Bioinformatics, McGill University, Montréal, QC H3A 2B4, Canada

*Correspondence

DOI 10.1016/j.molcel.2007.06.027

SUMMARY

We have performed a survey of soluble human protein complexes containing components of the transcription and RNA processing machineries using protein affinity purification coupled to mass spectrometry. Thirty-two tagged polypeptides yielded a network of 805 high-confidence interactions. Remarkably, the network is significantly enriched in proteins that regulate the formation of protein complexes, including a number of previously uncharacterized proteins for which we have inferred functions. The RNA polymerase II (RNAP II)-associated proteins (RPAPs) are physically and functionally associated with RNAP II, forming an interface between the enzyme and chaperone/scaffolding proteins. BCDIN3 is the 7SK snRNA methylphosphate capping enzyme (MePCE) present in an snRNP complex containing both RNA processing and transcription factors, including the elongation factor P-TEFb. Our results define a high-density protein interaction network for the mammalian transcription machinery and uncover multiple regulatory factors that target the transcription machinery.

INTRODUCTION

The information contained in mammalian genomes is decoded through the concerted action of a myriad of proteins that recognize and interpret specific signals at active genomic sites in such a way that mRNA precursors are synthesized and processed to form mature mRNA that

will then serve for protein synthesis. The machinery required for gene transcription includes RNA polymerase II (RNAP II) and its general transcription factors TFIIA, TFIIB, TFIID/STAGA, TFIIE, TFIIIF, TFIIH, TFIIS, and Mediator (Hampsey, 1998; Hahn, 2004; Coulombe and Burton, 1999). Many additional proteins regulating the activity of RNAP II at various stages of the transcription reaction have also been purified and characterized from various organisms. Among the regulatory proteins, many were found to modulate the structure of chromatin, the substrate of RNAP II in vivo (Smith and Peterson, 2005; Fischle et al., 2003).

Although information on the composition, function, and mechanisms of these protein complexes has accumulated in the literature, the protein-protein interactions that participate in their functional recruitment at their site of action along genomic DNA are often elusive. Emerging evidence shows that the RNAP II transcription apparatus is involved in this process, particularly through the CTD of RPB1, the largest subunit of RNAP II, which was shown to bind several regulatory complexes at various stages of transcription (Hampsey and Reinberg, 2003; Hirose and Manley, 2000). The CTD was found to interact with transcription elongation factors, chromatin-modifying complexes, pre-mRNA maturation enzymes, and other proteins, providing a molecular basis for the coupling between transcription and RNA processing. The pre-mRNA processing events, including 5' end capping, splicing, and 3' end cleavage/polyadenylation, were shown to occur cotranscriptionally (Proudfoot et al., 2002; Bentley, 2005; Aguilera, 2005). Indeed, during the transcription cycle, CTD phosphorylation stimulates efficient recruitment of 5' and 3' end processing factors to the pre-mRNA. The splicing machinery is comprised of small-nuclear ribonucleoprotein particles (snRNPs), each composed of various proteins and a specific snRNA (U1, U2, U4, U5, and U6 snRNAs), which together form the

spliceosome responsible for performing the catalytic events leading to exon excision and intron ligation (Zhou et al., 2002; Jurica and Moore, 2003). Many proteins such as the hnRNPs and the SR proteins regulate splice site selection (Black, 2003; Shin and Manley, 2004).

The spliceosomal U1-U5 snRNA genes are transcribed by RNAP II while U6 is synthesized by RNA polymerase III (RNAP III) (reviewed in Kiss [2004]). In contrast to the RNAP II-synthesized snRNAs, which receive a 7-methylguanosine (m⁷G) cap during transcription and are subsequently methylated posttranscriptionally to form the trimethylguanosine (TMG) cap, RNAP III-transcribed snRNAs acquire a methyl group on the gamma-phosphate of their 5' end posttranscriptionally (Reddy et al., 1992). Another snRNA synthesized by RNAP III, 7SK, is an abundant 331 nucleotide transcript found in association with the positive transcription elongation factor b (P-TEFb) (Nguyen et al., 2001; Yang et al., 2001). P-TEFb, which is composed of CDK9 and CCNT1/cyclin T1 or CCNT2/cyclin T2, stimulates the elongation phase of transcription by phosphorylating the CTD of RNAP II and reversing the effects of negative elongation factors (for recent reviews, see Peterlin and Price [2006] and Zhou and Yik [2006]). P-TEFb not only plays an important role in the transcription of cellular genes but is also a key factor for the expression of the human immunodeficiency virus type 1 (HIV-1) genome (Zhu et al., 1997). Previous studies have shown that a complex formed of the 7SK snRNA and proteins termed HEXIMs can interact with P-TEFb and repress its stimulatory activity on transcriptional elongation by RNAP II (Barboric et al., 2005; Blazek et al., 2005; Byers et al., 2005; Li et al., 2005; Yik et al., 2005).

Although the protein complexes actively involved in transcription on genomic DNA have been extensively studied, very little is known about the organization of the transcription machinery in the soluble nuclear compartment, prior to or following its association with genomic DNA, where its formation and recycling likely take place. To determine the composition and organization of the soluble machinery for mRNA formation and obtain insight into its mechanisms of formation and/or recycling, we performed a survey of the soluble protein complexes comprising key components of the RNAP II machinery. Tagged proteins were purified and their associating partners identified by mass spectrometry. High-confidence interactions were selected computationally and used to draw a map of interactions connecting these various complexes. The composition and organization of this network revealed important features of the machinery involved in mRNA formation in eukaryotes.

RESULTS

A Network of Protein Complexes Involving the Human Transcription and RNA Processing Machineries

To survey soluble protein complexes that comprise the transcription machinery and define their interaction

network, several subunits of RNAP II and basal transcription factors carrying a tandem affinity purification (TAP) tag were purified and their interaction partners identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (see Figure 1A for an overview of our procedure and see Figure S1A in the Supplemental Data available with this article online for a schematic representation of the expression system used). A number of identified interaction partners, including RNA processing factors, were also tagged and submitted to the same procedure, thereby enriching the interaction data set (see Figure S1B for examples). Of note, all the tagged transcription factors that we purified and tested were active in transcription reactions *in vitro* and were present in transcribed regions in chromatin immunoprecipitation (ChIP) experiments *in vivo* (M. Cojocaru, C.J., D. F., A.B., D.B., P. Côte, G.G.P., J.G., and B. Coulombe, unpublished data), indicating that bona fide, functional protein complexes were purified.

To select high-confidence interactions and minimize the number of false positives in our data set, we developed a computational method that, first, filters out spurious interactions caused by proteins that bind nonspecifically to our columns and very abundant cellular proteins that may have remained as contaminants after affinity purification and, second, attributes an interaction reliability (IR) score to each protein interaction detected by mass spectrometry (see the Supplemental Data for a description of the method; Figure S2 illustrates the overall procedure, and the Supplemental Experimental Procedures describes the computational algorithm used).

Figure 1B illustrates the ability of our IR score to separate true protein interactions from likely false positives. Of the 2008 putative protein interactions detected by mass spectrometry (see Table S1), 805 were judged highly reliable by our classifier (IR score above 0.6729) (Figure 1C), involving a total of 436 proteins. This set of highly reliable interactions contains 83% of the interactions supported by the literature, while excluding 83% of those that were judged likely false positives. As summarized in Figure 1C, 170 successful affinity purifications using 32 tagged proteins yielded 2008 putative protein interactions (see Figure 2 and Table S1 for a list of the tagged proteins). The 805 distinct interactions with an IR score above 0.6729 constitute a network of protein interactions in which we have high confidence.

Figure 2 shows the interaction network for the protein complexes comprising all the tagged transcription and RNA processing factors used in our experiments and their interaction partners. Not surprisingly, gene ontology (GO) annotations (Gene-Ontology database, <http://www.geneontology.org>) revealed that many proteins of the network are either transcription factors (Figure 2; blue nodes) or RNA processing factors (Figure 2; pink nodes), which are sometimes interconnected as should be expected for two closely related cellular machineries. Unexpectedly, however, GO annotations also revealed that a significant number of network components are proteins involved in

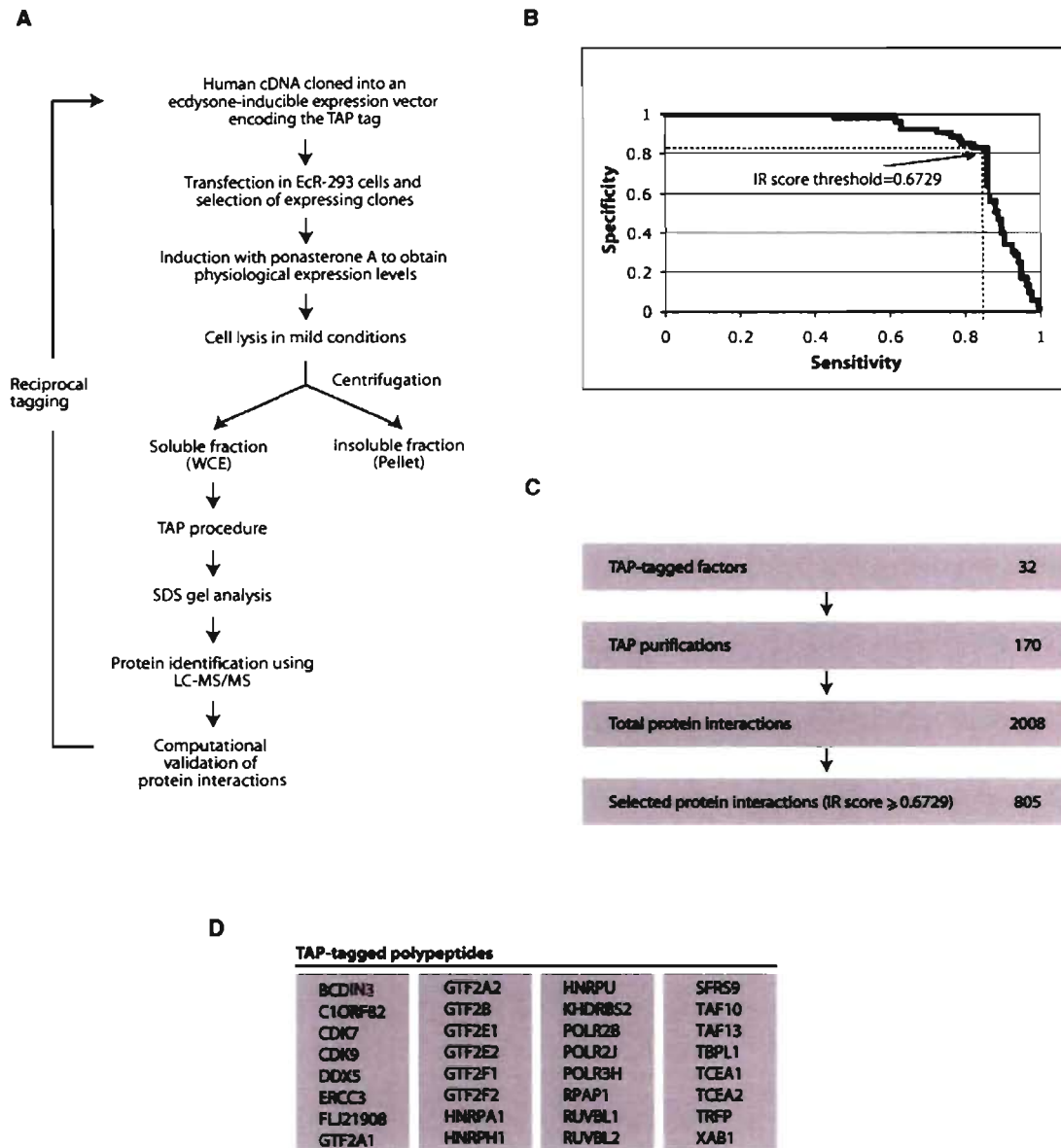


Figure 1. A Method for the Analysis of Protein Complexes in Human Cells

(A) Flow chart of the procedure used for the affinity purification and mass spectrometry identification of protein complexes in HEK293 cells. (B) Computational validation of protein interactions. Sensitivity/specificity ROC curve, parameterized by the IR score, for the discrimination between literature-supported interactions and dubious interactions (see Table S1). The chosen IR score threshold results in a sensitivity of 83% and a specificity of 83%. The IR was calculated based on the method described in the Supplemental Experimental Procedures.

(C) Summary of the overall screen for protein interactions.

(D) List of the 32 TAP-tagged polypeptides used in this study.

modulating the formation of protein complexes (P value = $1.8e-5$) (see Figure 2; green nodes). Importantly, our protein network also contains several proteins that are in the databases but have not been previously characterized (Figure 2; yellow nodes).

To further assess the efficiency of our affinity-purification method for the discovery of functionally relevant protein interaction partners, five proteins occupying key

positions in the overall interaction network, namely RNAP II-associated protein (RPAP)1, C1ORF82, FLJ21908, XAB1, and BCDIN3, were used in experiments aimed at validating their network connections both physically (reciprocal tagging and gel filtration) and functionally (bioinformatics, silencing of the yeast homolog, and biochemical assays). These five proteins were selected because our initial bioinformatic analysis revealed that they may

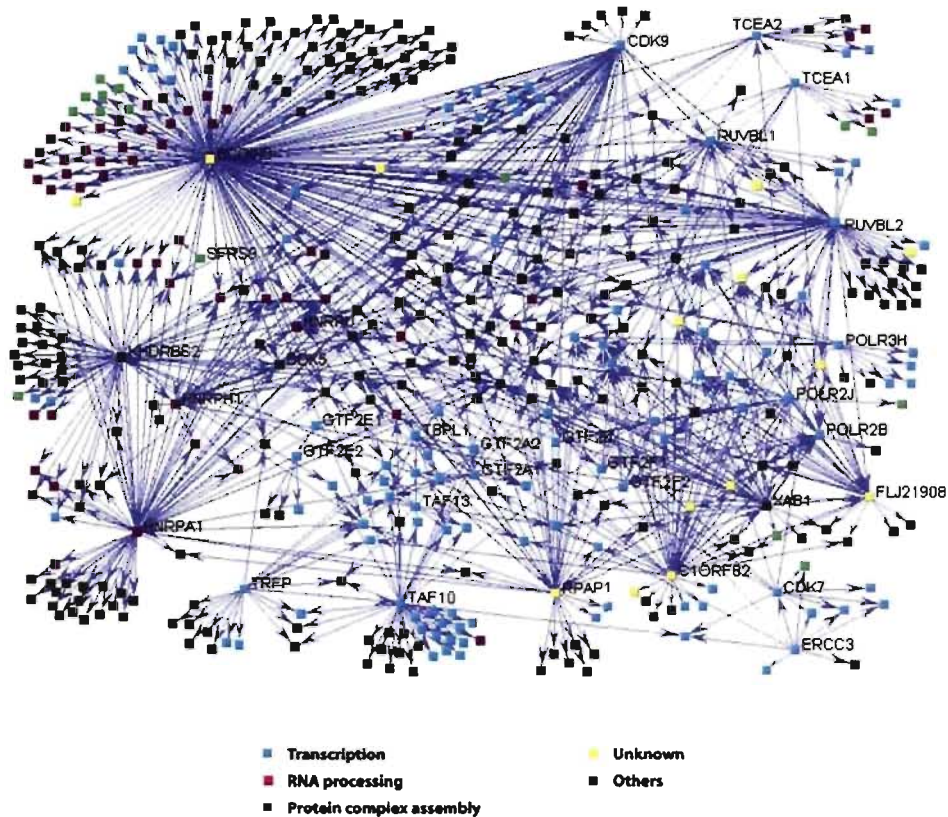


Figure 2. A Network of Protein Complexes Involving the RNAP II Basal Transcription Machinery

Overview of the interaction map showing the 805 validated interactions obtained using 32 tagged polypeptides. The interactions are represented as directional edges extending from the tagged protein using the web-based software VisANT (<http://visant.bu.edu>). The thickness of the line that connects two proteins is proportional to the IR score obtained for this interaction (see Table S1). The nodes are colored according to their GO annotations (<http://www.geneontology.org>). The tagged proteins are indicated.

represent unprecedented regulatory factors targeting two key components of the transcription machinery, RNAP II and P-TEFb. As shown in a previous study in yeast (Krogan et al., 2004), the affinity-purification procedure used is also efficient in identifying RNA components of the purified complexes.

The RPAPs-XAB1 Physically and Functionally Connect RNAP II to Regulators of Protein Complex Formation

Our protein interaction network reveals the existence of a group of four polypeptides, RPAP1 (Jeronimo et al., 2004), XAB1, C19ORF82, and FLJ21908, that are tightly connected to the enzyme RNAP II itself in the soluble nuclear compartment (Figure 3A). Gel filtration experiments using the tagged-RPB11/POLR2J eluate confirmed that RPAP1, FLJ21908, and XAB1 cofractionate with RNAP II (Figure 3B). Because of their physical association with RNAP II, which may be direct or indirect, the two previously uncharacterized proteins C19ORF82 and FLJ21908 were named RPAP2 and RPAP3, respectively. The RPAPs-XAB1 also interact with the regulatory complex

Mediator, GRINL1A/Gdown1, a recently discovered RNAP II interaction partner that mediates Mediator responsiveness during transcriptional activation (Hu et al., 2006), and the Integrator, a complex involved in the processing of RNAP II-synthesized snRNAs (Baillat et al., 2005) (Figure 3A).

Examination of the protein interaction network further reveals that the RPAPs-XAB1 are placed at the interface between RNAP II and a group of proteins able to assemble into molecular chaperone complexes, including the pre-foldins (C19ORF2/URI, PFDN6, PFDN2, and UXT) (Gstaiger et al., 2003; Vainberg et al., 1998; Zhao et al., 2005) and the AAA+ chaperone-like ATPases RUVBL1 and RUVBL2 (Kanemaki et al., 1999; Ikura et al., 2000) (Figure 3A). Notably, a bioinformatics analysis revealed that motifs unique to chaperones and scaffolding proteins are also present in the RPAPs-XAB1 proteins themselves (Figure 3C). RPAP1 contains an ARM domain (Jeronimo et al., 2004), which is a structural element known to mediate protein-protein interactions (Huber et al., 1997). FLJ21908/RPAP3 contains helix-turn-helix tetratricopeptide repeat (TPR) motifs, which characterize cofactors of

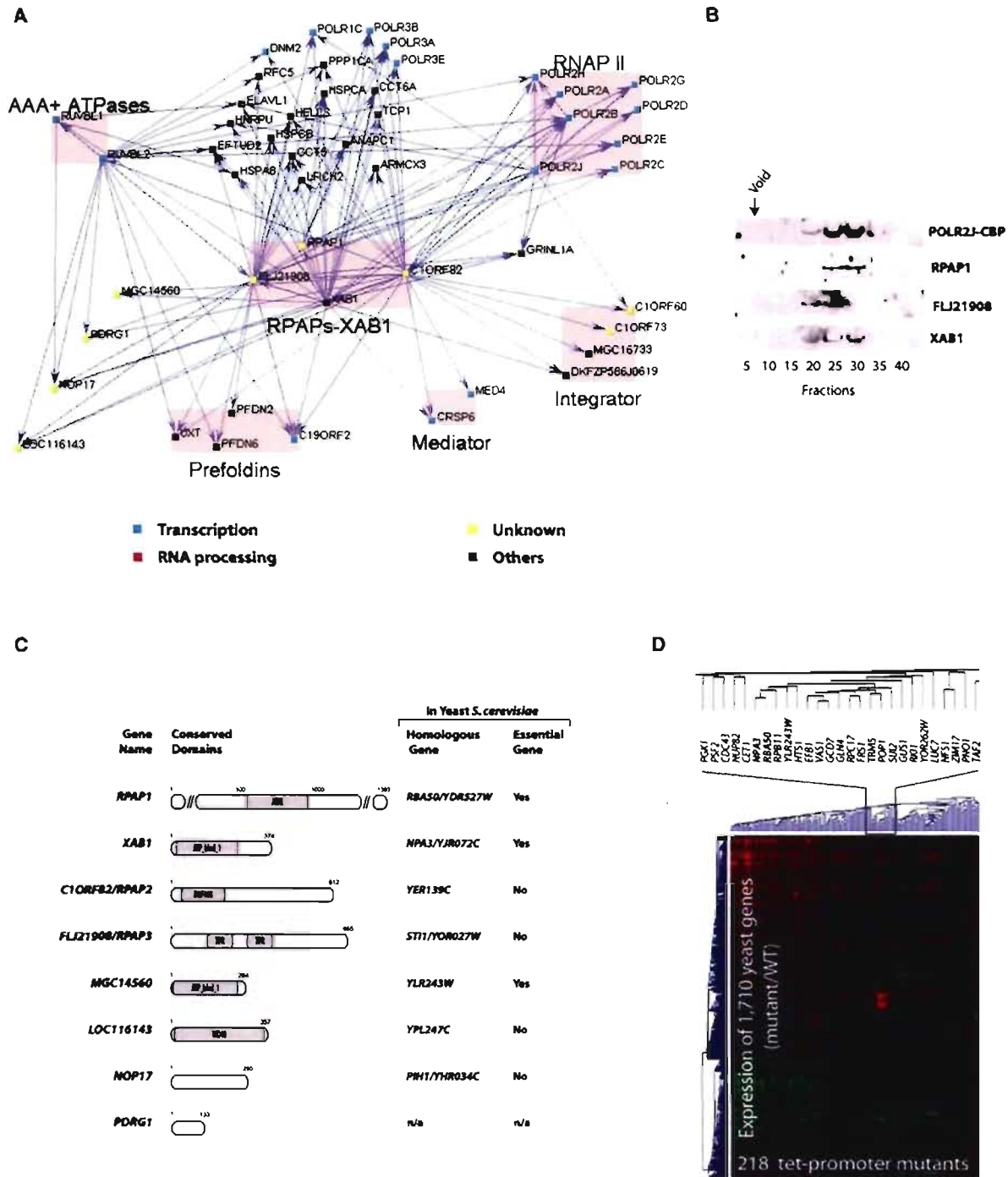


Figure 3. Previously Uncharacterized Protein Complexes Containing RNAP II in Association with Regulators of Protein Complex Assembly

(A) Network highlighting the interactions of RPAPs-XAB1 with RNAP II, the regulatory complexes Integrator and Mediator, and a group of proteins with chaperone/scaffolding activity.

(B) Elution profile of RNAP II and the RPAPs-XAB1 in gel filtration.

(C) Previously uncharacterized proteins are associated with RNAP II. For each human protein, a representation with putative functional domains provided by the NCBI's conserved domain database, the name of the closest yeast homologous gene as determined by a BLASTP search on the June 2005 assembly version of the UCSC genome browser, and gene essentiality in yeast provided by the SGD are shown. *ARM*, ARM repeat protein

chaperones (Smith, 2004). XAB1 is part of a family of polypeptides containing an ATP-binding domain, termed "ATP_bind_1," that are conserved among eukaryotes and archaea; of note, energy-driven conformational changes are central to the formation of many protein complexes. Although previous reports have shown that human XAB1 binds to the DNA repair factor XPA and participates in transcriptional regulation at methylated promoters (Nitta et al., 2000; Lembo et al., 2003), its mechanism of action remains undetermined. The RPAPs-XAB1 are also associated with a number of additional previously uncharacterized proteins (MGC14560, LOC116143, NOP17, and PDRG1) (Figure 3A). MGC14560 is another member of the ATP_bind_1 family, and LOC116143 contains a WD40 domain (Figure 3C), which is believed to coordinate multiprotein complex assemblies (Smith et al., 1999). Our results demonstrate that RNAP II is associated with a number of proteins known or predicted to regulate protein complex formation.

A number of interactions among the RPAPs-XAB1, RNAP II, and their other human interaction partners are also observed in yeast (*Saccharomyces* genome database, SGD). Interestingly, silencing the yeast homologs of XAB1 and MGC14560, namely *NPA3* and *YLR243W*, which are both essential for cell growth and are known to physically interact in yeast (see Figure 3C and SGD), had global effects on yeast gene expression similar to those of silencing either the RNAP II subunit *RPB11* or *RBA50* (the closest homologous gene to RPAP1 in yeast) (Figure 3D). These results along with their position at the interface of RNAP II regulatory complexes and molecular chaperones/scaffolding proteins suggest a role for the RPAPs-XAB1 in the formation of multicomponent transcription complexes.

A 7SK snRNP Complex Is Comprised of RNA Processing and Transcription Factors, Including the Elongation Factor P-TEFb, and the Previously Uncharacterized Protein BCDIN3

Previous studies have shown that two forms of P-TEFb complexes, one being large and inactive and the other being small and active, can be purified from mammalian cells. Indeed, the cyclin-dependent kinase activity of P-TEFb is inhibited upon binding of proteins termed HEXIMs and the 7SK snRNA (Barboric et al., 2005; Blazek et al., 2005; Byers et al., 2005; Li et al., 2005; Yik et al., 2005). P-TEFb also binds to the bromodomain protein Brd4 to form the transcriptionally active P-TEFb complex (Jang et al., 2005; Yang et al., 2005). Our protein interaction network revealed that, in the soluble fraction, besides its positive (Brd4) and negative (HEXIMs and 7SK snRNA) regulators, P-TEFb is associated with many other proteins including RNA processing factors, SART3 (also referred to

as Tip110), and the previously uncharacterized protein BCDIN3 (*Bicoid-interacting 3*, homolog [*Drosophila*]) (see Figure 4A and Table S1). Reciprocal tagging of BCDIN3 confirmed these interactions. Interestingly, SART3, a surface antigen in several cancers (Yang et al., 1999), has independently been described as (1) a coregulator of HIV-1 gene expression that directly associates with Tat (Liu et al., 2002), and (2) a negative regulator of transcriptional activation by the androgen receptor (Liu et al., 2004).

The affinity-purified complex was further characterized by running the BCDIN3 eluate on a gel filtration column (Figure 4B). This experiment confirmed that P-TEFb, HEXIM1, and SART3 cofractionate with BCDIN3. Because subunits of P-TEFb and the HEXIMs were found associated with BCDIN3, we reasoned that 7SK might also be present, because P-TEFb and HEXIM1 do not interact in the absence of 7SK (Li et al., 2005; Michels et al., 2003; Yik et al., 2003). RNA was isolated from a portion of a BCDIN3 eluate and probed for the presence of 7SK, U6, and U2 snRNAs. The probes specifically recognized their respective cellular RNAs from total HeLa cell RNA, but only 7SK and U6 were found in the BCDIN3-tagged eluate (Figure 4C). Both 7SK and U6 RNAs were partially degraded, presumably due to extensive incubations utilized in the affinity-purification procedure.

To assess the roles of protein-RNA and protein-protein interactions in maintaining the integrity of the 7SK snRNP, the complex was submitted to extensive digestion with RNase A prior to its purification on our affinity columns. This treatment led to a partial dissociation of SART3 and HEXIM1 (Figure 4D), suggesting that RNA-protein interactions contribute to, but are not essential for, the association of these proteins with the complex. This experiment also ruled out the possibility that our observed interactions are caused by contaminating RNA molecules associated nonspecifically with some protein components of the complex. The association of P-TEFb (and most other proteins of the eluate; data not shown) with tagged BCDIN3 is resistant to RNase A treatment (Figure 4D), indicating that P-TEFb and BCDIN3 are associated mainly through direct or indirect protein-protein interactions.

BCDIN3 Is a Conserved 7SK snRNA Methylphosphate Capping Enzyme

Bioinformatics analysis revealed that BCDIN3 contains a consensus S-adenosyl methionine (AdoMet)-binding domain (Figure 5A), AdoMet being the methyl donor used by methyltransferases (Lu, 2000). As in the other members of the major class (class I) of AdoMet-dependent methyltransferases, BCDIN3 has the four conserved amino acid sequence motifs (I, Ia, II, and III) that form a seven-strand twisted beta sheet (Kagan and Clarke, 1994). A database search with BCDIN3 revealed the

superfamily; *ATP_bind_1*, conserved hypothetical ATP-binding protein; *DUF408*, domain of unknown function; *TPR*, tetratricopeptide repeat domain; *WD40*, WD40 domain.

(D) Clustering analysis of mRNA expression profiles with tet-promoter alleles corresponding to *RBA50*, *NPA3*, *YLR243W*, and *RPB11*, in comparison to profiles from 214 other tet-promoter mutants representing diverse cellular functions (Mnaimneh et al., 2004).

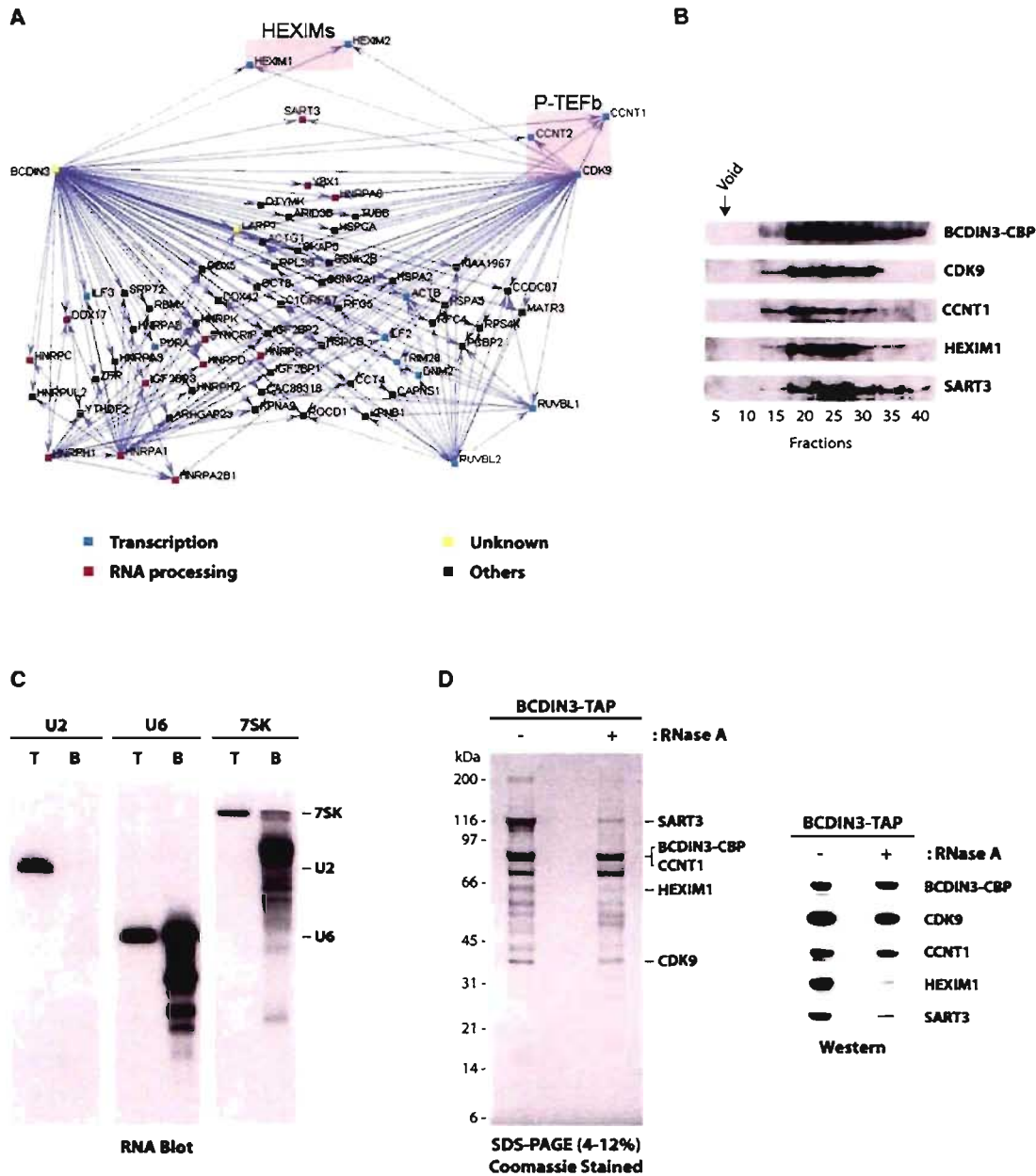


Figure 4. A Complex Comprising the Elongation Factor P-TEFb, the 7SK and U6 snRNAs, and the Previously Uncharacterized Protein BCDIN3

(A) Network highlighting the interactions of BCDIN3 and CDK9 with various RNA processing factors, P-TEFb (cyclin T1/cyclin T2 subunits), and its regulatory factors HEXIM1 and HEXIM2.

(B) Gel filtration analysis showing that P-TEFb, HEXIM1, and the RNA processing factor SART3 cofractionate with BCDIN3.

(C) The 7SK and U6 snRNAs are present in the BCDIN3-TAP eluate. RNA blots were performed on total RNA extracted from HeLa cells (T; 900 ng) or RNA extracted from the BCDIN3-TAP eluate (B; 60 ng) and probed with RNA oligos specific for U2, U6, or 7SK snRNAs. Migration of the three intact RNA species is indicated.

(D) The 7SK snRNP complex is partly resistant to extensive digestion with RNase A. SDS gel showing BCDIN3 affinity eluates prepared from cell extracts treated or not with RNase A. Western blots confirming the presence of BCDIN3, P-TEFb, HEXIM1, and SART3 in both eluates are shown.

existence of homologous proteins in metazoans (*Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans*), plants (*Arabidopsis thaliana*), and lower eukary-

otes (*Schizosaccharomyces pombe*) (Figure 5B; see **Experimental Procedures** for homolog accession numbers). The *Drosophila* BCDIN3 homologous protein, named

Proteomics Survey Reveals the 7SK Capping Enzyme

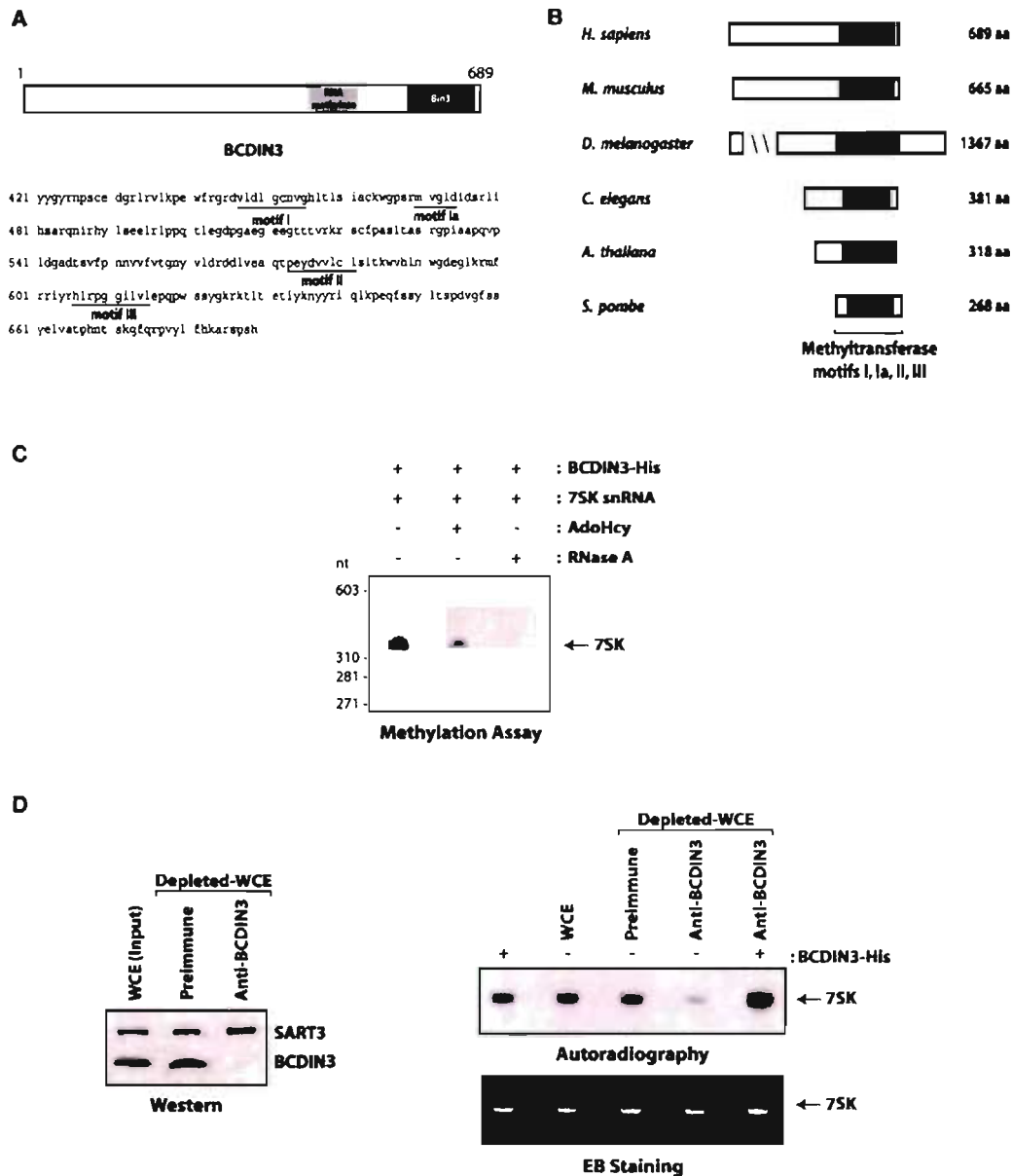


Figure 5. BCDIN3 Is a Conserved Methyltransferase that Targets the 7SK snRNA

(A) Linear representation of BCDIN3 with its *RNA methylase* and *Bin3* domains provided by the NCBI's conserved domain database. The amino acid sequence of the C-terminal region of BCDIN3 containing the putative S-adenosyl methionine (AdoMet)-binding domain is shown. The signature seven β strand methyltransferase motifs I, II, and III are indicated.

(B) Evolutionary conservation of human BCDIN3. A schematic representation of BCDIN3 and its protein homolog according to the NCBI's HomoloGene system is provided. The conserved regions containing the methyltransferase motifs I, II, and III are indicated.

(C) The 7SK snRNA is methylated *in vitro* by BCDIN3. Recombinant 7SK snRNA was incubated with His-tagged BCDIN3 and ^3H -AdoMet in the presence of the buffer alone, AdoHcy, or RNase A. The ^3H -methylated RNA product was detected by autoradiography.

(D) Immunodepletion of BCDIN3 impairs the 7SK methylation activity of whole-cell extracts and addition of recombinant BCDIN3-His to the depleted extract rescues 7SK methylation. HEK293 whole-cell extracts (WCE) were incubated or not with protein A Sepharose beads coupled to either pre-immune or anti-BCDIN3 serum, and supernatants were immunoblotted with antibodies directed against BCDIN3 and SART3 (as a loading control). WCE incubated or not with beads coated with preimmune or anti-BCDIN3 serum were assayed for methylation of recombinant 7SK snRNA in the absence or presence of His-tagged BCDIN3. RNA was visualized by ethidium bromide (EB) staining (lower panel), and the ^3H -methylated product was detected by autoradiography (upper panel).

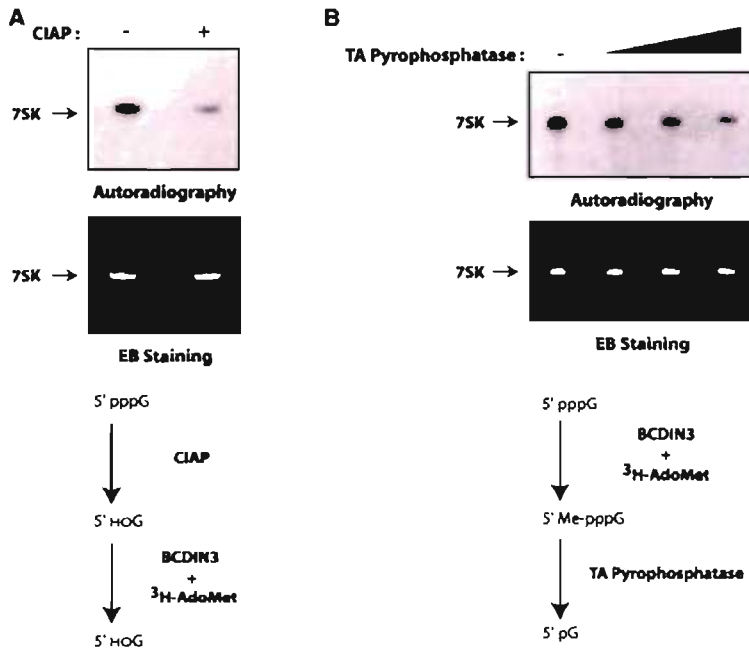


Figure 6. BCDIN3 is the 7SK snRNA MePCE

(A) Dephosphorylation of the 7SK snRNA impairs its methylation by BCDIN3. Recombinant 7SK snRNA was treated with (+) or without (–) calf intestinal alkaline phosphatase (CIAP) prior to its incubation with His-tagged BCDIN3 and ³H-AdoMet. The RNA was visualized by ethidium bromide (EB) staining (lower panel), and the methylated product was detected by autoradiography (upper panel). A schematic representation of the assay is shown.

(B) Enzymatic decapping of BCDIN3-methylated 7SK removes the radiolabeled methyl group. Recombinant 7SK was incubated with His-tagged BCDIN3 and ³H-AdoMet and then decapped or not (–) with different amounts of tobacco acid (TA) pyrophosphatase. The RNA was visualized by ethidium bromide (EB) staining (lower panel), and the methylated product was detected by autoradiography (upper panel). A schematic representation of the assay is shown.

Bin3 (*Bicoid-interacting protein 3*), was previously identified in association with Bicoid, a DNA-binding factor that directs pattern formation in the early developing embryo (Zhu and Hanes, 2000). BCDIN3 and its homologous proteins show the highest homology in their C-terminal regions, which contain the conserved AdoMet-dependent methyltransferase motifs (Figure 5B), as well as the Bin3 domain (Figure 5A). Notably, a BCDIN3 homolog is absent in *S. cerevisiae* (Figure 5B).

To start addressing the function of this previously uncharacterized methyltransferase, we searched for specific substrates within members of the 7SK snRNP complex. Because BCDIN3 harbors a putative *RNA methylase* domain (NCBI's conserved domain database) (Figure 5A), we investigated whether the 7SK snRNA could serve as a substrate for BCDIN3 in an *in vitro* methylation assay. His-tagged BCDIN3 expressed and purified from bacteria was incubated with recombinant 7SK in the presence of radiolabeled AdoMet (³H-AdoMet). Figure 5C shows that the 7SK is methylated by BCDIN3-His. Moreover, competition with AdoHcy, a potent competitive inhibitor of a number of methyltransferases (Ueland, 1982), confirmed the enzymatic nature of the BCDIN3 methylation reaction.

To confirm that BCDIN3 is a bona fide 7SK methyltransferase in human cells, we depleted endogenous BCDIN3 from HEK293 whole-cell extracts using a specific antibody and assessed the effect on the ability of the extract to support 7SK methylation. As shown in Figure 5D, the depletion of BCDIN3 resulted in a marked decrease in the 7SK methylation activity of the extract. Addition of recombinant BCDIN3 to the depleted extract fully restored the methylation activity. These results indicate that BCDIN3 is an active 7SK methyltransferase in human cell extracts.

Previous studies have shown that RNAP III-synthesized snRNAs such as 7SK are capped posttranscriptionally through the addition of a methyl group directly on the gamma-phosphate of their first 5' nucleotide (Gupta et al., 1990; Shumyatsky et al., 1990). However, up to now the enzyme responsible for this function had not been identified. Consequently, BCDIN3 is a candidate for being a methylphosphate capping enzyme. To test this hypothesis, we performed two independent experiments. First, recombinant 7SK was treated with calf intestinal alkaline phosphatase (CIAP), specifically removing the phosphate groups from the 5' end of RNA, prior to its incubation with His-tagged BCDIN3 in the *in vitro* methylation assay. Treatment of 7SK with CIAP decreased the methylation signal (Figure 6A, upper panel), showing that the presence of the 5' phosphates is required for efficient methylation by BCDIN3. Second, the 7SK snRNA was methylated *in vitro* by BCDIN3-His and the reaction product treated with tobacco acid (TA) pyrophosphatase, an enzyme that specifically removes the gamma and beta phosphates from the 5' end of RNA (Shinshi et al., 1976). Figure 6B shows that treatment with this decapping enzyme removes the radiolabeled methyl from the 7SK. To ensure that equal amounts of RNA were used and that RNA was not degraded during the dephosphorylation and decapping assays, the gels containing the methylation reactions were stained with ethidium bromide prior to their autoradiography (Figures 6A and 6B, lower panels). Essentially the same results were observed using endogenous BCDIN3 (data not shown). Taken together, these results demonstrate that BCDIN3 is capable of acting as a 7SK methylphosphate capping enzyme, hereby renamed MePCE.

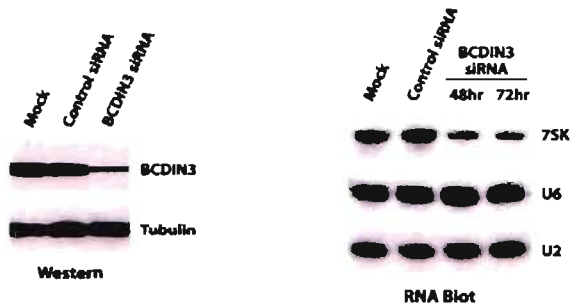


Figure 7. BCDIN3 Silencing Decreases the Steady-State Level of Cellular 7SK

BCDIN3 silencing was monitored by western blotting using extracts from HEK293 cells mock transfected or transfected with control or BCDIN3 siRNA (tubulin was used as a loading control). 7SK steady-state levels were assessed by RNA blotting using total RNA extracted from HEK293 cells mock transfected or transfected with control or BCDIN3 siRNA (48 hr and 72 hr posttransfection) and probed with RNA oligonucleotides that specifically detect 7SK, U6, or U2 snRNAs.

BCDIN3/MePCE Silencing Decreases the Steady-State Level of Cellular 7SK

Previous studies have shown that 7SK capping through gamma-phosphate methylation is important for protecting the RNA from exonucleolytic degradation. Indeed, it has been shown that the cap structure enhances the stability of U6 and 7SK snRNAs (Shumyatsky et al., 1990) and that uncapped U6 snRNA is rapidly degraded (Hamm et al., 1990). Moreover, other known cap structures such as m⁷G and TMG also enhance the half-life of RNAs (Cougnot et al., 2004). To determine whether BCDIN3/MePCE is involved in regulating 7SK steady-state levels, we used small interfering RNA (siRNA) to silence its expression in HEK293 cells and analyzed the effect on 7SK RNA levels. The results show that reducing the expression of BCDIN3/MePCE by about 70% in vivo caused a decrease in the steady-state level of cellular 7SK as determined by RNA blot (Figure 7). BCDIN3/MePCE knockdown had no effect on U2 and U6 snRNA levels (Figure 7). Together, our results indicate that BCDIN3/MePCE stabilizes the 7SK snRNA in human cells through the addition of a methyl-phosphate cap at its 5' end.

DISCUSSION

In this article, we report the development of a general procedure allowing (1) the purification of affinity-tagged proteins in native conditions, (2) the identification of purified proteins using sensitive mass spectrometry, and (3) the computational analysis of the data for minimizing the rate of both false positives and false negatives, and its use to define a high-confidence protein interaction data set for transcription and RNA processing factors. The results were used to build a high-density interaction network connecting components of the transcription and RNA processing machineries in human cells. Although the tran-

scription and RNA processing machineries have been studied for decades, our results define many additional proteins that interact with some of their key components. In particular, important transcription factors are associated with proteins that regulate the formation of other protein complexes. We believe that these proteins have not been identified in previous studies because they apparently do not regulate these machineries during active transcription, although this possibility remains to be tested. For example, the RPAPs-XAB1 are tightly associated both physically and functionally with RNAP II. Because they also interact with chaperones/scaffolding proteins and themselves contain motifs characteristic of chaperones/scaffolding proteins (see Figures 3A and 3C), our results suggest that we have identified and purified proteins involved in assembly of the protein complexes composing the cellular machinery for mRNA synthesis.

The composition of the 7SK snRNP, which was discovered more than two decades ago, has been elusive (Wassarman and Steitz, 1991; Gurney and Eliceiri, 1980). Association of the transcription elongation factor P-TEFb with a putative methyltransferase prompted us to use this tagged protein in reciprocal affinity-purification experiments. Our results confirmed that BCDIN3 associates with P-TEFb, its negative regulatory factors HEXIMs, and the RNA processing factor SART3. Our results also indicate that the 7SK and U6 snRNAs are part of the BCDIN3 affinity-purified complex. Indeed, previous studies have shown that HEXIM1 and HEXIM2 bind to the 7SK snRNA through their arginine-rich RNA-binding motifs (Byers et al., 2005; Yik et al., 2004, 2005). Of note, SART3 contains two RNA recognition motifs (RRMs) that bind directly to the U6 snRNA (Bell et al., 2002). Interestingly, it has been proposed that SART3 acts as an RNA chaperone. Indeed, SART3 interacts via its C-terminal domain with other components of the U6 snRNP, the LSM proteins, which assemble as heteroheptamers on the 3' end of the U6 snRNA (Achsel et al., 1999). In addition, it has also been shown that SART3 interacts through its TPR motifs with the U4/U6-specific factor PRPF3, suggesting that it provides the scaffolding interface for the assembly of the U4/U6 snRNP (Medenbach et al., 2004). It will be interesting to investigate whether SART3 participates in the formation of the 7SK snRNP.

In this study, we show that BCDIN3, a conserved protein that is part of a family of AdoMet-dependent methyltransferases, is a 7SK MePCE. To date, the methylated gamma-phosphate cap structure has been identified in three other RNAP III-transcribed small RNAs, namely U6, rodent B2, and plant U3 (Shumyatsky et al., 1990; Shimba et al., 1992; Liu et al., 1992). Our results indicate that siRNA-mediated silencing of MePCE expression reduces the steady-state level of cellular 7SK, suggesting that MePCE protects 7SK from exonucleolytic degradation. Interestingly, previous studies have shown that the activity of P-TEFb is highly modulated via alternative interactions with its positive (Brd4) and negative (HEXIMs and 7SK snRNA) regulators (Nguyen et al., 2001; Yang et al.,

2001, 2005; Yik et al., 2003; Michels et al., 2003, 2004). Although the role of 7SK capping by MePCE remains to be detailed, this structure might be involved in regulating the formation (e.g., assembly and/or transport) of transcriptionally active P-TEFb to be recruited to gene promoters. Further analyses are necessary to unravel the mechanism of action of the capped 7SK and the pathway controlling the association of capped 7SK with P-TEFb.

Despite decades of detailed biochemical and genetic analysis of the RNAP II basal transcription machinery, we identified many previously uncharacterized factors, some being involved in regulating the network of interactions formed by the transcription and RNA processing machineries. Expanding this network promises to reveal numerous additional interactions of general or regulatory importance and may be similarly revealing for other cellular functions.

EXPERIMENTAL PROCEDURES

Human Cell Lines Carrying Inducible TAP-Tagged Proteins

Selected human polypeptides were cloned into the mammalian expression vector pMZI (Zeghouf et al., 2004) carrying a TAP tag at its C terminus (Rigaut et al., 1999). Stable human embryonic kidney cell lines EcR-293 (derived from HEK293) carrying these constructs were produced as previously described (Jeronimo et al., 2004). The conditions for expression, affinity purification, and mass spectrometry identification of proteins are detailed in the **Supplemental Experimental Procedures**.

DNA Microarray Analysis of Tet-Promoter Mutants

DNA microarray analyses of tet-promoter mutants were performed as described (Mnaimneh et al., 2004). See the **Supplemental Experimental Procedures** for further details.

siRNA Knockdown and RNA Blot

BCDIN3 (ON-TARGETplus SMART pool) and control (siCONTROL Non-targeting pool) siRNAs (Dharmacon) were transfected in 50% confluent HEK293 cells using Lipofectamine 2000 (Invitrogen) at an siRNA final concentration of 100 nM. At various time intervals post-transfection, cells were lysed and BCDIN3 expression levels were monitored by western blotting (see the **Supplemental Experimental Procedures** for a description of the antibodies used).

RNA blot analyses of total RNAs from HeLa cells, HEK293 cells treated with siRNA, or BCDIN3-TAP eluate were carried out as previously described (Li et al., 2007) (see the **Supplemental Experimental Procedures** for further details).

Sequence Analysis

The proteins belonging to the Bin3 family (NCBI's conserved domain database; pfam06859) have the following accession numbers: *H. sapiens*, NP_062552; *M. musculus*, NP_659162; *D. melanogaster*, NP_724468; *C. elegans*, NP_496572; *A. thaliana*, NP_568752; and *S. pombe*, NP_596220.

Other Methods

See the **Supplemental Experimental Procedures** for methods relating to gel filtration chromatography, immunodepletion, RNA methyltransferase, dephosphorylation, and decapping assays.

Supplemental Data

Supplemental Data include Supplemental Discussion, Supplemental Experimental Procedures, Supplemental References, two figures,

and one table and can be found with this article online at <http://www.molecule.org/cgi/content/full/27/2/262/DC1/>.

ACKNOWLEDGMENTS

We are grateful to the members of our laboratory and Jacques Archambault for helpful discussions and comments. The SART3/Tip110 antibody was a kind gift from J.J. He. We thank Denis Faubert for mass spectrometry analysis and S. Clarke for helpful discussions. This work is supported by grants from the Canadian Institutes for Health Research (CIHR), Genome Canada, Génome Québec, Genome Prairie-Alberta, the Ontario Genomics Institute, the Fonds de la recherche en santé du Québec (FRSQ), and the Canadian Foundation for Innovation. G.G.P. holds a Canada Research Chair in Targeted Proteomics. B. Chabot holds a Canada Research Chair in Functional Genomics. C.J. and C.T. hold studentships from the CIHR and the FRSQ. G.C. holds a Charles H. Best fellowship.

Received: December 13, 2006

Revised: May 16, 2007

Accepted: June 22, 2007

Published: July 19, 2007

REFERENCES

- Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J.* **18**, 5789–5802.
- Aguilera, A. (2005). Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. *Curr. Opin. Cell Biol.* **17**, 242–250.
- Baillat, D., Hakimi, M.A., Naar, A.M., Shilatfard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276.
- Barboric, M., Kohoutek, J., Price, J.P., Blazek, D., Price, D.H., and Peterlin, B.M. (2005). Interplay between 7SK snRNA and oppositely charged regions in HEXIM1 direct the inhibition of P-TEFb. *EMBO J.* **24**, 4291–4303.
- Bell, M., Schreiner, S., Damianov, A., Reddy, R., and Bindereif, A. (2002). p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.* **21**, 2724–2735.
- Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.* **17**, 251–256.
- Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336.
- Blazek, D., Barboric, M., Kohoutek, J., Oven, I., and Peterlin, B.M. (2005). Oligomerization of HEXIM1 via 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. *Nucleic Acids Res.* **33**, 7000–7010.
- Byers, S.A., Price, J.P., Cooper, J.J., Li, Q., and Price, D.H. (2005). HEXIM2, a HEXIM1-related protein, regulates positive transcription elongation factor b through association with 7SK. *J. Biol. Chem.* **280**, 16360–16367.
- Cougot, N., van Dijk, E., Babajko, S., and Seraphin, B. (2004). "Cap-tabolism." *Trends Biochem. Sci.* **29**, 436–444.
- 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.
- Fischle, W., Wang, Y., and Allis, C.D. (2003). Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.* **15**, 172–183.
- Gstaiger, M., Luke, B., Hess, D., Oakeley, E.J., Wirbelauer, C., Blondel, M., Vigneron, M., Peter, M., and Krek, W. (2003). Control of

Proteomics Survey Reveals the 7SK Capping Enzyme

- nutrient-sensitive transcription programs by the unconventional prefoldin URI. *Science* 302, 1208–1212.
- Gupta, S., Busch, R.K., Singh, R., and Reddy, R. (1990). Characterization of U6 small nuclear RNA cap-specific antibodies. Identification of gamma-monomethyl-GTP cap structure in 7SK and several other human small RNAs. *J. Biol. Chem.* 265, 19137–19142.
- Gurney, T., Jr., and Eliceiri, G.L. (1980). Intracellular distribution of low molecular weight RNA species in HeLa cells. *J. Cell Biol.* 87, 398–403.
- Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* 11, 394–403.
- Hamm, J., Darzynkiewicz, E., Tahara, S.M., and Mattaj, I.W. (1990). The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell* 62, 569–577.
- Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* 62, 465–503.
- Hampsey, M., and Reinberg, D. (2003). Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* 113, 429–432.
- Hirose, Y., and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev.* 14, 1415–1429.
- Hu, X., Malik, S., Negroui, C.C., Hubbard, K., Velalar, C.N., Hampton, B., Grosu, D., Catalano, J., Roeder, R.G., and Gnat, A. (2006). A Mediator-responsive form of metazoan RNA polymerase II. *Proc. Natl. Acad. Sci. USA* 103, 9506–9511.
- Huber, A.H., Nelson, W.J., and Weis, W.I. (1997). Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* 90, 871–882.
- Ikura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* 102, 463–473.
- Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* 19, 523–534.
- Jeronimo, C., Langelier, M.F., Zeghouf, M., Cojocar, M., Bergeron, D., Baali, D., Forget, D., Mnaimneh, S., Davierwala, A.P., Pootoolal, J., et al. (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.
- Junca, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* 12, 5–14.
- Kagan, R.M., and Clarke, S. (1994). Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. *Arch. Biochem. Biophys.* 310, 417–427.
- Kanemaki, M., Kurokawa, Y., Matsu-ura, T., Makino, Y., Masani, A., Okazaki, K., Morishita, T., and Tamura, T.A. (1999). TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. *J. Biol. Chem.* 274, 22437–22444.
- Kiss, T. (2004). Biogenesis of small nuclear RNPs. *J. Cell Sci.* 117, 5949–5951.
- Krogan, N.J., Peng, W.T., Cagney, G., Robinson, M.D., Haw, R., Zhong, G., Guo, X., Zhang, X., Canadien, V., Richards, D.P., et al. (2004). High-definition macromolecular composition of yeast RNA-processing complexes. *Mol. Cell* 13, 225–239.
- Lembo, F., Pero, R., Angrisano, T., Vitiello, C., Iuliano, R., Bruni, C.B., and Chiarotti, L. (2003). MBDin, a novel MBD2-interacting protein, relieves MBD2 repression potential and reactivates transcription from methylated promoters. *Mol. Cell Biol.* 23, 1656–1665.
- Li, Q., Price, J.P., Byers, S.A., Cheng, D., Peng, J., and Price, D.H. (2005). Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. *J. Biol. Chem.* 280, 28819–28826.
- Li, Q., Cooper, J.J., Altwerger, G.H., Feldkamp, M.D., Shea, M.A., and Price, D.H. (2007). HEXIM1 is a promiscuous double-stranded RNA-binding protein and interacts with RNAs in addition to 7SK in cultured cells. *Nucleic Acids Res.* 35, 2503–2512.
- Liu, M.H., Busch, R.K., Buckley, B., and Reddy, R. (1992). Characterization of antibodies against methyl-pppN cap structure: plant U3 small nucleolar RNA is recognized by these antibodies. *Nucleic Acids Res.* 20, 4299–4304.
- Liu, Y., Li, J., Kim, B.O., Pace, B.S., and He, J.J. (2002). HIV-1 Tat protein-mediated transactivation of the HIV-1 long terminal repeat promoter is potentiated by a novel nuclear Tat-interacting protein of 110 kDa, Tip110. *J. Biol. Chem.* 277, 23854–23863.
- Liu, Y., Kim, B.O., Kao, C., Jung, C., Dalton, J.T., and He, J.J. (2004). Tip110, the human immunodeficiency virus type 1 (HIV-1) Tat-interacting protein of 110 kDa as a negative regulator of androgen receptor (AR) transcriptional activation. *J. Biol. Chem.* 279, 21766–21773.
- Lu, S.C. (2000). S-Adenosylmethionine. *Int. J. Biochem. Cell Biol.* 32, 391–395.
- Medenbach, J., Schreiner, S., Liu, S., Luhrmann, R., and Bindereif, A. (2004). Human U4/U6 snRNP recycling factor p110: mutational analysis reveals the function of the tetratricopeptide repeat domain in recycling. *Mol. Cell Biol.* 24, 7392–7401.
- Michels, A.A., Nguyen, V.T., Fraldi, A., Labas, V., Edwards, M., Bonnet, F., Lania, L., and Bensaude, O. (2003). MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol. Cell Biol.* 23, 4859–4869.
- Michels, A.A., Fraldi, A., Li, Q., Adamson, T.E., Bonnet, F., Nguyen, V.T., Sedore, S.C., Price, J.P., Price, D.H., Lania, L., and Bensaude, O. (2004). Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J.* 23, 2608–2619.
- Mnaimneh, S., Davierwala, A.P., Haynes, J., Moffat, J., Peng, W.T., Zhang, W., Yang, X., Pootoolal, J., Chua, G., Lopez, A., et al. (2004). Exploration of essential gene functions via titratable promoter alleles. *Cell* 118, 31–44.
- Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* 414, 322–325.
- Nitta, M., Saijo, M., Kodo, N., Matsuda, T., Nakatsu, Y., Tamai, H., and Tanaka, K. (2000). A novel cytoplasmic GTPase XAB1 interacts with DNA repair protein XPA. *Nucleic Acids Res.* 28, 4212–4218.
- Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* 23, 297–305.
- Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. *Cell* 108, 501–512.
- Reddy, R., Singh, R., and Shimba, S. (1992). Methylated cap structures in eukaryotic RNAs: structure, synthesis and functions. *Pharmacol. Ther.* 54, 249–267.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Serafini, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032.
- Shimba, S., Buckley, B., Reddy, R., Kiss, T., and Filipowicz, W. (1992). Cap structure of U3 small nucleolar RNA in animal and plant cells is different. gamma-Monomethyl phosphate cap structure in plant RNA. *J. Biol. Chem.* 267, 13772–13777.
- Shin, C., and Manley, J.L. (2004). Cell signalling and the control of pre-mRNA splicing. *Nat. Rev. Mol. Cell Biol.* 5, 727–738.
- Shinshi, H., Miwa, M., and Sugimura, T. (1976). Enzyme cleaving the 5'-terminal methylated blocked structure of messenger RNA. *FEBS Lett.* 65, 254–257.

- Shumyatsky, G.P., Tillib, S.V., and Kramerov, D.A. (1990). B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5' end. *Nucleic Acids Res.* **18**, 6347–6351.
- Smith, D.F. (2004). Tetratricopeptide repeat cochaperones in steroid receptor complexes. *Cell Stress Chaperones* **9**, 109–121.
- Smith, C.L., and Peterson, C.L. (2005). ATP-dependent chromatin remodeling. *Curr. Top. Dev. Biol.* **65**, 115–148.
- Smith, T.F., Gaitatzes, C., Saxena, K., and Neer, E.J. (1999). The WD repeat: a common architecture for diverse functions. *Trends Biochem. Sci.* **24**, 181–185.
- Ueland, P.M. (1982). Pharmacological and biochemical aspects of S-adenosylhomocysteine and S-adenosylhomocysteine hydrolase. *Pharmacol. Rev.* **34**, 223–253.
- Vainberg, I.E., Lewis, S.A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H.L., and Cowan, N.J. (1998). Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* **93**, 863–873.
- Wassarman, D.A., and Steitz, J.A. (1991). Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol. Cell. Biol.* **11**, 3432–3445.
- Yang, D., Nakao, M., Shichijo, S., Sasatomi, T., Takasu, H., Matsumoto, H., Mori, K., Hayashi, A., Yamana, H., Shirouzu, K., and Itoh, K. (1999). Identification of a gene coding for a protein possessing shared tumor epitopes capable of inducing HLA-A24-restricted cytotoxic T lymphocytes in cancer patients. *Cancer Res.* **59**, 4056–4063.
- Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **414**, 317–322.
- Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell* **19**, 535–545.
- Yik, J.H., Chen, R., Nishimura, R., Jennings, J.L., Link, A.J., and Zhou, Q. (2003). Inhibition of P-TEFb (CDK9/cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol. Cell* **12**, 971–982.
- Yik, J.H., Chen, R., Pezda, A.C., Samford, C.S., and Zhou, Q. (2004). A human immunodeficiency virus type 1 Tat-like arginine-rich RNA-binding domain is essential for HEXIM1 to inhibit RNA polymerase II transcription through 7SK snRNA-mediated inactivation of P-TEFb. *Mol. Cell. Biol.* **24**, 5094–5105.
- Yik, J.H., Chen, R., Pezda, A.C., and Zhou, Q. (2005). Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive positive transcription elongation factor b complexes for control of transcription. *J. Biol. Chem.* **280**, 16368–16376.
- Zeghouf, M., Li, J., Butland, G., Borkowska, A., Canadien, V., Richards, D., Beattie, B., Emili, A., and Greenblatt, J.F. (2004). Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes. *J. Proteome Res.* **3**, 463–468.
- Zhao, H., Wang, Q., Zhang, H., Liu, Q., Du, X., Richter, M., and Greene, M.I. (2005). UXT is a novel centrosomal protein essential for cell viability. *Mol. Biol. Cell* **16**, 5857–5865.
- Zhou, Q., and Yik, J.H. (2006). The yin and yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. *Microbiol. Mol. Biol. Rev.* **70**, 646–659.
- Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182–185.
- Zhu, W., and Hanes, S.D. (2000). Identification of drosophila bicoid-interacting proteins using a custom two-hybrid selection. *Gene* **245**, 329–339.
- Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M.B., and Price, D.H. (1997). Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev.* **11**, 2622–2632.

Molecular Cell, Volume 27

Supplemental Data

Systematic Analysis of the Protein Interaction

Network for the Human Transcription Machinery

Reveals the Identity of the 7SK Capping Enzyme

Célia Jeronimo, Diane Forget, Annie Bouchard, Qintong Li, Gordon Chua, Christian Poitras, Cynthia Thérien, Dominique Bergeron, Sylvie Bourassa, Jack Greenblatt, Benoit Chabot, Guy G. Poirier, Timothy R. Hughes, Mathieu Blanchette, David H. Price, and Benoit Coulombe

Supplemental Discussion

Computational Validation of Protein Interactions

Our purification procedure was specifically designed to preserve the integrity of the purified complexes as they exist in live human cells. Affinity purification of tagged proteins theoretically allows the isolation of all protein complexes containing the tagged polypeptide. This method does not, however, allow for the direct determination of the abundance of the purified complexes. In addition, the high sensitivity of mass spectrometry requires the development of methods that discriminate between specific and spurious interactions (Patil and Nakamura, 2005; Krogan et al., 2006; Gavin et al., 2006). This was accomplished through the development of an algorithm that selects high confidence interactions by assigning Interaction Reliability (IR) scores to each protein interaction (a schematic representation of our computational procedure is provided in Supplemental Figure S2). The sensitivity and specificity of the algorithm was evaluated using literature-based classification of protein interactions. We selected as high confidence interactions those for which the IR score exceeded a threshold (IR score

above 0.6729) predicted to miss as false-negatives only 17% of a set of literature-supported interactions while incorrectly retaining only 17% of a set of interactions without literature support as false-positives. The selected interactions were used to build a protein interaction map (see Figure 2 of the article).

Supplemental Experimental Procedures

Interaction Reliability Score

Let I_0 be the set of 2008 putative interactions detected by mass spectrometry and for which the Mascot (MS) score in the induced eluate was at least five times larger than in the equivalent non-induced eluate (Supplemental Table S1). Each interaction in I_0 was assigned an Interaction Reliability (IR) score, reflecting the likelihood that the interaction is correct, calculated as follows. Let $m(i,j)$ be the MS score for interaction between proteins P_i and P_j . Let $t(i,j)$ be the number of triangles in which interaction (P_i, P_j) is involved: $t(i,j) = |\{k: (i,k) \in I_0 \text{ and } (k,j) \in I_0\}|$. Interactions involved in a number of triangles are more likely to be correct than those involved in none. A set $P \subset I_0$ of 135 interactions strongly supported by the literature was extracted and used as representatives of true interactions (Supplemental Table S1). A set $N \subset I_0$ of 53 interactions that were judged to have low biological support were used as representatives of likely false-positives (Supplemental Table S1). The posterior probabilities $\Pr[m(i,j) | (i,j) \in P]$ and $\Pr[m(i,j) | (i,j) \in N]$, were then empirically estimated by fitting a gamma distribution to each subset. Similarly, posterior probabilities $\Pr[t(i,j) | (i,j) \in P]$ and $\Pr[t(i,j) | (i,j) \in N]$ were estimated using empirical frequencies. Finally, the reliability score for the predicted interaction (i,j) is given by the Naïve Bayes classifier:

$$\begin{aligned} \text{Reliability}(i,j) &= \Pr[(i,j) \in P \mid m(i,j), t(i,j)] \\ &= \frac{\Pr[m(i,j) \mid (i,j) \in P] * \Pr[t(i,j) \mid (i,j) \in P] * \Pr[(i,j) \in P]}{\sum_{x \in \{P,N\}} \Pr[m(i,j) \mid (i,j) \in x] * \Pr[t(i,j) \mid (i,j) \in x] * \Pr[(i,j) \in x]} \end{aligned}$$

where we set prior probabilities $\Pr[(i,j) \in P] = \Pr[(i,j) \in N] = 0.5$.

Expression of TAP-Tagged Proteins and Affinity Purification of Protein Complexes

Induction for 24-72 hours with 3-6 μM ponasterone A (Invitrogen) was used to express the TAP-tagged proteins at near physiological levels. Whole cell extracts prepared from induced and non-induced stable EcR-293 cell lines were subjected to purification by the TAP procedure as previously described (Jeronimo et al., 2004). Nuclease-treated (RNase A and/or DNase) whole cell extracts were used in some experiments to ensure that the purified interaction partners are associated with the tagged complex through protein-protein interactions (Robert et al., 2002).

Identification of Proteins by Mass Spectrometry

The TAP eluates were run on SDS gels, stained with silver or Sypro Ruby (Bio-Rad) and gel slices were excised and digested with trypsin as previously described (Jeronimo et al., 2004). The resulting tryptic peptides were purified and identified by tandem mass spectrometry (LC-MS/MS) with microcapillary reversed-phase high-pressure liquid chromatography coupled to an LCQ DecaXP (ThermoFinnigan), LTQ or LTQ-Orbitrap (ThermoElectron) quadrupole ion trap mass spectrometer with a nanospray interface.

Gel Filtration Chromatography

Affinity purified protein complexes were concentrated by dialysis in buffer F containing 10 mM HEPES pH 7.9, 100 mM NaCl, 0.1 mM EDTA, 5% glycerol and 0.5

mM DTT. An aliquot (50 μ L) of the concentrated eluate was fractionated on a Superose 6 PC 3.2/30 column (2.4 ml) previously equilibrated in buffer F using the ÄKTA FPLC system (GE Healthcare). The column was run in buffer F at a flow rate of 0.04 ml/minutes and 50- μ L fractions were collected. Aliquots of each five fractions were pooled, concentrated and analyzed by Western blot.

Antibodies

Primary antibodies anti-RPAP1, anti-FLJ21908 and anti-XAB1 were obtained from the CIM Antibody Core at Arizona State University (Tempe, Arizona). The rabbit anti-BCDIN3 antibody was generated using the BCDIN3 peptide VPPHQEAASGELRGGTERGPGRC and obtained from Chemicon (Millipore). The TAP-specific anti-Calmodulin Binding Peptide (CBP) antibody (clone C16T) was purchased from Upstate. Other primary antibodies were anti-SART3/Tip110 (gift from J.J. He, Indiana University School of Medicine, Indianapolis, Indiana) and anti-CDK9 (C-20, Santa Cruz). Anti-Cyclin T1 and anti-HEXIM1 antibodies have been previously characterized (Michels et al., 2003). The horseradish peroxidase-conjugated secondary antibodies were purchased from GE Healthcare.

DNA Microarray Analysis of Tet-Promoter Mutants

Mutant and isogenic wild type cultures were grown in parallel in SC medium with 10 μ g/ml doxycycline for a total of 22 hours. Total RNA was prepared by hot acidic phenol:chloroform extraction followed by ethanol precipitation. Microarrays were composed of 70-mer oligonucleotides, each specific to a different yeast gene (Operon Technologies), which were spotted onto poly-L-lysine-coated slides. Each array was normalized using grid-by-grid lowess smoothing. All measurements were taken in fluor-

reversed pairs (e.g., each time a mutant was analyzed, it was hybridized to two arrays, combined by averaging the normalized $\log(\text{ratio})$).

RNA Blot

Total RNAs from HeLa cells, HEK 293 cells treated with siRNA or BCDIN3-TAP eluate were extracted using mirVana miRNA isolation kit (Ambion). From an aliquot of the BCDIN3 eluate containing about 10 μg of protein, 3 μg of RNA was recovered. Probes for 7SK, U6 and U2 were 5'-end labeled using T4 PNK (NEB), following manufacturer's instructions. RNAs were separated on a 6% denaturing gel, and transferred to a Nytran N membrane (Whatman). The membrane was blocked with ULTRAhyb (Ambion), and probed with labeled 7SK, U6, or U2 oligos, following the manufacturer's instruction. RNA probes used for 7SK, U6 and U2 were as follows (numbers in parenthesis indicate the regions to which the probes hybridize):

7SK (121-160): GGGGAUGGUCGUCCUCUUCGACCGAGCGCGCAGCUUCGGG

U6 (60-99): GGAACGCUUCACGAAUUUGCGUGUCAUCCUUGCGCAGGGG

U2 (97-136): CGGAGCAAGCUCCUAUCCAUCUCCCUGCUCCAAAAAUCC

Immunodepletion

Protein A-sepharose beads (GE Healthcare) were saturated with BSA (Roche) and incubated with preimmune or anti-BCDIN3 serum for 12 hours at 4°C. The antibody-bound beads were then divided into three 50 μl -aliquots and each aliquot was used for one round of depletion. A total of 100 μl of HEK 293 whole cell extracts containing approximately 0.5 mg of protein was added to the first aliquot of antibody-bound beads and rocked for 2 hours at 4°C. After a brief spin, the supernatant was transferred to the next aliquot of beads and the process repeated. After three rounds of depletion, the

supernatants were analyzed by Western blotting and assayed in the *in vitro* methyltransferase assay.

Methyltransferase Assay

The RNA methyltransferase assay was performed in 20 μ l of total reaction mixture, containing 20 mM Tris-HCl (pH 8.0), 0.5 mM DTT, 2 mM EDTA, 50 mM KCl, 5% Glycerol, 100 units of RNAGuard RNase inhibitor (GE Healthcare), 5 μ Ci of 3 H-AdoMet (TRK865, GE Healthcare) (63.0 Ci/mmol), 250 ng of purified BCDIN3-His and 200 ng of recombinant 7SK snRNA prepared as described previously (Li et al., 2005). When using HEK 293 whole cell extracts as a source of enzyme, the final volume of the reaction was increased to 50 μ l. When mentioned, 100 μ M of S-adenosyl-L-homocysteine (AdoHcy) (Sigma) and 0.5 mg/ml of RNase A (USB) were used. After a 1-hour incubation at 30°C, the reaction was terminated by adding 200 μ l of a stop solution (0.1 M NaOAc (pH 5.2), 0.5% SDS, 2 mM EDTA, 100 μ g/ml tRNA). The reaction was extracted with phenol/chloroform and the RNA was ethanol precipitated and analyzed in a 6% urea-polyacrylamide gel. The gel was fixed (45% methanol, 10% acetic acid), treated with Amplify (NAMP100, GE Healthcare) and exposed to Hyperfilm-MP film (GE Healthcare) at -80°C.

Dephosphorylation and Decapping Assay

Recombinant 7SK snRNA was incubated for 1 hour at 37°C with Calf Intestinal Alkaline Phosphatase (CIAP) (Invitrogen), following manufacturer's instructions and supplemented by 100 units of RNAGuard RNase inhibitor (GE Healthcare). The volume of the reaction was brought to 200 μ l with nuclease-free H₂O containing 20 μ g of

glycogen (Roche), phenol:chloroform extracted and the RNA ethanol precipitated, air-dried and resuspended in H₂O.

Five pmoles of *in vitro* capped 7SK snRNA was digested with Tobacco Acid (TA) pyrophosphate (2.5, 7.5 and 15 units) at 37°C for 1 hour according to the manufacturer's instructions (Epicentre Biotechnologies), supplemented by 100 units of RNAGuard RNase inhibitor (GE Healthcare). The reaction was stopped by adding 200 µl of a stop solution, phenol/chloroform extracted and ethanol precipitated as mentioned above. The RNA was analyzed in a 6% urea-polyacrylamide gel and ethidium bromide stained. The gel was then fixed (45% methanol, 10% acetic acid), treated with Amplify (NAMP100, GE Healthcare) and exposed to Hyperfilm-MP film (GE Healthcare) at -80°C.

Supplemental References

Gavin,A.C., Aloy,P., Grandi,P., Krause,R., Boesche,M., Marzioch,M., Rau,C., Jensen,L.J., Bastuck,S., Dumpelfeld,B., Edlmann,A., Heurtier,M.A., Hoffman,V., Hoefert,C., Klein,K., Hudak,M., Michon,A.M., Schelder,M., Schirle,M., Remor,M., Rudi,T., Hooper,S., Bauer,A., Bouwmeester,T., Casari,G., Drewes,G., Neubauer,G., Rick,J.M., Kuster,B., Bork,P., Russell,R.B., and Superti-Furga,G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631-636.

Jeronimo,C., Langelier,M.F., Zeghouf,M., Cojocar,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.

Krogan,N.J., Cagney,G., Yu,H., Zhong,G., Guo,X., Ignatchenko,A., Li,J., Pu,S., Datta,N., Tikuisis,A.P., Punna,T., Peregrin-Alvarez,J.M., Shales,M., Zhang,X., Davey,M., Robinson,M.D., Paccanaro,A., Bray,J.E., Sheung,A., Beattie,B., Richards,D.P., Canadien,V., Lalev,A., Mena,F., Wong,P., Starostine,A., Canete,M.M., Vlasblom,J., Wu,S., Orsi,C., Collins,S.R., Chandran,S., Haw,R., Rilstone,J.J., Gandi,K., Thompson,N.J., Musso,G., St Onge,P., Ghanny,S., Lam,M.H., Butland,G., Altaf-Ul,A.M., Kanaya,S., Shilatifard,A., O'Shea,E., Weissman,J.S., Ingles,C.J., Hughes,T.R., Parkinson,J., Gerstein,M., Wodak,S.J., Emili,A., and Greenblatt,J.F. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637-643.

Li,Q., Price,J.P., Byers,S.A., Cheng,D., Peng,J., and Price,D.H. (2005). Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. *J. Biol Chem.* **280**, 28819-28826.

Michels,A.A., Nguyen,V.T., Fraldi,A., Labas,V., Edwards,M., Bonnet,F., Lania,L., and Bensaude,O. (2003). MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol Cell Biol* **23**, 4859-4869.

Patil,A. and Nakamura,H. (2005). Filtering high-throughput protein-protein interaction data using a combination of genomic features. *BMC. Bioinformatics.* **6**, 100.

Robert,F., Blanchette,M., Maes,O., Chabot,B., and Coulombe,B. (2002). A human RNA polymerase II-containing complex associated with factors necessary for spliceosome assembly. *J. Biol. Chem.* **277**, 9302-9306.

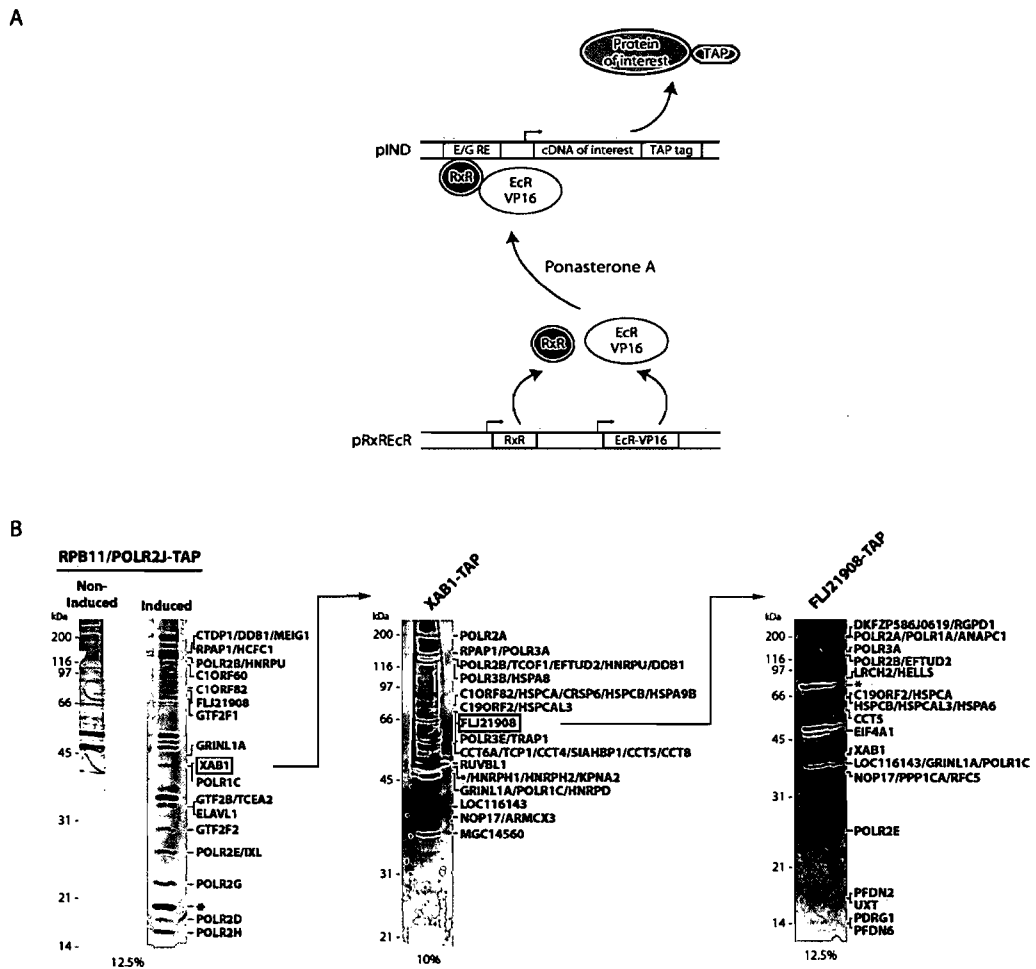


Figure S1. A Method for the Purification of Protein Complexes in Human Cells

(A) Schematic representation of the ecdysone-inducible system used to express, at physiological levels, the tagged protein upon addition of the inducer Ponasterone A. Expression was maintained near physiological levels through the use of the ecdysone-inducible system, which allows, by varying the dose of the inducer and the time of induction, calibration of the amount of the tagged protein made within the cells, as determined by Western blotting ((Jeronimo et al., 2004) and data not shown).

(B) Examples of SDS gels showing the composition of affinity purified protein complexes in reciprocal tagging experiments. Most bands that are not labeled contain proteins that have been either identified in a control eluate prepared from a non-induced cell extract or represent interactions that had IR scores below our selected threshold. The tagged protein is indicated by an asterisk. The proteins targeted in our reciprocal tagging experiments are boxed.

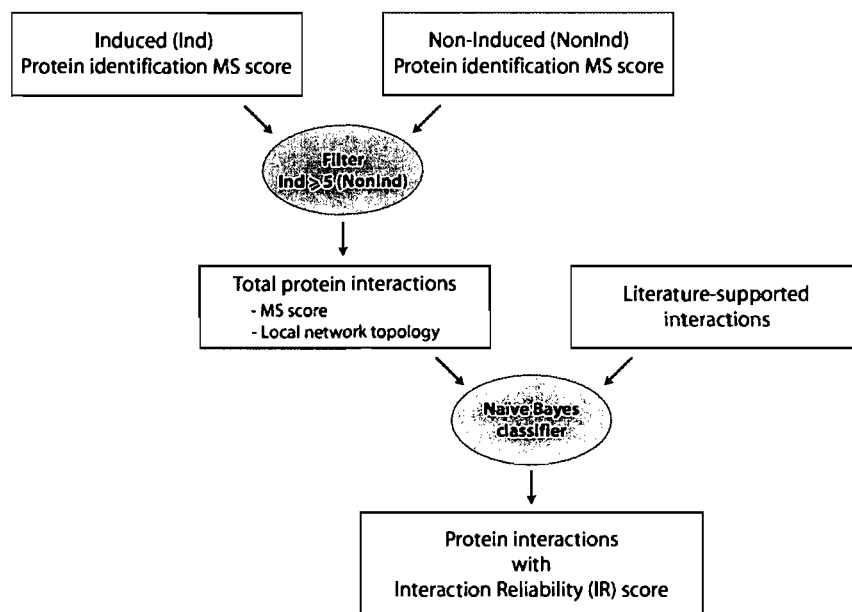


Figure S2. Computational Validation of Protein Interactions

For each purification, the non-specific interactions were filtered out by selecting proteins with a MS score in the induced (Ind) eluate at least 5 fold higher than in the equivalent non-induced (NonInd) eluate ($\text{Ind} \geq 5 (\text{NonInd})$), thereby eliminating proteins that bind non-specifically to our columns and very abundant cellular proteins that may remain as contaminants after the affinity purification steps. Each putative protein interaction selected at this stage was then assigned an IR score based on two criteria: (i) the MS score and (ii) the number of triangles, as defined by the local topology of the network (see Supplemental Experimental Procedures and Supplemental Table S1). The IR score was computed using a Naïve Bayes classifier, trained on a set of 135 positive interactions validated from the literature and a set of 53 negative interactions that were judged likely to be false-positives (based on the literature, but independently of the results of our experiments) (Supplemental Table S1).

CHAPITRE 4

LOCALIZATION GÉNOMIQUE DE LA MACHINERIE GÉNÉRALE DE L'ARN POLYMÉRASE II : ÉVIDENCE D'UN RÔLE POUR TFIIF ET RPB7 AUX STADES D'INITIATION ET D'ÉLONGATION DE LA TRANSCRIPTION

4.1 Préambule

Ma contribution à cet article est la suivante. J'ai effectué le clonage et l'établissement des lignées stables pour les facteurs généraux de transcription TFIIA, TFIIB, TFIIE et TFIIH portant une étiquette TAP et utilisés dans les expériences de ChIP. J'ai purifiée tous ces facteurs ainsi que TFIIF, TBPL1 et l'ARN Pol II par double chromatographie (figure 1A et Suppl. figure 1) et j'ai confirmé par des analyses de Western blot les composants identifiés dans le complexe TFIIH-TAP (figure 1A). J'ai montré que l'ARN Pol II purifiée en association avec le facteur TFIIS/TCEA1 était hypophosphorylée, la forme de l'enzyme recrutée aux promoteurs géniques (figure 1B). J'ai effectué l'alignement de séquences entre les protéines TCEA1 et TCEA2, montrant que la région C-terminale, connue pour être nécessaire et suffisante pour leur activité *in vitro*, est hautement conservée (figure 1C). Finalement, j'ai montré à la figure 1D que les complexes protéiques purifiés par la méthode TAP étaient actifs dans l'essai de transcription *in vitro*. J'ai participé à la rédaction de l'article avec Marilena Cojocaru et le Dr Benoit Coulombe.

Marilena Cojocaru a réalisée les expériences de localisation de l'ARN Pol II et des facteurs généraux de transcription le long des gènes étudiés par ChIP (figures 2, 3 et 4). La figure S1 a été effectuée en collaboration avec Diane Forget (TAF10, TAF13 et TRFP) et Annie Bouchard (TCEA1). Les analyses de MS ont été réalisées par les groupes

des Drs Jack Greenblatt et Guy G. Poirier (figure S1). Dominique Bergeron a développé les outils bioinformatiques pour l'analyse des interactions protéine-protéine présentées à la figure S1. Pierre Côte a contribué à la sélection des gènes étudiés.

4.2 Article

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



Genomic location of the human RNA polymerase II general machinery: evidence for a role of TFIIF and Rpb7 at both early and late stages of transcription

Marilena COJOCARU*, Célia JERONIMO*, Diane FORGET*, Annie BOUCHARD*, Dominique BERGERON*, Pierre CÔTE*, Guy G. POIRIER†, Jack GREENBLATT‡ and Benoit COULOMBE*¹

*Institut de Recherches Cliniques de Montréal, 110 avenue des Pins Ouest, Montréal, QC, Canada H2W 1R7, †Centre Hospitalier Universitaire de QC, Université Laval, Québec, QC, Canada, and ‡Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, Canada

The functions ascribed to the mammalian GTFs (general transcription factors) during the various stages of the RNAPII (RNA polymerase II) transcription reaction are based largely on *in vitro* studies. To gain insight as to the functions of the GTFs in living cells, we have analysed the genomic location of several human GTF and RNAPII subunits carrying a TAP (tandem-affinity purification) tag. ChIP (chromatin immunoprecipitation) experiments using anti-tag beads (TAP-ChIP) allowed the systematic localization of the tagged factors. Enrichment of regions located close to the TIS (transcriptional initiation site) versus further downstream TRs (transcribed regions) of nine human genes, selected for the minimal divergence of their alternative TIS, were analysed by QPCR (quantitative PCR). We show that, in contrast with reports using the yeast system, human TFIIF (transcription factor IIF) associates both with regions proximal to the TIS and with further downstream TRs, indicating

an *in vivo* function in elongation for this GTF. Unexpectedly, we found that the Rpb7 subunit of RNAPII, known to be required only for the initiation phase of transcription, remains associated with the polymerase during early elongation. Moreover, ChIP experiments conducted under stress conditions suggest that Rpb7 is involved in the stabilization of transcribing polymerase molecules, from initiation to late elongation stages. Together, our results provide for the first time a general picture of GTF function during the RNAPII transcription reaction in live mammalian cells and show that TFIIF and Rpb7 are involved in both early and late transcriptional stages.

Key words: chromatin immunoprecipitation, general transcription factor, genomic location, RNA polymerase II, tandem-affinity purification, transcription factor IIF (TFIIF).

INTRODUCTION

Since their first purification two decades ago, the set of mammalian factors required for transcription by RNAPII (RNA polymerase II) have drawn a great deal of attention. First believed to be required for the transcription of all class II genes, it is now generally accepted that the so-called GTFs (general transcription factors) TFI (transcription factor I) A, TFIIB, TFIID [also called TBP (TATA-box-binding protein)], TFIIE, TFIIF, TFIIH and TFIIIS/TCEA1 [transcription elongation factor A (SII) 1] form the core of the machinery that participates in the transcription of a large fraction of class II genes (reviewed in [1]). Using this wide definition, the TFIID–STAGA (SPT3–TAF9–GCN5 acetyltransferase)–TFTC [TATA-box-binding protein-free TAF (TBP-associated transcription factor)-containing]–PCAF [p300/CREB (cAMP-response-element-binding protein)-binding protein-associated factor] and Mediator complexes can also be classified as GTFs because they also participate in the transcription of a large proportion of mammalian genes [2,3]. However, studies in yeast have revealed that some GTFs are not

required for the transcription of a significant number of genes [4,5]. Similar studies in mammalian cells are still lacking.

Many reports have focused on defining the specific function of the GTFs during transcription. Biochemical analyses have revealed that the transcription reaction involves a number of successive steps that lead to the formation of a pre-mRNA [1]. In the first step, RNAPII positions its catalytic centre near the TIS (transcriptional initiation site) of promoter DNA. This first step is preceded by the formation of a pre-initiation complex that contains, in addition to RNAPII, a number of general initiation factors including TFIID, TFIIB, TFIIF, TFIIE and TFIIH. Assembly of the pre-initiation complex requires specific binding of some general initiation factors to core promoter elements such as binding of TBP to the TATA box, TFIIB to the BRE (TFIIB-recognition element) and TAFs of TFIID to the Inr (initiator element) and DPE (downstream promoter element). Both TFIIA and the Mediator complex have been shown to stimulate transcriptional initiation in reconstituted systems. Formation of the pre-initiation complex is accompanied by topological changes including bending and wrapping of promoter

Abbreviations used: BCR, breakpoint cluster region; CDK7, cyclin-dependent kinase 7; ChIP, chromatin immunoprecipitation; CTD, C-terminal domain; ENO1, enolase 1; ERCC, excision repair cross-complementing rodent repair deficiency, complementation group; GO, gene ontology; HEK-293 cell, human embryonic kidney cell; EcR 293 cells, HEK-293 cells stably transfected with the vector pVgRXR, expressing upon induction the heterodimeric ecdysone receptor; GTF, general transcription factor; HSPA8, heat-shock 70 kDa protein 8; MS/MS, tandem MS; PCAF, p300/CREB (cAMP-response-element-binding protein)-binding protein-associated factor; QPCR, quantitative PCR; RAP, RNA polymerase II-associated protein; RNAPII, RNA polymerase II; SFRS2, splicing factor, arginine/serine-rich 2; SRB7, suppressor of RNA polymerase B 7; STAGA, SPT3–TAF9–GCN5–acetyltransferase; TAF, TBP-associated transcription factor; TAP, tandem-affinity purification; TBP, TATA-box-binding protein; TCEA1, transcription elongation factor A (SII) 1; TCR, transcription-coupled DNA repair; TFI, transcription factor I; TFTC, TATA-box-binding protein-free TAF-containing; TIS, transcriptional initiation site; TR, transcribed region; XPB, xeroderma pigmentosum group B; XPG, xeroderma pigmentosum group G (Cockayne syndrome).

¹ To whom correspondence should be addressed, at Laboratory of Gene Transcription and Proteomics, Institut de Recherches Cliniques de Montréal, 110 avenue des Pins Ouest, Montréal, QC, Canada H2W 1R7 (email [redacted]).

DNA around the protein core of the complex [6]. In the second step, promoter DNA is melted between nucleotides -9 and $+2$ in such a way that the template strand becomes accessible for NTP polymerization. Both TFIIF and TFIIE were shown to participate in promoter melting [7]. Two distinct single-stranded DNA helicases, XPB (xeroderma pigmentosum group B)/ERCC (excision repair cross-complementing rodent repair deficiency, complementation group) 3 and XPD (xeroderma pigmentosum group D)/ERCC2, have been identified as components of TFIIF. XPB/ERCC3 was attributed a major role in promoter DNA melting prior to initiation [8]. During the third step, RNAPII enters a cycle of abortive initiation events wherein the enzyme synthesizes many short 2–10 nt transcripts. TFIIF is responsible for the melting of the template DNA during promoter escape and for the phosphorylation of the CTD (C-terminal domain) of the Rpb1 subunit of RNAPII [9]. Crystallographic structures of RNAPII suggest that the dissociable heterodimer formed by Rpb4 and Rpb7 stabilizes the RNAPII complex in a clamp-closed conformation, which is believed to favour efficient initiation [10]. This view is consistent with the hypothesis that the Rpb4–Rpb7 heterodimer is dispensable for transcript elongation [11], as RNAPII adopts the same conformation when bound to DNA even in the absence of Rpb4–Rpb7. Formation of the mRNA is completed through transcript elongation and termination of the transcription reaction. While some reports have proposed a role for both TFIIE and TFIIF in transcript elongation [12,13], a large body of information has accumulated to support a role for TFIIF and TFIIS/TCEA1 after promoter clearance (reviewed in [14]). TFIIS/TCEA1 can rescue paused or arrested elongation complexes through stimulation of an intrinsic endoribonuclease activity that allows transcript cleavage by backtracked RNAPII [15]. TFIIF was shown to stimulate the rate of NTP addition by RNAPII and to modulate the activity of TFIIS/TCEA1 [16]. However, most of these studies have been performed using *in vitro* systems and the information on the roles of the various GTFs *in vivo* remains fragmentary.

To gain insight into the function of the GTFs in live mammalian cells and determine whether specific GTFs act at early versus late stages of the transcription reaction *in vivo*, we have affinity-tagged many components of the RNAPII transcription machinery, expressed the tagged polypeptides at physiological levels in human cells using an inducible system and analysed their genomic location by CHIP (chromatin immunoprecipitation) coupled with QPCR (quantitative PCR) detection of the enriched regions. The location of RNAPII, TFIIA, TFIIB, TFIIE, TFIIF, TFIID–STAGA–TFTC–PCAF, TFIIF, TFIIS/TCEA1 and the Mediator was analysed both in the region proximal to TIS and further downstream in TRs (transcribed regions) of the nine human genes that were selected according to (i) their active transcriptional status in HEK-293 cells (human embryonic kidney cells), (ii) the minimal divergence of their TIS and (iii) their belonging to various functional classes according to GO (gene ontology) terms. Our results shed light on the function of the GTFs in living cells by showing that although all the factors, including TFIIS/TCEA1, can associate with regions proximal to the TIS, only TFIIF, TFIIS/TCEA1, the Mediator and STAGA–TFTC–PCAF complexes and the heterodimer Rpb4–Rpb7 are detected with RNAPII in TRs situated further downstream along actively transcribed genes.

EXPERIMENTAL

Generation of cell lines expressing tagged polypeptides

TAP (tandem-affinity purification)-tagged GTF and RNAPII subunits were cloned and expressed in human EcR 293 cells

(HEK-293 cells stably transfected with the vector pVgRXX, expressing upon induction the heterodimeric ecdysone receptor) as previously described [17]. The following C-terminal TAP-tagged polypeptides were analysed: RNAPII (Rpb2, Rpb4, Rpb7 and Rpb11), TFIIF [RAP (RNAPII-associated protein) 30 and RAP74], TFIIB, TFIIA (the γ -subunit and the $\alpha\beta$ precursor), TFIIE (TFIIE34 and TFIIE56), TFIIF [XPB/ERCC3 and CDK7 (cyclin-dependent kinase 7)], TFIID–STAGA–TFTC–PCAF (TAF10 and TAF13), the Mediator [SRB7 (suppressor of RNA polymerase B 7) and TRFP (TBP-related factor proximal homologue)] and TFIIS/TCEA1. Near physiological expression levels were obtained by inducing the cells for 24–48 h with 3–6 μM Ponasterone A (Invitrogen). In the case of TFIIS/TCEA1, two conditions were used in parallel: the cells were either exposed to a total dose of 12.5 J/m² of UVC radiations and allowed to recover for 20 min at 37 °C or left unexposed, with a total of 48 h of culture in both cases. For the heat shock, the cells were incubated for 1 h at 42 °C in parallel with matched controls.

Affinity purification and identification of proteins by MS

Protein-affinity purification was performed as we have previously described [17] (detailed protocols are available on our website at <http://www.ircm.qc.ca/microsites/hupi/en/733.html>). The TAP eluates were run on SDS gels, stained with silver or SYPRO Ruby (Bio-Rad) and gel slices were excised and digested with trypsin as previously described [17]. The resulting tryptic peptides were purified and identified by either MALDI–TOF–MS (matrix-assisted laser-desorption ionization–time-of-flight MS) [18] or LC (liquid chromatography)–MS/MS (tandem MS) with microcapillary reversed-phase HPLC coupled with an LCQ DecaXP (Thermo Finnigan) or LTQ (Thermo Electron) quadrupole ion-trap mass spectrometer with a nanospray interface. Resulting peptide MS/MS spectra were interpreted using the MASCOT (Matrix Science) software and searched against proteins in either the NCBI (National Center for Biotechnology Information) non-redundant protein database or the Uniref protein database [19].

Western blots and antibodies

Proteins from TAP eluates were separated by SDS/PAGE and transferred on to a PVDF membrane and probed with the appropriate antibodies. The horseradish peroxidase-conjugated secondary antibodies were purchased from GE Healthcare. Signals were detected using the ECL[®] (enhanced chemiluminescence) system (GE Healthcare).

Transcription assay

In vitro transcription reactions were performed as previously described [17] using highly purified calf thymus RNAPII and recombinant TBP, TFIIB, TFIIE, TFIIF and TFIIF. RNAPII, TBP, TFIIE and TFIIF were independently omitted from the reconstituted system and replaced by the eluate of affinity-purified Rpb11–TAP, TFIIA $\alpha\beta$ –TAP, TFIIE56–TAP and XPB–TAP. Control reactions in which an eluate from non-induced cells is added were included in each case.

CHIP

TAP–CHIP experiments were performed as previously described [17]. Briefly, 10⁸ cells were cross-linked with formaldehyde at a final concentration of 1%. A cross-link time of 10 min at room temperature (21 °C) was generally allowed before adding glycine

to a final concentration of 125 mM. A reduced cross-link time of 1–2 min was necessary in the case of the cells expressing the TAP-tagged TFIIE56, TAF10, TAF13 and SRB7. Chromatin was sonicated to obtain fragments of an average length of 450 bp. After immunoprecipitation with IgG beads, washing and elution, the cross-links were reversed at 65 °C, and the DNA was purified. The enrichment of regions proximal to the TIS and further downstream TRs for the nine genes was evaluated by QPCR as previously described [17], using two control regions: an internal sequence of an expressed sequence tag located at chr17:49926700–49927999 and a sequence located in a 1.330 kbp gene-less region of the 13q21.33 band [20]. To ensure that the enrichment obtained was specific to the TAP-tagged factors, the immunoprecipitated DNA from the untagged EcR 293 parental cells was systematically analysed in parallel. These control experiments allowed us to estimate the background level, which ranged between 0.8- and 1.4-fold enrichment over control regions. In pilot experiments (results not shown), we found that treating the samples with proteinase K at a final concentration of 0.2 µg/µl during the cross-link reversal greatly decreased non-specific enrichment of TRs, which can result at this step, as reported previously [21].

The genes to be analysed were selected using the databases DBTSS (DataBase of Transcription Start Site; April 2003 version, http://dbtss.hgc.jp/samp_home.html) and MGC (Mammalian Gene Collection; March 2003 version, <http://mgc.nci.nih.gov/>). The selection criteria was the minimal divergence of their alternative TIS, empirically defined as the number of full-length cDNAs situated at a distance of less than 40 bp from the 5'-ends provided by RefSeq. Primers were designed by using the Primer3 algorithm (see Table 1 for the sequence of the primers and Supplementary Figure 2 at <http://www.BiochemJ.org/bj/409/bj4090139add.htm> for the genomic location of the PCR products, determined on the UCSC Genome Browser [22]). The fragments to be analysed have been chosen based on three criteria: (i) to include the TIS or a position situated at a distance of at least 1000 bp from the TIS; (ii) to rank between those with the best score generated by the primer design algorithm; and (iii) to be unique. All PCR-amplified fragments are unique as verified by BLAT on the April 2003 assembly version of the UCSC Genome Browser (<http://genome.ucsc.edu/>), by QPCR melting-curve analysis and by agarose-gel electrophoresis (results not shown).

In the present study, we consider that a region is occupied by a given factor when: (i) reproducible and statistically significant higher enrichment values are obtained for that region compared with control regions and (ii) low background signals are obtained for the same region in a parallel control ChIP experiments conducted with the parental cells in which the tagged factor is not expressed.

RESULTS AND DISCUSSION

TAP-tagged components of the RNAPII machinery are functional

The specific function of the various GTFs that assist mammalian RNAPII during the successive stages of the transcription reaction is largely based on *in vitro* studies (reviewed in [1]). Although these biochemical data have proved invaluable to understand transcription at a fine molecular level, the function of the various GTFs in living mammalian cells remains mainly unaddressed. We reasoned that a systematic analysis of the genomic location of many individual components of the general transcription machinery would help us to understand the roles of the GTFs *in vivo*. We took advantage of our expanding collection of cell

lines programmed to express, upon induction, physiological levels of polypeptides carrying a TAP tag, including four subunits of RNAPII, two subunits of each TFIIA, TFIIE, TFIIF, TFIID–STAGA–TFTC–PCAF, TFIIH and the Mediator as well as TFIIB and TFIIS/TCEA1.

To ensure that the tagged polypeptides expressed at physiological levels in human cells behave properly (e.g. the tag does not impair their activity), we performed two lines of experiments. First, we affinity-purified all the tagged proteins in native conditions, analysed the eluates using MS and defined their network of interactions (see Figure 1A for an example and see Supplementary Figure 1 at <http://www.BiochemJ.org/bj/409/bj4090139add.htm> for details). Consistent with our previously published results [17], RNAPII is linked to TFIIB, TFIIF, FCP1 (TFIIF-interacting CTD phosphatase subunit 1)/CTDPI (RNAPII CTD polypeptide A phosphatase subunit 1) and RPAP1 (RNAPII-associated protein 1) *in vivo*. Both RNAPII and TFIIF are connected to the Mediator complex, which itself is linked to TFIID–STAGA–TFTC–PCAF subunits. TFIID–STAGA–TFTC–PCAF associates with TFIIA and TRFs (TBP-related factors) *in vivo*. Tagged TFIIS/TCEA1 co-purifies with RNAPII subunits, as previously described [23]. Surprisingly, however, the RNAPII molecules that interact with TFIIS/TCEA1 appear to be mainly in the hypophosphorylated I₂ form (Figure 1B); purifications from whole-cell lysates prepared in the presence of phosphatase inhibitors indicated that this finding is not merely the consequence of CTD dephosphorylation during the preparation of the cell extracts (results not shown). Because RNAPII with a hypophosphorylated Rpb1 CTD represents the form of the enzyme that is recruited to promoter DNA prior to transcription initiation, our results support the notion that TFIIS/TCEA1 can associate with RNAPII at promoters *in vivo* (see below). Also surprising is the finding that tagged Rpb1 co-purifies with TFIIS.1/TCEA2 (see Supplementary Figure 1). Comparison of the amino acid sequence of TFIIS/TCEA1 with TFIIS.1/TCEA2 shows that the C-terminal region, the portion of the polypeptide known to be necessary and sufficient for activity *in vitro* [24], is highly conserved (Figure 1C). A previous study has shown that the N-terminal region of human TFIIS/TCEA1 interacts with RNAPII holoenzyme, whereas its C-terminal portion binds free RNAPII [25]. To our knowledge, this is the first time that the protein encoded by the TFIIS.1/TCEA2 gene is purified and shown to associate with RNAPII. Notably, both TFIIE and TFIIH are only weakly linked to each other and to other GTFs in our network (Supplementary Figure 1). Both CDK7 and XPB/ERCC3 co-purify with the DNA repair factor ERCC5/XPG [xeroderma pigmentosum group G (Cockayne syndrome)]; XPB/ERCC3 co-purifies with BCR (breakpoint cluster region) (Figure 1A), the polypeptide forming the chimaeric BCR-ABL (Abelson tyrosine kinase fusion protein) protein in the Philadelphia chromosome. ERCC5/XPG connects TFIIH to TFIID–STAGA–TFTC–PCAF (Supplementary Figure 1). Together, these results indicate that the tagged factors have the ability to associate with their endogenous interaction partners *in vivo* and reveal new aspects of the TFIIS–RNAPII interaction in live mammalian cells.

Secondly, the ability of the purified GTFs and RNAPII to support transcription was assessed. The GTFs and RNAPII affinity-purified using their tagged subunits were able to replace their highly purified counterparts in *in vitro* transcription assays (Figure 1D). Together, the results of the proteomic analysis and the biochemical assays support the notion that the tagged polypeptides associate with their cognate partners to form functionally active transcription factors in mammalian cells.

Table 1 Sequences of the primers used to analyse the location of various polypeptides

The position of the PCR products relative to the assigned TIS is indicated for each primer set. Fw, forward; Re, reverse; FTL, ferritin, light polypeptide; GNB2L1, guanine-nucleotide-binding protein (G-protein), β -polypeptide 2-like 1; GTF2F2, general transcription factor II F, polypeptide 2, 30 kDa; POLA2, polymerase (DNA-directed), α 2; POLR2E, polymerase (RNA) II (DNA-directed) polypeptide E; TIMM17A, translocase of inner mitochondrial membrane 17 homologue A (yeast).

Official gene symbol	Gene length (bp)	Primer	Sequence	PCR product position	
				Start	End
<i>POLA2</i>	35 000	TIS-Fw	5'-aaagcaaggggaagggtt-3'	-61	+123
		TIS-Re	5'-gaatggaggggcagaaat-3'		
		TR-Fw	5'-ggctctgcaaatatgaatc-3'	+988	+1199
		TR-Re	5'-atgactccttaccactgctctac-3'		
<i>GTF2F2</i>	163 566	TIS-Fw	5'-ttcttcagttatgctgacc-3'	-238	0
		TIS-Re	5'-ttacctgccagaacctg-3'		
		TR-Fw	5'-ctaagaggcttcttctgctg-3'	+978	+1186
		TR-Re	5'-actattctgggtatgacagg-3'		
<i>FTL</i>	1559	TIS-Fw	5'-gctgagactcctatgtct-3'	-157	+33
		TIS-Re	5'-acactgtgaaagcaagagac-3'		
		TR-Fw	5'-tatagaagccactggaagat-3'	+957	+1176
		TR-Re	5'-gtgaaatgaggctctgaa-3'		
<i>TIMM17A</i>	15 141	TIS-Fw	5'-aaccaatgcatatagacctt-3'	-80	+84
		TIS-Re	5'-gcaagagaatgcaagac-3'		
		TR-Fw	5'-caggaccttaatactactctgg-3'	+910	+1094
		TR-Re	5'-tatftacagtgctgagctctac-3'		
<i>GNB2L1</i>	6964	TIS-Fw	5'-gaatgtgctgttcagagt-3'	-200	+55
		TIS-Re	5'-atggcttagagaaactagca-3'		
		TR-Fw	5'-aaagggctctglatctg-3'	+998	+1165
		TR-Re	5'-tlaactagagatgctgctt-3'		
<i>SFRS2</i>	3215	TIS-Fw	5'-gccaatcagaagglttcal-3'	-86	+17
		TIS-Re	5'-gcacctgaglaacaactgg-3'		
		TR-Fw	5'-gaagglccaagltcaagt-3'	+1005	+1194
		TR-Re	5'-gattccagacattaccat-3'		
<i>POLR2E</i>	7170	TIS-Fw	5'-aactgccctctctgtaag-3'	-161	+90
		TIS-Re	5'-atgatggctctgctggtt-3'		
		TR-Fw	5'-gagalagggttctctgct-3'	+812	+1093
		TR-Re	5'-gaccagatgatcctgagag-3'		
<i>ENO1</i>	17 717	TIS-Fw	5'-ggtaggggaalgaglac-3'	-3	+151
		TIS-Re	5'-accgaggtagcaglaaag-3'		
		TR 1-Fw	5'-taggccaagaaggatgat-3'	+1140	+1289
		TR 1-Re	5'-gaatlagggacacglaaat-3'		
		TR 2-Fw	5'-gaacaagacctcaattgcta-3'	+3428	+3702
		TR 2-Re	5'-agcactggactaaalactgg-3'		
		TR 3-Fw	5'-gagaattgaaactcttc-3'	+5373	+5558
		TR 3-Re	5'-gtgactcacagatggtag-3'		
		TR 4-Fw	5'-gcacaaglttagagggtta-3'	+17106	+17321
		TR 4-Re	5'-cagctccttcaattct-3'		
<i>HSPA8</i>	4643	TIS-Fw	5'-ctgtgattgggtctgta-3'	-141	+88
		TIS-Re	5'-aaccttgagcagagggtt-3'		
		TR 1-Fw	5'-acctggaglccallgtagta-3'	+1175	+1463
		TR 1-Re	5'-aataccattatccctgcaa-3'		
		TR 2-Fw	5'-ctgaaatctggataacgtaggag-3'	+2026	+2324
		TR 2-Re	5'-ctctccctgtattctactggac-3'		
		TR 3-Fw	5'-gtaccattlgtagcaagltc-3'	+3265	+3549
		TR 3-Re	5'-tacagctctctgttctcaatg-3'		
		TR 4-Fw	5'-gtcaggaggagaagaggllatta-3'	+4152	+4439
		TR 4-Re	5'-atgtgggaacaalgctactac-3'		

TAP-ChIP as a systematic method for the localization of proteins along mammalian genomic DNA

The genomic location analysis of transcription factors by ChIP in a systematic fashion is known to be difficult mainly because of the requirement of antibodies able to efficiently immunoprecipitate proteins after a cross-linking step. TAP-ChIP experiments used IgG beads, which target the Protein A moiety of the TAP tag,

to pull down the TAP-tagged polypeptides covalently cross-linked to chromatin fragments. This method allowed us to systematically localize many components of the RNAPII general transcription machinery. Pilot experiments have revealed that the second affinity purification step (e.g. calmodulin beads) does not significantly improve the enrichment values of DNA fragments under our conditions (results not shown). After purification, enriched DNA sequences were analysed using QPCR with primer

		TIS								Downstream TR											
		FTL	GNB2L1	ENOI1	POLA2	SFRS2	TIMM17A	POLR2E	HSP8A	FTL	GNB2L1	ENOI1	POLA2	SFRS2	TIMM17A	POLR2E	HSP8A	GTF2F2			
RNAPII	Rpb11	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	8/8	8/9
TFIIA	TFIIA $\alpha\beta$	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	4/7	0/8
TFIIB		█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	7/8	0/9
TFIIE	E56	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	3/7	0/8
TFIIH	XPB	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	1/7	6/9
TFIID	STAGA TAF10	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	4/8	1/9
	TAF13	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	5/8	0/9
Mediator	SRB7	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	5/8	1/9
TFIIS	TFIISexp	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	3/7	3/8

Figure 2 Occupancy of regions proximal to the TIS and further downstream TRs of selected human genes by the GTF and RNAPII subunits obtained by ChIP

Fold enrichment of regions located close to the TIS (left-hand matrix) and further downstream TRs (right-hand matrix) over control regions are represented for each TAP-tagged factor, including subunits of RNAPII (Rpb11) and the GTFs TFIIA (TFIIA $\alpha\beta$), TFIIB, TFIIE (TFIIE56), TFIIH (XPB), TFIID/STAGA (TAF10 and TAF13), the Mediator (SRB7) and TFIIS/TCEA1. Colour code: black: > 5-fold enrichment; grey: significantly enriched with less than 5-fold enrichment values; white: not significantly enriched; X: not determined. The number of regions situated close to the TIS and further downstream TRs occupied by each factor is indicated as a ratio (observed/total tested). The results shown for TFIIS/TCEA1 were obtained by the analysis of cells cross-linked after an irradiation with a total dose of 12.5 J/m² of UVC (TFIISexp). Student's *t* tests were performed to determine the significance of the enrichment values obtained in four experiments, including at least two independent ChIP assays.

belonging to various GO classes, (ii) their activity in HEK-293 cells as verified by quantitative RT (reverse transcriptase)-PCR (results not shown) and (iii) our ability to design primer sets that are efficient in QPCR detection.

In the present study, the regions to be analysed by QPCR were selected in such a way that early versus late transcription events can be monitored. A distinction between early steps of the reaction (e.g. promoter binding, first-bond formation and promoter clearance) is difficult to address principally because of the limitation imposed by the resolution of the ChIP technique, which depends on the size of the sonicated fragments (approx. 450 bp in our case). In addition, it is conceivable that the efficiency of bringing down a promoter fragment would be much higher when a protein is associated with a tightly bound, engaged polymerase as compared with a loosely bound polymerase in a pre-initiation complex.

The results of our TAP-ChIP experiments (Figure 2) indicate that, as expected, TFIIA, TFIIB, TFIID, TFIIE, TFIIH and the Mediator assemble with RNAPII in regions proximal to the TIS of transcribed genes in live mammalian cells. The analysis of a second subunit confirmed the results obtained for these factors (see Supplementary Figure 3 at <http://www.BiochemJ.org/bj/409/bj4090139add.htm>). Moreover, when N-tagged factors or specific antibodies were used in validation experiments, we obtained similar location profiles as when using the C-tagged version (results not shown). The lack of detection of initiation factors in some regions close to the TIS may reflect a low rate of initiation events (e.g. under the detection limit of the ChIP procedure used here).

Our results also indicate that the transcription elongation factor TFIIS/TCEA1 is not stably recruited along the tested genes under normal growth conditions (results not shown). Notably,

TFIIS/TCEA1 was detected at a number of regions close to the TIS and in further downstream TRs after cell exposure to UVC radiations (Figure 2), suggesting a role for TFIIS/TCEA1 both at early and late stages of transcription. Cyclobutane pyrimidine dimers, the major DNA lesions induced by UVC radiations, are known to arrest the polymerase and to induce the recruitment of TFIIS/TCEA1 *in vitro* [26]. Whether the association of TFIIS/TCEA1 with genomic DNA is simply a consequence of an increased number of stalled elongating RNAPII molecules along damaged DNA or the signature of an active role for this factor in TCR (transcription-coupled DNA repair), as suggested by genetic studies in yeast [27], is not clear. A recent study in human cells showed that TFIIS is recruited to lesion-arrested RNAPII via the CSA and CSB (Cockayne syndrome type A and type B) proteins, suggesting a non-specific role of TFIIS in TCR [28]. Although the location of TFIIS/TCEA1 at regions close to the TIS does not necessarily imply that it has a function in transcription initiation, the finding that TFIIS/TCEA1 also associates with the hypophosphorylated form of RNAPII (see Figure 1B) fully supports the conclusion that it plays a role in very early stages of transcription, as recently proposed in yeast TFIIS/TCEA1 [29].

Two TAFs, TAF10 and TAF13, were detected along our tested genomic regions with very similar occupancy profiles of regions close to the TIS but distinct profiles in further downstream TRs (Figure 2). Both TAFs were detected at half of the regions close to the TIS analysed, indicating that the complex recruited is TFIID, TAF13 being a TFIID-specific subunit [30]. Of note, at some regions close to the TIS (three out of eight), neither TAF10 nor TAF13 was detected, which is in agreement with a previous report showing that these TAFs are required for the expression of only a fraction of the yeast genes [31]. Our results indicate that TAF10, a component of both TFIID and the STAGA-TFTC-PCAF complexes (reviewed in [30]), is recruited to TRs (see Figure 2 and results not shown) and suggest a role for this TAF during transcriptional elongation of specific genes. The observation that TAF13 is not recruited to these regions suggests that the TAF10-containing complex present at these locations is STAGA, TFTC or PCAF. While the function of SAGA (SPT-ADA-GCN5 acetyltransferase) at different stages of the transcription reaction is supported by a number of reports (see [32] for a recent review), its human homologue complexes STAGA-TFTC-PCAF are only poorly characterized. In support of our observation, the STAGA complex has been implicated in diverse transcription-coupled processes such as chromatin modification, pre-mRNA splicing and DNA repair [33].

The Mediator subunit SRB7 was detected at a fraction of the regions close to the TIS that we have tested (Figure 2), which is in agreement with data reported by Gromoller and Lehming [34] showing that Srb7p is essential for the activation of a subset of genes and with recent observations in yeast that Mediator subunits are recruited to the promoter of highly transcribed genes [35,36]. Unexpectedly, SRB7 was also localized in one TR, suggesting a role in late stages of transcription for this co-activator.

TFIIF is involved in post-initiation stages of transcription *in vivo*

As shown in Figure 3, our results indicate that the TFIIIF subunit RAP30 co-localizes with RNAPII in many regions located close to the TIS but also in further downstream TRs, demonstrating for the first time a role for this factor in transcriptional elongation *in vivo*. The analysis of the other TFIIIF subunit RAP74 confirmed this observation (see Supplementary Figure 3). This result contrasts with *in vivo* results in yeast [18], but fully supports previous *in vitro* results (reviewed in [1,37–39]) showing that

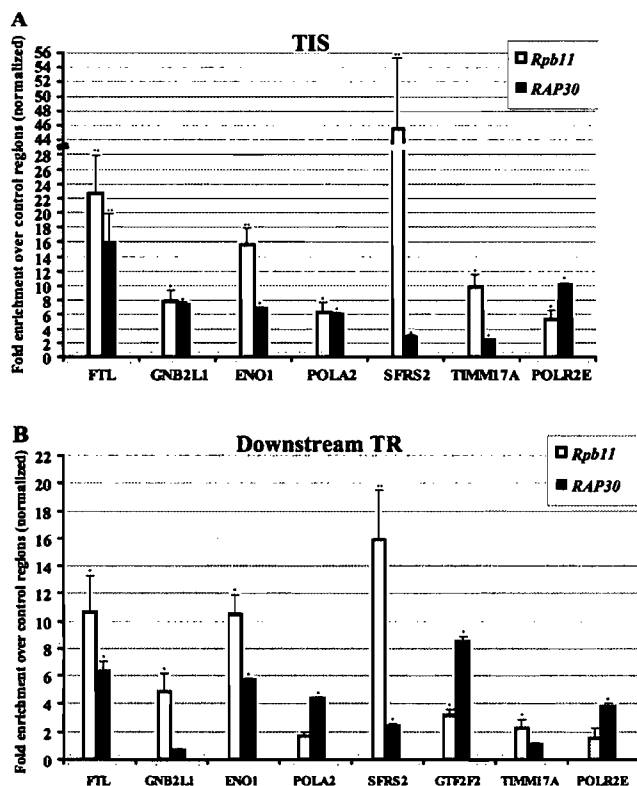


Figure 3 Occupancy of regions proximal to the TIS and further downstream TRs of selected human genes by the TFIIF subunit RAP30 and the RNAPII subunit Rpb11 as determined by ChIP

Fold enrichment of regions situated close to the TIS (A) and further downstream TRs (B) over control regions are represented for nine selected genes. Student's *t* tests were performed to determine the significance of the enrichment values obtained in four experiments, including at least two independent ChIP assays (**P* < 0.05, ***P* ≤ 0.001).

TFIIF acts at both the initiation and elongation stages of the transcription reaction and those of genetic screens, indicating a complex interdependence between TFIIF and the elongation factor TFIIS/TCEA1 [40].

It is interesting to note the gene-specific differences in the relative occupancy of TFIIF and RNAPII (Figure 3): while the RNAPII subunit Rpb11 showed higher occupancy values than the TFIIF subunit RAP30 at most of the regions located close to the TIS, the relative occupancy ratio is inverted in the case of the POLR2E gene where RAP30 showed a higher signal than Rpb11. Furthermore, RNAPII/TFIIF occupancy ratios significantly differ between regions close to the TIS and those situated further downstream in TRs as, for half of the downstream TRs occupied by RAP30, this polypeptide showed higher occupancy signals than Rpb11. This finding may reflect the presence of transcription complexes with different polypeptide composition at these positions, affecting the accessibility of the tag and/or the cross-linking efficiency. This is supported by a recent paper showing that RNAPII stalled close to the promoter *in vitro* cannot respond to TFIIF and can enter in productive elongation only after the P-TEFb (positive transcription elongation factor b) has functioned, at which point it begins to elongate and move rapidly away from the promoter [41]. Another possible explanation is that TFIIF is preferentially recruited on certain sequences in order to support elongation and to suppress pausing of the polymerase as previously suggested [16].

Rpb7 associates with both the regions proximal to the TIS and with further downstream TRs of active genes

Previous reports showed that Rpb7 is required for the initiation but not the elongation stage of transcription *in vitro* [42,43]. *In vivo* studies provided evidence suggesting that human Rpb7 can interact with transcriptional activators such as the oncogenic fusion protein EAD-Fli1 [44] and regulate transcription of genes such as VEGF (vascular endothelial growth factor) [45]. Nevertheless, no previous study addressed the direct implication of this RNAPII subunit at distinct stages of transcription *in vivo*.

ChIP experiments using a TAP-tagged version of Rpb7 show that it associates with regions close to the TIS and, surprisingly, with further downstream TRs as well (Figure 4). As Rpb4 and Rpb7 are known to form a heterodimer that can dissociate from the polymerase (reviewed in [46]), we generated and used cell lines expressing the TAP-tagged version of Rpb4 in order to confirm the Rpb7 results. Indeed, Rpb4 was also detected in regions close to the TIS and in further downstream TRs (see Supplementary Figure 3). Neither Rpb7 nor Rpb4 was shown to have a direct function in transcription elongation to date (see [46] for a recent review).

Because the Rpb4–Rpb7 dimer was previously suggested to have a role in the transcriptional response to stress, we next addressed the possible implication of Rpb7 in heat shock. As shown in Figure 4(C), and similarly to the core RNAPII subunit Rpb11, Rpb7 occupies all the regions close to the TIS after a heat shock of 1 h at 42°C. We noted a significant increase (40% or higher) in the enrichment of most regions close to the TIS (e.g. seven out of eight) by tagged Rpb7 when comparing heat shock with non-heat-shock conditions. Similar results were obtained with the tagged version of Rpb4, for which we observed a significant increase (> 50%) in the enrichment of regions close to the TIS after a heat shock (results not shown). Notably, even at regions close to the TIS showing a decreased Rpb11 occupancy after heat shock [Figure 4(C), for example SFRS2 (splicing factor, arginine/serine-rich 2)], a higher Rpb7 occupancy is observed. Conversely, in the case of POLR2E, a marked decrease in Rpb11 occupancy is accompanied by a lower Rpb7 occupancy after heat shock. This observation argues in favour of a function for the Rpb4–Rpb7 heterodimer in the stabilization of RNAPII during the early stages of transcription under stress conditions.

The same TRs were occupied by Rpb7 before and after heat shock, with similar or increased enrichments, suggesting that the requirement for this subunit during the elongation stage of transcription is maintained under stress conditions (Figure 4D). Of note, Rpb7 and Rpb4 were both detected in one TR which was not occupied by these polypeptides under normal conditions (Figures 4B and 4D and results not shown). In order to address in more detail the requirement of these subunits for transcriptional elongation under stress conditions, we designed primer sets along the entire TR of two constitutive heat-shock genes, known to present increased expression levels in response to a heat shock: ENO1 (enolase 1) and the HSPA8 (heat-shock 70 kDa protein 8). The change in the occupancy pattern observed for the RNAPII core subunit Rpb11 is in agreement with an increased expression of these genes in response to a heat shock, presenting higher occupancy values compared with normal conditions (Figure 4) and becoming detectable in downstream TRs which are not occupied under normal condition (Figures 4E and 4F). In contrast with the core subunit Rpb11, Rpb7 occupied only the TRs proximal to the TIS under normal conditions (up to approx. +1200) of both genes, while Rpb11 is detected much further downstream under the same conditions (Figures 4E and 4F). This

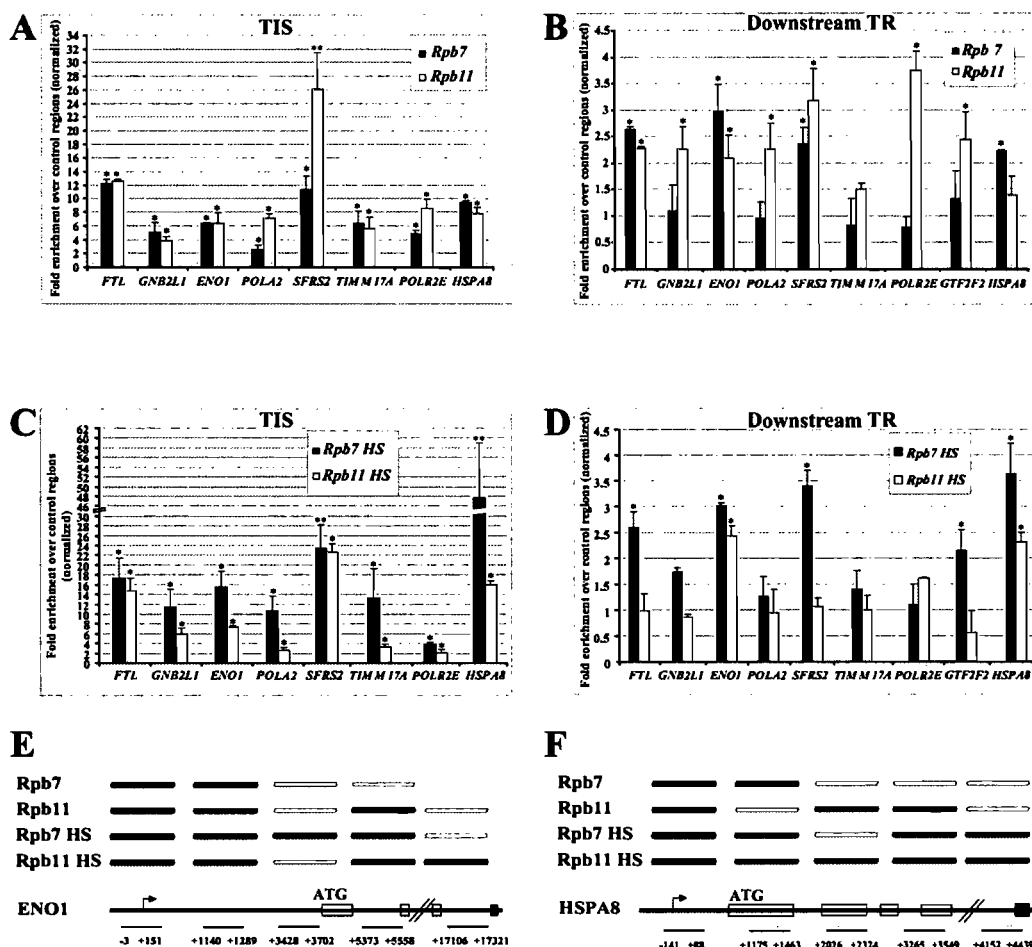


Figure 4 Occupancy of regions proximal to the TIS and further downstream TRs of selected human genes by Rpb7 as determined by ChIP

(A, B) Fold enrichment of regions situated close to the TIS (A) and further downstream TRs (B) over control regions are represented for Rpb7 and Rpb11 when cells are cultured under normal conditions. (C, D) Fold enrichment of regions situated close to the TIS (C) and further downstream TRs (D) after a heat shock (1 h at 42°C) (HS) are represented for Rpb7 and Rpb11. (E, F) The distribution along the genes ENO1 (E) and HSPA8 (F) is represented for Rpb7 and Rpb11 cultured under normal conditions or after a heat shock (HS). Colour code: black: > 10-fold enrichment; dark grey: between 5- and 10-fold enrichment; light grey: significantly enriched with less than 5-fold enrichment values; white: not significantly enriched. Student's *t* tests were performed to determine the significance of the enrichment values obtained in four experiments, including at least two independent ChIP assays (**P* < 0.05, ***P* < 0.001).

does not seem to be related to a detection limitation in the case of Rpb7, since occupancy values for these two RNAPII subunits are either similar or Rpb7 presents even higher occupancy values than Rpb11 in the TRs proximal to the TIS of these two genes. Interestingly, after heat shock, both Rpb7 and Rpb11 are detected down to the end of both TRs. Moreover, Rpb7 is detected at the 3' extremity of the HSPA8 gene, which is in agreement with the hypothesis of its involvement in the formation of 3' termini of the mRNA [47]. We cannot rule out the possibility that Rpb7 remains associated with the elongating polymerase all along the genes under normal conditions, but if this is the case, its conformation or its position in the elongation complex must be different from those under heat shock. Our results showing that Rpb4 is recruited to TRs (Supplementary Figure 3) confirm that the Rpb4–Rpb7 heterodimer has a function during transcriptional elongation. It has been previously shown that Rpb4 genetically interacts with the elongation factor TFIIS/TCEA1 and with the RNAPII subunit Rpb9, which also plays a role in elongation [48,49]. Moreover, it has been shown that yeast Rpb4 is a regulator of transcription-coupled DNA repair pathways

[50]. Taken together, our results suggest that the Rpb4–Rpb7 heterodimer is involved in both the initiation and elongation stages of transcription, having a specific role in stress conditions such as heat shock.

In conclusion, we used a systematic affinity tagging procedure coupled with ChIP experiments and gene-specific QPCR detection of the enriched genomic regions to determine the location of components of the general transcription machinery along active class II genes in human cells. Based on the location of the various polypeptides, our results help define their function during the transcription reaction in living mammalian cells.

We are grateful to the members of our laboratory, François Robert and the members of his laboratory and Jacques Archambault for helpful discussions and comments on this paper. We are also grateful to Jean-Marc Egly (IGBMC, Strasbourg, France) for antibodies. We thank Diane Bourque for artwork and Julie Edwards for a critical reading of this paper. This work is supported by grants from the Canadian Institutes for Health Research, Genome Canada and Genome Québec. M. C. and C. J. hold studentships from the Canadian Institutes for Health Research and the Fonds de recherche en Santé du Québec.

REFERENCES

- 1 Hahn, S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11**, 394–403
- 2 Müller, F. and Tora, L. (2004) The multicoloured world of promoter recognition complexes. *EMBO J.* **23**, 2–9
- 3 Bjorklund, S. and Gustafsson, C. M. (2005) The yeast Mediator complex and its regulation. *Trends Biochem. Sci.* **30**, 240–244
- 4 Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S. and Young, R. A. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728
- 5 Fan, X., Chou, D. M. and Struhl, K. (2006) Activator-specific recruitment of mediator *in vivo*. *Nat. Struct. Mol. Biol.* **13**, 117–120
- 6 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
- 7 Pan, G. and Greenblatt, J. (1994) Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. *J. Biol. Chem.* **269**, 30101–30104
- 8 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
- 9 Zurita, M. and Merino, C. (2003) The transcriptional complexity of the TFIIF complex. *Trends Genet.* **19**, 578–584
- 10 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
- 11 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
- 12 Mason, P. B. and Struhl, K. (2003) The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation *in vivo*. *Mol. Cell. Biol.* **23**, 8323–8333
- 13 Blau, J., Xiao, H., McCracken, S., O'Hare, P., Greenblatt, J. and Bentley, D. (1996) Three functional classes of transcriptional activation domain. *Mol. Cell. Biol.* **16**, 2044–2055
- 14 Sims, III, R. J., Belotserkovskaya, R. and Reinberg, D. (2004) Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* **18**, 2437–2468
- 15 Wang, D. and Hawley, D. K. (1993) Identification of a 3'→5' exonuclease activity associated with human RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 843–847
- 16 Zhang, C. and Burton, Z. F. (2004) Transcription factors IIF and IIS and nucleoside triphosphate substrates as dynamic probes of the human RNA polymerase II mechanism. *J. Mol. Biol.* **342**, 1085–1099
- 17 Jeronimo, C., Langelier, M. F., Zeghouf, M., Cojocaru, M., Bergeron, D., Baali, D., Forget, D., Mnaimneh, S., Davierwala, A. P., Pootoolal, J. et al. (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
- 18 Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emilii, A., Shilatifard, A., Buratowski, S. and Greenblatt, J. F. (2002) RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell. Biol.* **22**, 6979–6992
- 19 Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. et al. (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **33**, D154–D159
- 20 Nobrega, M. A., Ovcharenko, I., Afzal, V. and Rubin, E. M. (2003) Scanning human gene deserts for long-range enhancers. *Science* **302**, 413
- 21 Nagy, P. L., Cleary, M. L., Brown, P. O. and Lieb, J. D. (2003) Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6364–6369
- 22 Hsu, F., Pringle, T. H., Kuhn, R. M., Karolchik, D., Diekhans, M., Haussler, D. and Kent, W. J. (2005) The UCSC Proteome Browser. *Nucleic Acids Res.* **33**, D454–D458
- 23 Fish, R. N. and Kane, C. M. (2002) Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta* **1577**, 287–307
- 24 Nakanishi, T., Shimoaraiso, M., Kubo, T. and Natori, S. (1995) Structure–function relationship of yeast S-II in terms of stimulation of RNA polymerase II, arrest relief, and suppression of 6-azauracil sensitivity. *J. Biol. Chem.* **270**, 8991–8995
- 25 Pan, G., Ao, T. and Greenblatt, J. (1997) Interaction of elongation factors TFIIS and elongin A with a human RNA polymerase II holoenzyme capable of promoter-specific initiation and responsive to transcriptional activators. *J. Biol. Chem.* **272**, 24563–24571
- 26 Donahue, B. A., Yin, S., Taylor, J. S., Reines, D. and Hanawalt, P. C. (1994) Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8502–8506
- 27 Wong, J. M. and Ingles, C. J. (2001) A compromised yeast RNA polymerase II enhances UV sensitivity in the absence of global genome nucleotide excision repair. *Mol. Gen. Genet.* **264**, 842–851
- 28 Foustieri, M., Vermeulen, W., van Zeeland, A. A. and Mullenders, L. H. (2006) Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II *in vivo*. *Mol. Cell* **23**, 471–482
- 29 Prather, D. M., Larschan, E. and Winston, F. (2005) Evidence that the elongation factor TFIIS plays a role in transcription initiation at GAL1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **25**, 2650–2659
- 30 Martinez, E. (2002) Multi-protein complexes in eukaryotic gene transcription. *Plant Mol. Biol.* **50**, 925–947
- 31 Huisinga, K. L. and Pugh, B. F. (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol. Cell* **13**, 573–585
- 32 Timmers, H. T. and Tora, L. (2005) SAGA unveiled. *Trends Biochem. Sci.* **30**, 7–10
- 33 Martinez, E., Palhan, V. B., Tjernberg, A., Lymar, E. S., Gamper, A. M., Kundu, T. K., Chait, B. T. and Roeder, R. G. (2001) Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors *in vivo*. *Mol. Cell. Biol.* **21**, 6782–6795
- 34 Gromoller, A. and Lehming, N. (2000) Srb7p is essential for the activation of a subset of genes. *FEBS Lett.* **484**, 48–54
- 35 Zhu, X., Wren, M., Sinha, I., Rasmussen, N. N., Linder, T., Holmberg, S., Ekwall, K. and Gustafsson, C. M. (2006) Genome-wide occupancy profile of mediator and the srb8-11 module reveals interactions with coding regions. *Mol. Cell* **22**, 169–178
- 36 Andrau, J. C., van de Pasch, L., Lijnzaad, P., Bijma, T., Koerkamp, M. G., van de Peppel, J., Werner, M. and Holstege, F. C. (2006) Genome-wide location of the coactivator mediator: binding without activation and transient Cdk8 interaction on DNA. *Mol. Cell* **22**, 179–192
- 37 Woychik, N. A. and Hampsey, M. (2002) The RNA polymerase II machinery. Structure illuminates function. *Cell* **108**, 453–463
- 38 Orphanides, G., Lagrange, T. and Reinberg, D. (1996) The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657–2683
- 39 Hampsey, M. (1998) Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* **62**, 465–503
- 40 Fish, R. N., Ammerman, M. L., Davie, J. K., Lu, B. F., Pham, C., Howe, L., Ponticelli, A. S. and Kane, C. M. (2006) Genetic interactions between TFIIF and TFIIS. *Genetics* **173**, 1871–1884
- 41 Cheng, B. and Price, D. H. (2007) Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *J. Biol. Chem.* **282**, 21901–21912
- 42 Orlicky, S. M., Tran, P. T., Sayre, M. H. and Edwards, A. M. (2001) Dissociable Rpb4–Rpb7 subassembly of RNA polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J. Biol. Chem.* **276**, 10097–10102
- 43 Rosenheck, S. and Choder, M. (1998) Rpb4, a subunit of RNA polymerase II, enables the enzyme to transcribe at temperature extremes *in vitro*. *J. Bacteriol.* **180**, 6187–6192
- 44 Zhou, H. and Lee, K. A. (2001) An hsrPB4/7-dependent yeast assay for trans-activation by the EWS oncogene. *Oncogene* **20**, 1519–1524
- 45 Na, X., Duan, H. O., Messing, E. M., Schoen, S. R., Ryan, C. K., di Sant'Agnes, P. A., Golemis, E. A. and Wu, G. (2003) Identification of the RNA polymerase II subunit hsrPB7 as a novel target of the von Hippel–Lindau protein. *EMBO J.* **22**, 4249–4259
- 46 Croder, M. (2004) Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. *Trends Biochem. Sci.* **29**, 674–681
- 47 Mitsuizawa, H., Kanda, E. and Ishihama, A. (2003) Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res.* **31**, 4696–4701
- 48 Wery, M., Shematorova, E., Van Driessche, B., Vandenhaute, J., Thuriaux, P. and Van Mullem, V. (2004) Members of the SAGA and mediator complexes are partners of the transcription elongation factor TFIIS. *EMBO J.* **23**, 4232–4242
- 49 Awrey, D. E., Weibaecher, R. G., Hemming, S. A., Orlicky, S. M., Kane, C. M. and Edwards, A. M. (1997) Transcription elongation through DNA arrest sites: A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J. Biol. Chem.* **272**, 14747–14754
- 50 Li, S. and Smerdon, M. J. (2002) Rpb4 and Rpb9 mediate subpathways of transcription-coupled DNA repair in *Saccharomyces cerevisiae*. *EMBO J.* **21**, 5921–5929

Received 4 June 2007/20 August 2007; accepted 12 September 2007

Published as BJ Immediate Publication 12 September 2007, doi:10.1042/BJ20070751

CHAPITRE 5

LARP7 EST UNE COMPOSANTE STABLE DE LA snRNP 7SK ALORS QUE P-TEFb, HEXIM1 ET hnRNP A1 SONT ASSOCIÉS DE FAÇON RÉVERSIBLE

5.1 Préambule

J'ai contribué à l'article du chapitre 5 en réalisant les expériences montrant que LARP7, un polypeptide à fonction inconnue possédant des domaines LA et RRM de liaison à l'ARN (figure 1A), est un composant de la ribonucléoprotéine 7SK•HEXIM•MEPCE•P-TEFb (figures 1C et 1D). J'ai aussi effectué les expériences d'interférence par l'ARN double brin (siARN) présentées aux figures 4A et 4B montrant que LARP7 joue un rôle dans la stabilité du snARN 7SK dans les cellules humaines.

Annie Bouchard a purifié le complexe LARP7-TAP présenté à la figure 1B. Brian Krueger, Sarah A. Byers, Courtney E. Searcey et Jeffrey J. Cooper, membres de l'équipe du Dr David Price, ont produit l'anticorps spécifique de LARP7 (figure 1E), réalisé les expériences de séparation du complexe 7SK•HEXIM•MEPCE•P-TEFb•LARP7 sur gradient de glycérol avant et après le traitement avec flavopiridol (figure 2), immunoprécipité LARP7 confirmant que cette protéine est une composante stable de la snRNP 7SK (figures 3A et 3B), montré que le siARN de LARP7 augmente P-TEFb libre dans les cellules (figures 4C, 4D et 4E), et que le siARN de 7SK défait le complexe de grande taille de P-TEFb (figure 5). Charlotte Barrandon, étudiante chez le Dr Olivier Bensaude, a réalisé les expériences de co-immunoprécipitation de la snRNP 7SK à l'aide des anticorps dirigés contre HEXIM1, hnRNP A1 et LARP7 avant et après le traitement avec DRB (figure 3C). Bibhuti B. Roy, chercheur associé chez le Dr Éric Cohen, a effectué l'essai de transactivation du promoteur LTR du virus VIH-1 par la protéine Tat présenté à la figure 4F.

5.2 Article

Krueger, B.J., **Jeronimo, C.**, Roy, B.B., Bouchard, A., Barrandon, C., Byers, S.A., Searcey, C.E., Cooper, J.J., Bensaude, O., Cohen, E.A., Coulombe, B., Price, D.H. (2008) LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. **Nucleic Acids Res.** Advance Access published on February 16, 2008, doi:10.1093/nar/gkn061.

LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated

Brian J. Krueger¹, Célia Jeronimo², Bibhuti Bhusan Roy³, Annie Bouchard², Charlotte Barrandon⁴, Sarah A. Byers¹, Courtney E. Searcey¹, Jeffrey J. Cooper⁵, Olivier Bensaude⁴, Éric A. Cohen³, Benoit Coulombe² and David H. Price^{1,5,*}

¹Molecular and Cellular Biology Program, University of Iowa, Iowa City, Iowa, USA, ²Gene Transcription and Proteomics Laboratory, Institut de recherches cliniques de Montréal, ³Human Retrovirology Laboratory, Institut de recherches cliniques de Montréal, Montréal, Québec, Canada H2W 1R7, ⁴UMR 8541 CNRS, Ecole Normale Supérieure, 75230 Paris Cedex 05, France and ⁵Biochemistry Department, University of Iowa, Iowa City, Iowa, USA

Received December 17, 2007; Revised and Accepted January 30, 2008

ABSTRACT

Regulation of the elongation phase of RNA polymerase II transcription by P-TEFb is a critical control point for gene expression. The activity of P-TEFb is regulated, in part, by reversible association with one of two HEXIMs and the 7SK snRNP. A recent proteomics survey revealed that P-TEFb and the HEXIMs are tightly connected to two previously-uncharacterized proteins, the methyphosphate capping enzyme, MEPCE, and a La-related protein, LARP7. Glycerol gradient sedimentation analysis of lysates from cells treated with P-TEFb inhibitors, suggested that the 7SK snRNP reorganized such that LARP7 and 7SK remained associated after P-TEFb and HEXIM1 were released. Immunodepletion of LARP7 also depleted most of the 7SK regardless of the presence of P-TEFb, HEXIM or hnRNP A1 in the complex. Small interfering RNA knockdown of LARP7 in human cells decreased the steady-state level of 7SK, led to an initial increase in free P-TEFb and increased Tat transactivation of the HIV-1 LTR. Knockdown of LARP7 or 7SK ultimately caused a decrease in total P-TEFb protein levels. Our studies have identified LARP7 as a 7SK-binding protein and suggest that free P-TEFb levels are determined by a balance between release from the large form and reduction of total P-TEFb.

INTRODUCTION

The human positive transcription elongation factor b (P-TEFb), which is composed of Cdk9 and cyclin T1 or cyclin T2 (1–3), stimulates the elongation phase of transcription by reversing the effects of negative elongation factors [for recent reviews see (4,5)]. P-TEFb plays an important role in the transcription of cellular genes (6), and is also a key factor for the expression of the human immunodeficiency virus type 1 (HIV-1) genome (7–9). Previous studies have shown that a complex containing the 7SK small nuclear RNA (snRNA), a 332-nucleotide transcript synthesized by RNAPIII (10,11), and the RNA binding proteins HEXIM1 (12,13) or HEXIM2 (14,15) can interact with P-TEFb and inhibit its kinase activity (16). Signal transduction pathways have been implicated in the general release of P-TEFb from the large form during cardiac hypertrophy (17,18) and upon treatment of cells with the differentiation agent, HMBA (19). Also, transfection of cells with the HIV transactivator, Tat, leads to release of P-TEFb from the large form and the formation of a Tat•P-TEFb complex (20). Recent results from several labs indicate that P-TEFb may play a critical role during development. Poised polymerases have been found on most human genes in embryonic stem cells (21) and on most developmental control genes in *Drosophila* (22,23).

Recently, soluble human protein complexes containing components of the transcription and RNA processing machineries were analyzed using protein affinity purification coupled to mass spectrometry. Thirty-two tagged

*To whom correspondence should be addressed. Tel: +1 319 335 7910; Fax: +1 319 384 4770; Email: [REDACTED]

polypeptides yielded a network of 805 high-confidence interactions (24). This study revealed that besides its positive (Brd4) (25,26) and negative (HEXIMs and 7SK snRNA) regulators, P-TEFb is tightly connected to many other proteins, including the previously uncharacterized protein BCDIN3 [Bicoid-interacting 3, homolog (*Drosophila*)]. BCDIN3 is a conserved methyltransferase that has the ability to add a methyl group on the γ -phosphate of 7SK and because of this was renamed the methyl phosphate capping enzyme, MEPCE (24). The addition of this unusual mono-methyl cap structure to RNAPIII-synthesized snRNAs, such as 7SK, was previously shown to occur post-transcriptionally and to be important for protecting the RNA from exonucleolytic degradation (27). Indeed, it has been shown that the cap structure enhances the stability of U6 and 7SK snRNAs and that uncapped U6 snRNA is rapidly degraded (28). In support for a role of capping by MEPCE on 7SK stability, silencing of MEPCE was shown to decrease the steady-state level of cellular 7SK *in vivo* (24). Here we follow-up on another previously uncharacterized protein, LARP7, which was discovered to be connected with P-TEFb and HEXIM proteins (24).

MATERIALS AND METHODS

Affinity purification of a human LARP7-containing complex

The cDNA encoding human LARP7 (Invitrogen; accession number BC066945) was cloned into the mammalian expression vector pMZI (29) carrying a tandem affinity purification (TAP) tag at its C-terminus and a stable human embryonic kidney (HEK) cell line EcR-293 (derived from HEK 293) carrying this construct was produced (30). The conditions for expression, affinity purification, mass spectrometry identification of proteins and gel filtration chromatography were as previously described (24).

Generation of affinity purified LARP7 antibodies

C-terminally His-tagged recombinant LARP7 was expressed in *Escherichia coli* BL21 CodonPlus (DE3) RIL cells (Stratagene) from a LARP7 containing pET23a expression vector (Novagene). Cells were induced overnight with 0.1 mM isopropyl 1-thio- β -D-galactopyranoside at 18°C. Purification on Ni-NTA resin and mono S was carried out as previously described for the purification of HEXIM proteins (14). One fraction from the Mono S elution containing predominately a 38 kDa, C-terminal proteolysis product of LARP7 was used as an antigen to generate sheep antibodies (Elmira Biologicals). The LARP7 antibodies were affinity purified from the crude serum by the method previously described for purification of HEXIM2 antibodies (14).

Glycerol gradient analysis

HeLa cells were grown to 90% confluence in two T-150 flasks. Cells were then treated for 1 h with 500 nM Flavopiridol or mock-treated with carrier (0.004%

DMSO). Cells were harvested, lysed with a buffer containing 150 mM NaCl, and subjected to glycerol gradient fractionation as previously described (31). Gradient fractions were resolved by 9% SDS-PAGE followed by transfer to 0.45 μ m nitrocellulose membranes. The antibodies used for western blotting were: sheep anti-cyclin T1 (ab27963; Abcam), rabbit anti-Cdk9 (sc-8338; Santa Cruz Biotechnology), affinity-purified sheep anti-HEXIM1 (Abcam), rabbit anti-MEPCE (24) and affinity-purified sheep anti-LARP7. Western blots were incubated with the appropriate horseradish peroxidase-conjugated secondary antibody (Sigma) and then treated with Super Signal Dura West Extended Duration Substrate (Pierce). Blots were imaged using a cooled charge-coupled camera (UVP).

Immunoprecipitation

Glycerol gradient analyses of Flavopiridol-treated or untreated HeLa cells were performed as described before. Fractions 8–11, which contain the majority of LARP7 and the large form of P-TEFb, were pooled. To immunoprecipitate LARP7, affinity purified LARP7 antibodies were covalently attached to Actigel ALD beads (Sterogene) according to the manufacturer's instructions. Glycerol gradient pools were pre-cleared with beads alone at 4°C for 1 h. The flow through from the pre-clearing was then added to beads bound with affinity purified LARP7 or beads alone and rotated for 1 h at 4°C. The beads were washed with IP buffer [10 mM HEPES, 2 mM MgCl₂, 10 mM KCl, 0.1% NP-40, 0.5 mM EDTA, 150 mM NaCl, 0.1% PMSF, and 1 unit/ml Complete EDTA Free Protease Inhibitor Cocktail (Roche)]. Trizol reagent (Invitrogen) was then added directly to the beads to extract immunoprecipitated 7SK according to the manufacturer's protocol. Extracted RNA was resolved on a denaturing 6% TBE acrylamide RNA gel, ethidium bromide (EtBr) stained, and visualized using a cooled charge-coupled camera (UVP). RNAs were transferred from the acrylamide gels to 0.2 μ m Nytran N nylon membranes (Whatman) and then blocked for 1 h at 65°C in Ultrahybe Hybridization Buffer (Ambion). Radioactively tagged RNA oligos for U2, U6 or 7SK (31) were then added to the hybridization buffer and incubated overnight at 65°C. Blots were visualized using a Packard InstantImager and then exposed to film.

Co-immunoprecipitation of LARP7 associated proteins was similar; however, protein G Sepharose beads (Sigma) were used as the immunoprecipitation substrate to obtain better yields of LARP7 which was bound tightly by the antibodies. After immunoprecipitation, beads were resuspended in SDS loading buffer and boiled for 10 min at 100°C to release the bound protein from the beads. The immunoprecipitated proteins were resolved, blotted and probed with LARP7, Cdk9, MEPCE, and HEXIM1 antibodies as described before.

For co-immunoprecipitation from DRB-treated cells, cells were treated with 100 μ M DRB for 1 h, and whole cell extracts were prepared using the same lysis buffer as was

used for glycerol gradient analysis. Co-immunoprecipitations were carried out as described before.

Electrophoretic mobility shift assay and 7SK *in vitro* transcription

The mobility shift assay was carried out as previously described for the binding of HEXIM1 to RNA (31). In short, 0, 10 or 30 ng of recombinant LARP7 was combined in binding buffer (25 mM HEPES pH 7.6, 15% glycerol, 60 mM KCl, 0.1 mM EDTA, 5 mM DTT, 0.01% NP-40, 100 ng/ml BSA, 200 ng/rxn yeast tRNA) with 1 ng α -³²P-UTP labeled 7SK for 20 min at room temperature and then resolved on a 4% native Tris polyacrylamide gel. 7SK was transcribed *in vitro* from a cut pCR2.1 (Invitrogen) template driven by a T7 promoter using T7 polymerase (Stratagene) in the presence of α -³²P-UTP.

Small interfering RNA knockdown and luciferase assay

LARP7-specific small interfering RNA (siRNA) (custom synthesized by Integrated DNA Technologies, target sequence: 5'-ACAAGCGAGUAAACAUAUA-3'), MEPCE-specific siRNA (24) and control siRNA (Dharmacon siCONTROL Non-targeting pool), 7SK-specific siRNA (custom synthesized by Integrated DNA Technologies, targeting nucleotides 220-238) were transfected in 50%-confluent HEK 293 cells, TZM-bl cells or HeLa cells using Lipofectamine 2000 (Invitrogen) at a final siRNA concentration of 100 nM. TZM-bl cells, HeLa-cell derivatives that are stably transduced with a LTR-driven firefly luciferase cassette, were obtained from the NIH AIDS Reference and Reagent Program. Cells were transfected with the siRNAs twice at a 24 h interval. At various time intervals post-transfection, cells were lysed and LARP7 expression levels were monitored by western blotting. RNA blot analysis of total RNAs from LARP7-TAP eluate, HEK 293 cells mock-treated or treated with siRNA were carried out as previously described.

For luciferase assays, HEK 293 cells or TZM-bl cells were first mock-transfected or transfected with the control, MEPCE, LARP7 or double (MEPCE and LARP7) siRNAs as indicated before. Following an incubation of 5 h, the media was changed and a second siRNA transfection was performed 24 h later using similar conditions. After an incubation of 3 h, cells were co-transfected with 0.5 μ g of the luciferase reporter construct and 0.1 μ g of the Tat expressing plasmid or control plasmid. The plasmid encoding Tat (SV CMV Tat), the control plasmid (SV CMVexPA), as well as the LTR-driven firefly luciferase plasmid (pGL2-43-LTR-luc) were described previously (32). Transfected cells were collected 24 h post-final transfection and lysed using the luciferase assay system buffer (Promega). The protein concentration in cell lysates was determined using the Bio-Rad Protein Assay kit (Bio-Rad Laboratories). The luciferase activity was measured using the Dual-Luciferase[®] Reporter Assay System (Promega).

RESULTS

LARP7 is associated with P-TEFb, HEXIM1/2, MEPCE and 7SK RNA

In a recent study (24), a protein interaction network of human transcription and RNA processing machinery was constructed. P-TEFb was found to be associated with its negative regulators (HEXIMs and 7SK snRNA), the 7SK capping enzyme (MEPCE), and among other proteins, a previously uncharacterized protein, LARP7. A bioinformatics analysis revealed that LARP7 is a member of a family of about a dozen La related proteins generated from seven human genes. Of all the family members LARP7 is the most closely related to La. In their amino terminal halves, both La and LARP7 contain the La domain and an associated RNA recognition motif (RRM1) (Figure 1A). A recent structure determination using a co-crystal of La and an RNA oligo (33) strongly suggests that both the La domain and RRM1 are involved in the interaction of La with the 3' end of newly synthesized RNAPIII transcripts (34,35). Both proteins also contain a second RRM (RRM3) in their carboxyl terminal half (Figure 1A). Previous studies have shown that the La protein is a multifunctional protein that binds to and plays key roles in RNA metabolism (35,36). In the study presented here, we further characterize the association of LARP7 with the 7SK snRNP and determine the functional consequences of knocking down LARP7 in human cells.

To better identify the interaction partners of LARP7, a human HEK 293 cell line was stably transfected with a construct that expressed C-terminally TAP-tagged LARP7 at about the same level as the endogenous LARP7. The proteins found in affinity purified LARP7 complexes were identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS) after separation of the proteins on an SDS gel. Figure 1B shows the composition of the TAP-tagged LARP7 eluate. Proteins that have been either identified in a control eluate prepared from the non-induced cell extract or represent interactions that had Interaction Reliability scores below a stringent threshold (24) were not listed. RNA extracted from the TAP-tagged LARP7 eluate was subjected to northern analysis. Consistent with the results obtained earlier with MEPCE (24), 7SK and U6 snRNAs, but not U2 snRNA were pulled down with LARP7 (Figure 1C). As we have previously shown for the MEPCE-TAP eluate (24), both 7SK and U6 RNAs were found partially degraded when compared to their respective cellular counterpart isolated from total cell RNA. This is due to extensive incubations of the RNPs in crude extracts. Also relatively less U6 came down with LARP7 compared to that brought down by MEPCE suggesting that LARP7 might preferentially associate with 7SK or that MEPCE might preferentially associate with U6. Combining the data obtained from the previous study in which MEPCE and Cdk9 were tagged (24) with the data obtained here by TAP tagging LARP7 allowed us to construct an interaction network for LARP7, MEPCE and P-TEFb (Figure 1D). Note that arrows point from the tagged protein to the protein discovered to interact. Many of the

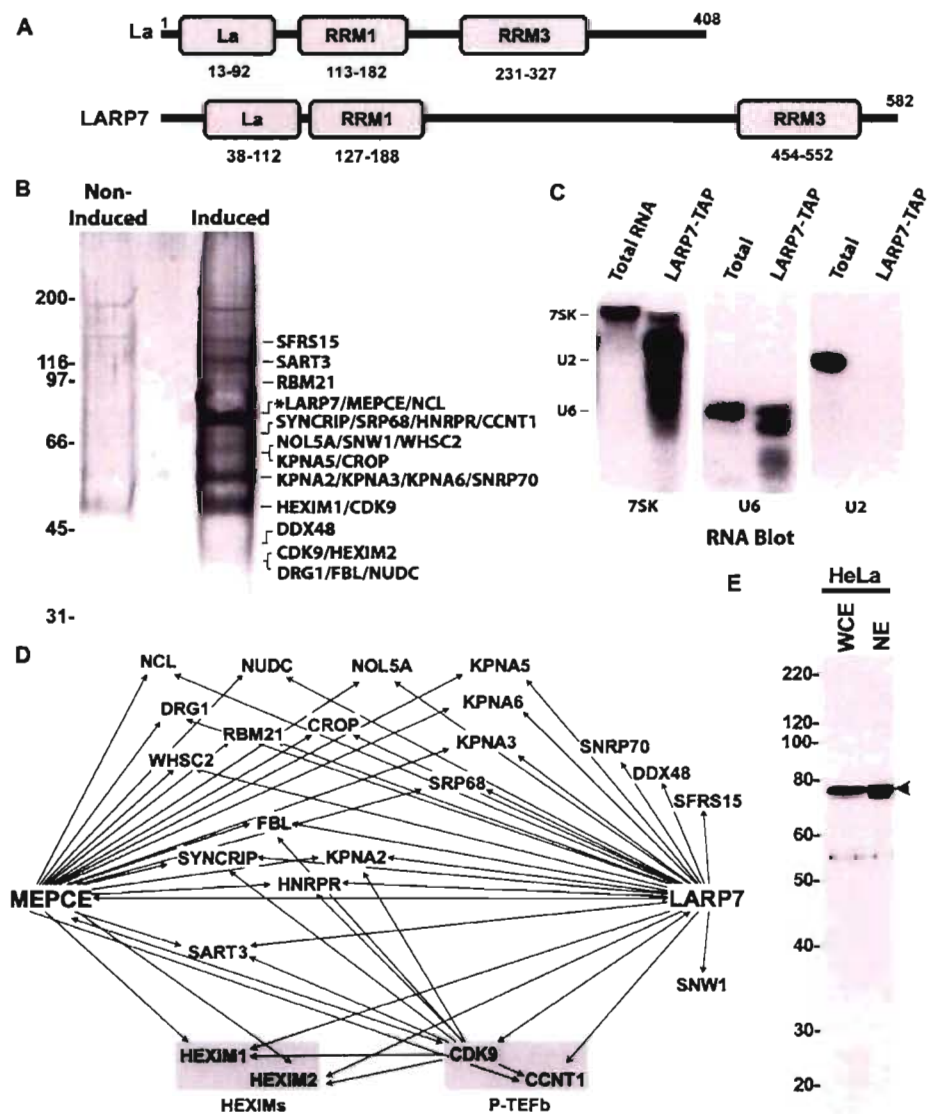


Figure 1. LARP7 is associated with P-TEFb, HEXIM, MEPCE, 7SK snRNA and U6 snRNA. (A) Schematic comparison of La with LARP7 with information obtained by querying NCBI's Conserved Domain Database. La—La domain, RRM1—RNA recognition motif I, RRM3—RNA recognition motif 3. (B) SDS gel showing the composition of an affinity-purified LARP7-TAP eluate. The tagged protein is indicated by an asterisk. (C) Northern blots were performed on total RNA extracted from HEK 293 cells (T; 900 ng) or RNA extracted from the LARP7-TAP eluate (L; 60 ng) and probed with RNA oligos specific for U2, U6 or 7SK snRNAs. (D) Network highlighting the interactions of LARP7 with various RNA processing factors, P-TEFb (CCNT1/cyclin T1 and Cdk9 subunits), its negative regulatory factors (HEXIM1 and HEXIM2) and the 7SK snRNA capping enzyme MEPCE. (E) Western blot of HeLa whole cell extract (WCE) or HeLa nuclear extract (NE) probed with the affinity purified LARP7 antibody generated in sheep. Arrowhead—LARP7.

proteins discovered interact with both MEPCE and LARP7. Not surprisingly, LARP7 is tightly connected to P-TEFb (cyclin T1 and Cdk9 subunits), its negative regulatory factors (HEXIM1 and HEXIM2) and MEPCE (Figure 1D). Together, our results indicate that LARP7 may be a part of the 7SK•HEXIM•MEPCE•P-TEFb snRNP. To facilitate further analysis of the endogenous LARP7 protein, we developed a highly specific LARP7 antibody in sheep that detects a band near the predicted molecular weight (69 kDa) of LARP7 (Figure 1E).

LARP7 co-sediments with large form P-TEFb and 7SK

To analyze the potential association of LARP7 and MEPCE with P-TEFb, HEXIM1 and 7SK, glycerol gradient sedimentation analysis of HeLa cell lysates was carried out. Otherwise identical cultures were either mock-treated or treated with 500 nM Flavopiridol for 1 h. Flavopiridol is a potent P-TEFb inhibitor that leads to inhibition of RNAPII transcription and to the release of P-TEFb and HEXIM1 from the large 7SK containing snRNP (37). As expected, in control cells most of the

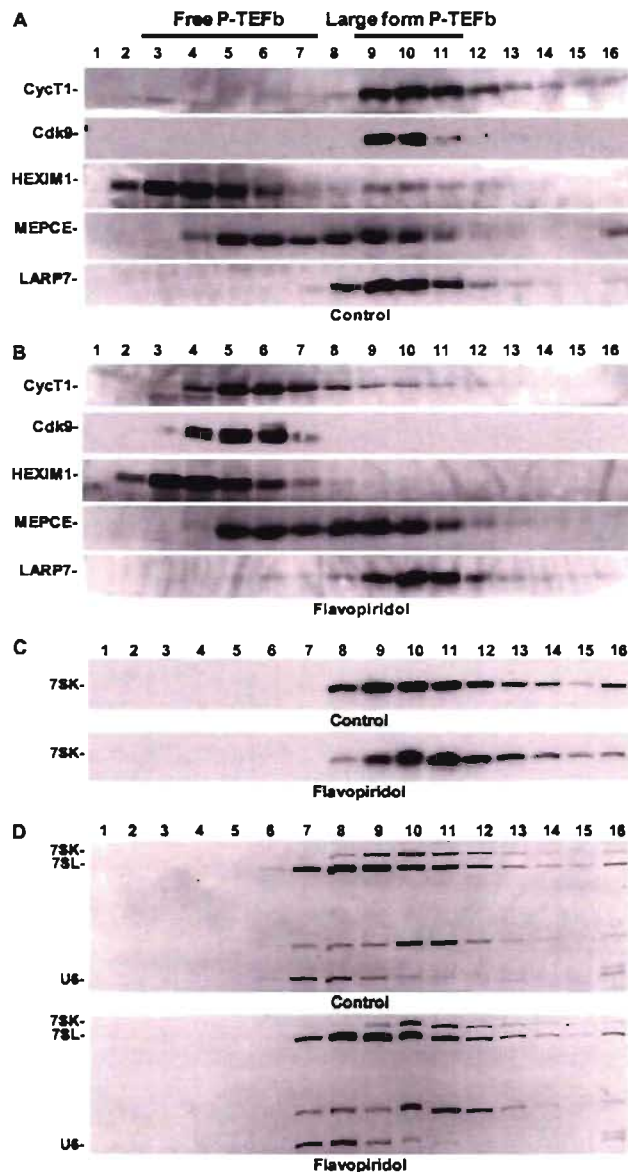


Figure 2. LARP7 co-sediments with the 7SK snRNP before and after P-TEFb and HEXIM1 release. (A) Glycerol gradient sedimentation analysis of untreated HeLa cells. Western blotting was performed with the indicated antibodies. (B) Gradient analysis of cells treated with 500 nM Flavopiridol for 1 h. (C) Northern blot of 7SK from gradient samples used in A and B. (D) Ethidium bromide stain of the RNA gels in C prior to northern transfer.

P-TEFb (cyclin T1 and Cdk9) and about 20% of HEXIM1 were found in fractions 9–11 that contain the 7SK snRNP with inactive P-TEFb (Figure 2A). The 7SK capping enzyme, MEPCE, was previously found in a complex containing P-TEFb, HEXIM1 and 7SK (24). Here slightly less than half of MEPCE was found to co-sediment with P-TEFb in control cells (Figure 2A) and the rest exhibited a slower sedimentation (found toward the top of the gradient). This observation is not surprising since MEPCE was found associated predominantly with

U6 snRNA and could potentially be associated with other small RNAs that it targets for capping (24,38). LARP7, on the other hand, completely co-sedimented with P-TEFb and HEXIM1 in the 7SK snRNP (Figure 2A). Flavopiridol treatment led to the expected release of P-TEFb and HEXIM1 from the 7SK snRNP and subsequent shift towards the top of the gradient (Figure 2B). MEPCE sedimentation was unaffected. However, LARP7 shifted about one fraction down the gradient (Figure 2B).

To determine if this shift down the gradient was consistent with LARP7 maintaining an association with 7SK after release of P-TEFb and HEXIM1, we extracted RNA from each of the glycerol gradient fractions and subjected them to northern analysis. In control cells, 7SK appeared in fractions 8–14, peaking in fraction 9; however, after Flavopiridol treatment, 7SK shifted down the gradient and peaked in fraction 10 (Figure 2C). To be sure that the shift in 7SK was not due to differences in the sedimentation or fractionation of the two samples, we visualized other RNAs with EtBr staining of the gel before the transfer. The pattern of sedimentation of other RNAs was identical between the two gradients, but as was found by the northern analysis, the pattern of 7SK changed (Figure 2D). These data strongly suggest that LARP7 remains in a stable complex with the 7SK snRNP even after Flavopiridol treatment and the release of P-TEFb and HEXIM1 from the snRNP. The shift down the gradient is likely due to association of hnRNP proteins (39–41).

LARP7 is a stable component of the 7SK snRNP

Results presented so far strongly suggest that LARP7 forms a complex with the large form of P-TEFb and remains in complex with 7SK after P-TEFb release. To determine if endogenous LARP7 is actually a component of these complexes, we carried out a series of immunoprecipitation experiments using an affinity purified LARP7 antibody. Glycerol gradient fractions 8–11, which contain ~80% of LARP7 (Figure 2A and B), were pooled from either control or Flavopiridol-treated cells and passed over beads that lacked antibody or had LARP7 antibody covalently attached. Analysis by either northern blot or EtBr staining of the RNA gel gave the same result. About 90% of 7SK was retained on the beads containing LARP7 antibody, but not on control beads (Figure 3A). Here, because care was taken to avoid ribonucleases, all the 7SK detected was intact. This indicates that LARP7 interacts with full length 7SK and not just the degradation products detected after the TAP purification earlier. Flavopiridol treatment of cells and the subsequent release of P-TEFb and HEXIM1 did not change the association of 7SK with LARP7 (Figure 3A). Pulldown of 7SK from these samples was highly specific, and did not affect the binding of the other small RNAs found in the gradient fractions (Figure 3A, ethidium bromide). Immunoprecipitation of LARP7 also resulted in the immunodepletion of HEXIM1 and Cdk9 from control gradient samples, and as expected these two proteins were absent from the samples from Flavopiridol-treated cells (Figure 3A, western blot). Some, but not all, of the

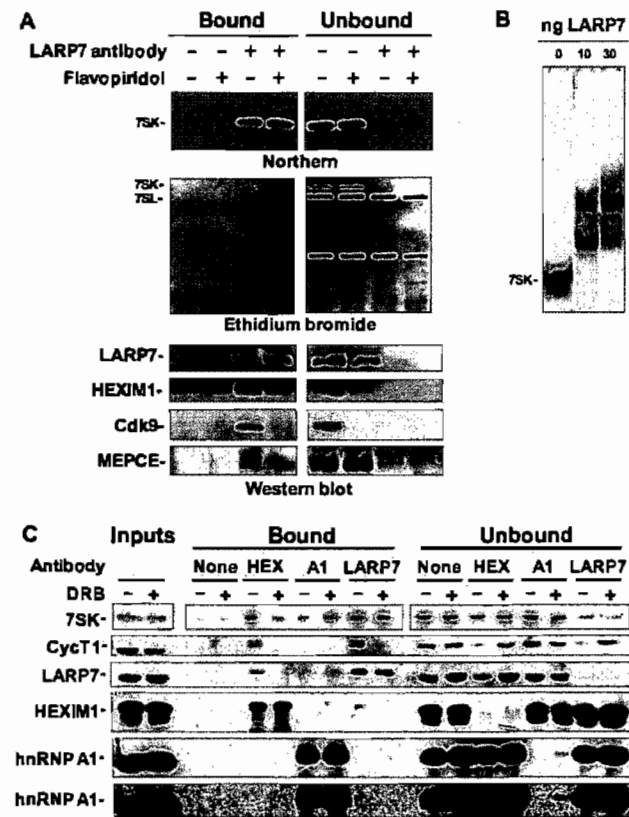


Figure 3. Immunoprecipitation of LARP7 confirms that it is a stable component of the 7SK snRNP. (A) Glycerol gradient fractions were pooled and subjected to LARP7 immunoprecipitation as described in the text. Nucleic acid and protein bound to the beads were either Trizol extracted or released in SDS loading buffer and analyzed by northern blotting, ethidium bromide staining and western blotting with the indicated antibodies. (B) Autoradiograph from the EMSA of *E. coli* expressed recombinant LARP7 bound to radioactively labeled 7SK. (C) Co-immunoprecipitation of the 7SK snRNP with HEXIM1 (HEX), hnRNP A1 (A1) and LARP7 antibodies before and after DRB treatment. Samples were then subjected to northern blotting for 7SK and western blotting with the indicated antibodies. Two exposures of the hnRNP A1 western blot are shown.

MEPCE was pulled down with the LARP7 antibody and treatment of cells with Flavopiridol caused less MEPCE to be pulled down (Figure 3A, western blot).

To determine if LARP7 associates directly with 7SK, we generated recombinant LARP7 in *E. coli* and used the protein in an electrophoretic mobility shift assay (EMSA) with radioactively labeled 7SK as the probe (Figure 3B). As increasing amounts of LARP7 were added, two 7SK mobility shifts were observed. The two RNPs could be due to the binding of LARP7 to one or two molecules of 7SK or one or two molecules of LARP7 binding to a single molecule of 7SK. From these results we conclude that LARP7 binds directly to 7SK.

To further examine the proteins associated with the 7SK snRNP before and after release of P-TEFb, immunoprecipitations from whole cell extracts of control cells or cells treated with another P-TEFb inhibitor, DRB, were performed using LARP7, HEXIM1 and hnRNP A1

antibodies. Regardless of the presence of the large form of P-TEFb, LARP7 was associated with a significant fraction of 7SK (Figure 3C). Immunoprecipitation of the 7SK snRNP from control cells using a HEXIM1 antibody verified that LARP7 is associated with this complex before but not after release of P-TEFb (Figure 3C). Recent research has shown that HEXIM and P-TEFb are replaced by hnRNP proteins after release from the 7SK snRNP (39,40). Here we confirm that the association of hnRNP A1 with 7SK increases after release, and show that only after release does hnRNP A1 associate with LARP7. Note that the immunoprecipitation with the LARP7 antibody brought down only a small fraction (less than 5%) of the hnRNP A1 protein which is more clearly seen with a longer exposure (Figure 3C, bottom blot). The low fraction bound to 7SK is expected since hnRNP A1 is found in many other RNPs. Overall, these results demonstrate that LARP7 is a stable component of the 7SK snRNP while HEXIM1, P-TEFb and hnRNP A1 are not.

Knockdown of LARP7 decreases the steady-state level of 7SK while increasing free P-TEFb and enhancing Tat-mediated transcription

Given the well-established role of the La protein in the protection of RNAPIII transcripts from 3' exonucleolytic degradation (34), we asked whether knockdown of LARP7 expression by siRNA would have a negative effect on 7SK snRNA steady-state levels in human cells. Figure 4A shows that reducing the expression of LARP7 by about 70% *in vivo* (Figure 4A) caused a decrease in the steady-state level of 7SK (Figure 4B). LARP7 knockdown had no effect on U2 or U6 snRNA levels (Figure 4B). This result indicates that LARP7 stabilizes the 7SK snRNA in human cells.

Since knockdown of LARP7 resulted in reduced stability of 7SK, we next determined the effect of LARP7 knockdown on the ratio of the forms of P-TEFb and on P-TEFb activity. Glycerol gradient analysis was carried out on control cells or cells treated with either LARP7 or MEPCE siRNA. siRNA treatment resulted in a 70% reduction of both LARP7 and MEPCE (Figure 4C and D). Knockdown of either protein resulted in a small, but significant increase in the amount of free P-TEFb; however, the effect of MEPCE knockdown was less dramatic (Figure 4E). Because the release of P-TEFb is a highly regulated process, we speculated that the increase in free P-TEFb mediated by the loss of LARP7 and MEPCE and subsequent loss of 7SK would affect transcription. LARP7 and MEPCE siRNAs were introduced into human HEK 293 cells, followed by the transient transfection of an HIV-1 LTR-luciferase reporter construct in the presence of the HIV-1 transactivator, Tat. As shown in Figure 4F, knockdown of LARP7 caused a consistent ~2-fold increase in Tat-mediated transactivation of the HIV-1 LTR. Knockdown of MEPCE had little or no effect either alone or in combination with knockdown of LARP7. Essentially the same results were observed in TZM-bl cells, a stable HeLa-derived cell line containing an integrated luciferase

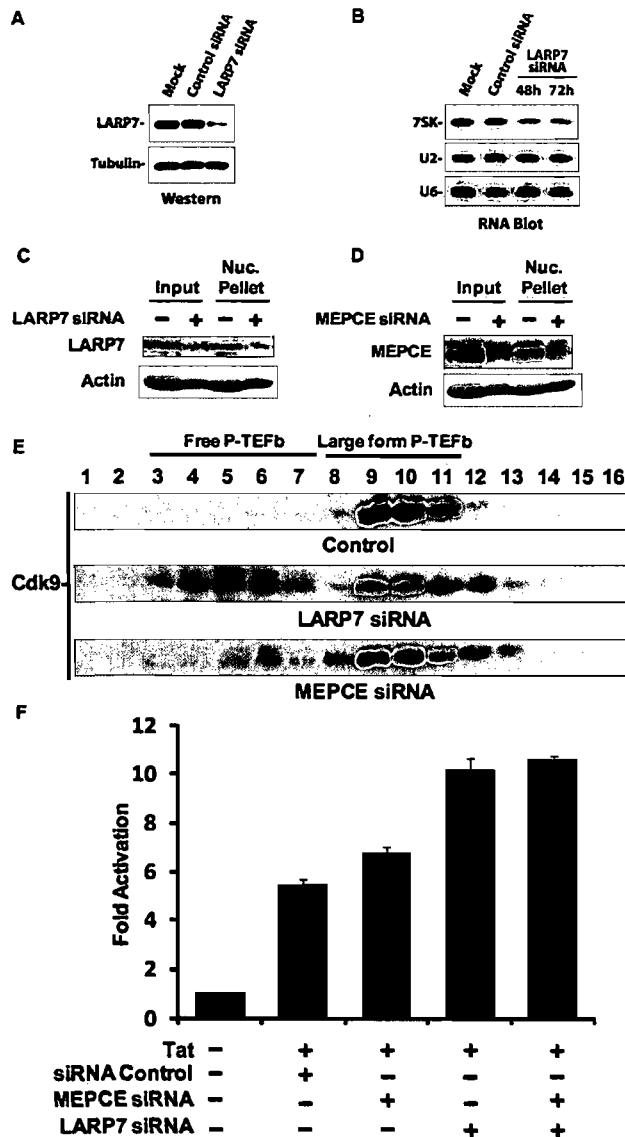


Figure 4. Knockdown of LARP7 decreases 7SK levels, increases free P-TEFb, and enhances Tat transactivation. (A) LARP7 knockdown (72 h post-transfection) was monitored by western blotting in extracts from HEK 293 cells mock-transfected or transfected with control or LARP7 siRNA (tubulin was used as a loading control). (B) 7SK steady-state levels were assessed by RNA blotting using total RNA extracted from HEK 293 cells mock-transfected or transfected with control or LARP7 siRNA (48 and 72 h post-transfection) and probed with RNA oligonucleotides that specifically detect 7SK, U6 or U2 snRNAs. (C) Knockdown of LARP7 in HeLa cell glycerol gradient input or the leftover nuclear pellet 72 h post-transfection was determined by western blot and comparison with an actin loading control. (D) Same conditions as C, except the siRNA used targeted MEPCE. (E) Glycerol gradient analysis of the large form of P-TEFb after siRNA knockdown of LARP7 and MEPCE using the Cdk9 antibody. (F) HEK 293 cells were mock-transfected or transfected twice with control, MEPCE, LARP7 or double (MEPCE and LARP7) siRNAs and an HIV-1 LTR-luciferase reporter construct. A Tat-expressing plasmid was co-transfected where indicated. Luciferase activity was measured 24 h after transfection. Results are presented as fold activation relative to transfections in the absence of the Tat expression plasmid. $N = 3$.

reporter gene driven by the HIV-1 LTR (data not shown). The relatively weak enhancement of Tat-mediated transactivation was most likely due to the 25–50% efficiency of LARP7 knockdown in treated cells (data not shown). Of note, these results compare well with those obtained in an earlier study of the effect HEXIM1 knockdown on transcription from the HIV-1 promoter, which also led to a ~2-fold increase (12). Together, these data provide evidence that LARP7 plays an important role in stabilizing and maintaining the 7SK snRNP. Knockdown of LARP7 led to both a reduction in 7SK and to an increase in active P-TEFb.

Knockdown of 7SK disrupts the large form of P-TEFb

To determine if the level of 7SK could directly impact the amount of free P-TEFb, siRNA directed against 7SK was used to knock down the snRNA. Previously, 7SK siRNA, but not a non-specific siRNA, was shown to specifically reduce 7SK even though it is a nuclear RNA (42). Total RNA was isolated 6, 12, 24, 48 or 72 h post-transfection from cells treated with siRNA or Lipofectamine 2000 alone. Northern blotting was performed to determine the extent of 7SK knockdown. Knockdown was rapid and persistent, with 7SK levels in siRNA-treated cells reduced to 21% of control cells at 6 h and 16% at 72 h (Figure 5A). The effect of knockdown of 7SK on the forms of P-TEFb was then examined by glycerol gradient sedimentation analysis of cell lysates made 48 h after mock transfection or transfection with 7SK siRNA. In the control cells, the vast majority of P-TEFb was found in the large form, with a small amount in the free form (Figure 5B, Control). The ratio of the free and large forms of P-TEFb was dramatically shifted by 7SK knockdown, with very little P-TEFb remaining in the large form (Figure 5B, 7SK siRNA). We conclude that reduction in the amount of 7SK in cells leads to the release of P-TEFb from the large form and is similar to the effect seen after LARP7 knockdown (Figure 4E) except that when 7SK siRNA is used the reduction in 7SK is much more rapid. However, at the 48 h time point analyzed the absolute signals for cyclin T1 and Cdk9 in 7SK siRNA-treated cells were reduced compared to control cells (Figure 5B).

To explore the possibility that 7SK siRNA treatment affected the overall level of P-TEFb, cells were transfected with siRNA or mock-transfected, lysed with SDS after 64 h, and the lysates were analyzed by northern and western blotting. Northern blotting for 7SK confirmed that it was reduced to ~20% of control cells (Figure 5C). Knockdown of 7SK caused a reduction in both Cdk9 and cyclin T1 to 30% (Figure 5C), indicating that the absolute amount of P-TEFb is dependent on 7SK levels. Evidently, the prolonged release of active P-TEFb from the disintegrating large form triggers the reduction of P-TEFb. Additionally, reduction of free P-TEFb could be part of the explanation for why only low levels of released P-TEFb were detected 72 h after the LARP7 siRNA treatment of cells (Figure 4E).

Finally, to determine if LARP7 knockdown had a similar effect as 7SK knockdown on total P-TEFb levels, HeLa cells were mock-transfected or transfected with 10,

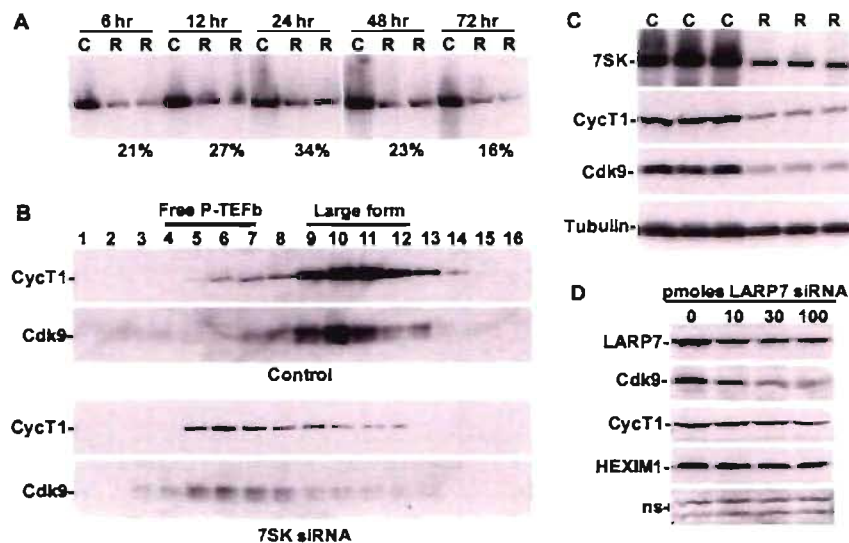


Figure 5. Direct knockdown of 7SK disrupts the large form of P-TEFb and decreases total P-TEFb. (A) HeLa cells were mock-treated or treated with 7SK siRNA for the indicated periods of time. One control and two independent siRNA treatments are shown for each time point. C-Control, R-siRNA-treated. (B) HeLa cells were mock-treated or treated with 7SK siRNA for 48 h, subjected to glycerol gradient sedimentation, and analyzed by western blot using the indicated antibodies. (C) HeLa cells were mock-treated or treated with 7SK siRNA for 64 h. Quantitative northern blotting for 7SK and western blotting for cyclin T1 and Cdk9 were performed. C-Control, R-siRNA-treated. (D) HeLa cells were treated with the indicated amount of LARP7 siRNA for 48 h, lysed in SDS loading buffer and western blotted with the indicated antibodies. Non-specific bands (ns) are provided as a loading control.

30 or 100 pmoles of LARP7 siRNA for 48 h. LARP7 knockdown resulted in a significant decrease in both Cdk9 and cyclin T1 levels, paralleling the 7SK results, while HEXIM1 was unchanged (Figure 5D). HEXIM1 levels likely remained the same due to the fact that HEXIM1 in the large form is a small fraction of the total HEXIM1 (Figure 2A). Loss of the 7SK snRNP by either 7SK or LARP7 knockdown underscores the importance of 7SK stability in maintaining proper control over P-TEFb.

DISCUSSION

We have examined the association of LARP7 with 7SK, P-TEFb, HEXIM1, MEPCE and hnRNP A1. Although the study began with a TAP-tagged LARP7 pulldown, all other experiments involved analysis of endogenous LARP7. Glycerol gradient sedimentation and immunoprecipitation experiments were used to compare the 7SK snRNP in control cells with the snRNP in cells treated with P-TEFb inhibitors or siRNAs. Our study demonstrated that LARP7 is a stable component of two major 7SK snRNPs that are present before and after the release of P-TEFb (Figure 6). Treatment of cells with P-TEFb inhibitors caused the P-TEFb•HEXIM1•MEPCE•LARP7•7SK complex to be converted into an hnRNPA1•LARP7•7SK complex. Only a small (substoichiometric) amount of MEPCE was found to associate with 7SK and LARP7 after release of P-TEFb. Knockdown of LARP7 led to a reduction of 7SK, the release of P-TEFb from the large form and the activation of transcription from the HIV LTR. Additionally, reduction of 7SK by direct siRNA knockdown or by knockdown of LARP7

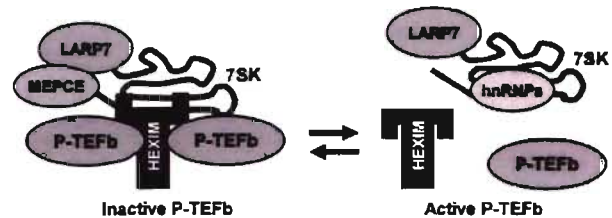


Figure 6. Model of the 7SK snRNP before and after release of P-TEFb. Prior to treatment with P-TEFb inhibitors or cellular stress, the 7SK snRNP contains P-TEFb, a HEXIM1 dimer, MEPCE, LARP7 and 7SK. Following release, the 7SK snRNP is composed of, LARP7, 7SK and hnRNP proteins.

caused a time-dependent reduction in the total amount of P-TEFb. This suggests that there is a cellular mechanism that regulates the level of free P-TEFb when it is not inhibited by the 7SK snRNP. Based on bioinformatic analysis LARP7 is the most closely related family member to the founding protein, La. The similarities in the domain structure of La and LARP7 suggest that they may have similar properties. La has been found in association with many nascent RNAPIII transcripts including 7SK (40,41,43). We propose that LARP7 replaces La during 7SK snRNP maturation. This would explain our finding that ~90% of the 7SK snRNP was immunoprecipitated with LARP7 antibodies (Figure 3). This represents a difference between LARP7 and La in that La is not present in the final mature form of most snRNPs (44–47). The results presented here show that LARP7 protects 7SK from degradation similarly to La's role in protecting other RNPs (34). It is possible that this is accomplished through the binding of the La domain and RRM1 of LARP7 to

both the 3' and 5' ends of 7SK (33). In addition to its role in RNA protection, La has been implicated in nuclear localization and nucleolar trafficking of RNAs such as U6 and 7SL (36). Binding of LARP7 to 7SK may partially explain why 7SK is predominantly nuclear (48). Interestingly, recent analysis of the crystal structure of La shows that the La domain contains hallmarks of a DNA binding protein and the surfaces that were assumed to be involved in RNA binding are actually free (33). In conjunction with this finding, La has been shown to localize to RNAPIII promoters (49). It would be intriguing if LARP7 had a similar localization to DNA. If so then it could either be used to target inactive P-TEFb to positions on the chromatin for local release to enhance RNAPII transcription, or position 7SK on chromatin for local regulated inhibition of P-TEFb.

Binding of La to its RNA targets is a regulated process that is dependent on the 5' and 3' ends of the RNA and the phosphorylation state of La. A 5'-monomethyl-guanosine capping of U6 and B2 RNA has been shown to reduce the affinity of La for these RNAs while adenylation of these transcripts completely releases La (50,51). The recent discovery of MEPCE, which monomethylates U6 and 7SK on their 5' termini may provide a mechanism for 7SK release from La and the subsequent association of LARP7. However, the status of the 5' end of 7SK bound by LARP7, as well as, the methylation preference of LARP7 has yet to be determined. Alternatively, it is possible that 5' end capping of 7SK might affect the way that LARP7 and 7SK interact and that could impact the association or dissociation of P-TEFb and HEXIM1. It is also possible that modification of the 3' end of 7SK could be used for regulation of the protein composition of the 7SK snRNP.

Approximately 70% of 7SK is post-transcriptionally adenylated on the 3' end (52) and this modification has been shown to affect La binding to other RNAPIII snRNAs. This modification may provide a mechanism for release of 7SK from La or release of P-TEFb from the LARP7•7SK snRNP. Release of P-TEFb from the LARP7•7SK snRNP could be accomplished by altering the association of LARP7 with 7SK in such a way that the secondary structure of the RNA is affected and prevents HEXIM and P-TEFb binding. It is possible that the 25% of 7SK that is associated with P-TEFb (Figure 3A, Northern and 3B, 7SK) might be the fraction of 7SK, ~30%, that is not adenylated. Phosphorylation of Ser366 of La has been shown to reduce 5' end binding (53) providing another feasible regulatory mechanism.

Results from the 7SK and LARP7 knockdown experiments performed here provide evidence for the coordination of two cellular mechanisms regulating the level of free P-TEFb. When the level of free P-TEFb was increased by disruption of the large form during siRNA-mediated knockdown of LARP7, an increase in P-TEFb activity was observed. After several days of LARP7 or 7SK siRNA treatment less total P-TEFb was found. This is consistent with an initial increase in P-TEFb activity as the large form is disrupted followed by a slower cellular compensation that leads to reduction in the total amount of P-TEFb protein by increasing turnover or decreasing

synthesis of Cdk9 and cyclin T1. The mechanism for the reduction of excess free P-TEFb is likely to be proteasome mediated decay because when P-TEFb subunits were overexpressed most of the P-TEFb was quickly ubiquitylated and degraded by the proteasome (54,55). Degradation of the free P-TEFb would explain why we found only a slight increase in free P-TEFb after LARP7 knockdown. The idea that the amount of free P-TEFb arises from a balance between release from the large form and degradation of the free form is supported by the finding that knockdown of either Cdk9 or cyclin T1 results in a reduction of the total amount of P-TEFb in the cell without grossly affecting transcription of a reporter gene or cell viability (56). In the future, it is important to determine how these mechanisms are used to obtain the selective P-TEFb function required for control of transcription during differentiation and development.

Mechanisms regulating the balance of free versus inactive P-TEFb have been previously characterized. Knockdown of HEXIM1 was shown to have no effect on the absolute levels of P-TEFb or the amount of P-TEFb in the large form because HEXIM1 was functionally replaced by HEXIM2 (14,15). Two studies have shown that long-term treatment of cells with HMBA, caused an initial increase in free active P-TEFb, but ultimately resulted in a shift in the equilibrium back to the large inactive form (19,57). Together these studies provide support for a variety of cellular mechanisms that maintain the P-TEFb equilibrium in cells.

Comparative genomic analysis shows that LARP7 is closely related to the *Drosophila melanogaster* protein multi sex combs (mxc). The domain structure of mxc is very similar to that of LARP7 and contains a La motif and two RRM. In flies, mxc is a member of the Polycomb group (PcG) of genes which regulate gene expression during differentiation and development (58). Mxc appears to function as a tumor suppressor because mutations were shown to result in uncontrolled malignant growth (59). Additionally, there is evidence that LARP7, studied under the alias DKFZP564K112, is a marker of gastric cancer as determined by microsatellite instabilotyping (60). Genetic studies in flies show that complete loss of mxc results in cell death after a few divisions while hypomorphic expression results in abnormalities during development (61). Mxc is also implicated in *Drosophila* hematopoiesis by controlling plasmacyte proliferation and differentiation (62). It would be interesting to see if the phenotypes observed in flies are related to an interaction between fly P-TEFb and its LARP7 homolog mxc. With these results in mind, it is important that we determine how LARP7 regulates P-TEFb function during cell differentiation, development and cancer progression.

ACKNOWLEDGEMENTS

We are grateful to Guy Poirier and Denis Faubert for mass spectrometry analysis, Mathieu Blanchette and Christian Poitras for help with bioinformatics and Benoit Chabot for assistance with siRNA design. This work was supported by NIH GM35500 (D.H.P.); the

Canadian Institutes for Health Research (B.C. and E.A.C.), the Natural Sciences and Engineering Research Council of Canada, Genome Canada and G enome Qu ebec (B.C.); the Association pour la Recherche sur le Cancer, Agence Nationale de Recherche sur le SIDA, Agence Nationale pour la Recherche, Fondation pour la Recherche M edicale (O.B.); the CIHR and the Fonds de la recherche en sant e du Qu ebec (C.J.) and the University of Iowa Molecular and Cellular Biology Training Grant (B.J.K.). Funding to pay the Open Access publication charges for this article was provided by NIH GM35500.

Conflict of interest statement. None declared.

REFERENCES

- Peng, J., Marshall, N.F. and Price, D.H. (1998) Identification of a cyclin subunit required for the function of *Drosophila* P-TEFb. *J. Biol. Chem.*, **273**, 13855–13860.
- Marshall, N.F., Peng, J., Xie, Z. and Price, D.H. (1996) Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.*, **271**, 27176–27183.
- Marshall, N.F. and Price, D.H. (1995) Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.*, **270**, 12335–12338.
- Peterlin, B.M. and Price, D.H. (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell*, **23**, 297–305.
- Zhou, Q. and Yik, J.H. (2006) The Yin and Yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. *Microbiol. Mol. Biol. Rev.*, **70**, 646–659.
- Chao, S.H. and Price, D.H. (2001) Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. *J. Biol. Chem.*, **276**, 31793–31799.
- Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H. and Jones, K.A. (1998) A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell*, **92**, 451–462.
- Mancebo, H.S., Lee, G., Flygare, J., Tomassini, J., Luu, P., Zhu, Y., Peng, J., Blau, C., Hazuda, D., Price, D. *et al.* (1997) P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. *Genes Dev.*, **11**, 2633–2644.
- Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M.B. and Price, D.H. (1997) Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev.*, **11**, 2622–2632.
- Nguyen, V.T., Kiss, T., Michels, A.A. and Bensaude, O. (2001) 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature*, **414**, 322–325.
- Yang, Z., Zhu, Q., Luo, K. and Zhou, Q. (2001) The 7SK small nuclear RNA inhibits the CDK9/cyclin T kinase to control transcription. *Nature*, **414**, 317–322.
- Yik, J.H., Chen, R., Nishimura, R., Jennings, J.L., Link, A.J. and Zhou, Q. (2003) Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol. Cell*, **12**, 971–982.
- Michels, A.A., Nguyen, V.T., Fraldi, A., Labas, V., Edwards, M., Bonnet, F., Lania, L. and Bensaude, O. (2003) MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol. Cell Biol.*, **23**, 4859–4869.
- Byers, S.A., Price, J.P., Cooper, J.J., Li, Q. and Price, D.H. (2005) HEXIM2, a HEXIM1-related protein, regulates positive transcription elongation factor b through association with 7SK. *J. Biol. Chem.*, **280**, 16360–16367.
- Yik, J.H., Chen, R., Pezda, A.C. and Zhou, Q. (2005) Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive positive transcription elongation factor b complexes for control of transcription. *J. Biol. Chem.*, **280**, 16368–16376.
- Michels, A.A., Fraldi, A., Li, Q., Adamson, T.E., Bonnet, F., Nguyen, V.T., Sedore, S.C., Price, J.P., Price, D.H., Lania, L. *et al.* (2004) Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J.*, **23**, 2608–2619.
- Sano, M. and Schneider, M.D. (2004) Cyclin-dependent kinase-9: an RNAPII kinase at the nexus of cardiac growth and death cascades. *Circ. Res.*, **95**, 867–876.
- Sano, M., Abdellatif, M., Oh, H., Xie, M., Bagella, L., Giordano, A., Michael, L.H., DeMayo, F.J. and Schneider, M.D. (2002) Activation and function of cyclin T-Cdk9 (positive transcription elongation factor-b) in cardiac muscle-cell hypertrophy. *Nat. Med.*, **8**, 1310–1317.
- Contreras, X., Barboric, M., Lenasi, T. and Peterlin, B.M. (2007) HMBA releases P-TEFb from HEXIM1 and 7SK snRNA via PI3K/Akt and activates HIV transcription. *PLoS Pathog.*, **3**, 1459–69.
- Sedore, S.C., Byers, S.A., Biglione, S., Price, J.P., Maury, W.J. and Price, D.H. (2007) Manipulation of P-TEFb control machinery by HIV: recruitment of P-TEFb from the large form by Tat and binding of HEXIM1 to TAR. *Nucleic Acids Res.*, **35**, 4347–4358.
- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. and Young, R.A. (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, **130**, 77–88.
- Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J. and Adelman, K. (2007) RNA polymerase is poised for activation across the genome. *Nat. Genet.*, **39**, 1507–1511.
- Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M. and Young, R.A. (2007) RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.*, **39**, 1512–1516.
- Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J. *et al.* (2007) Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol. Cell*, **27**, 262–274.
- Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N. and Ozato, K. (2005) The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell*, **19**, 523–534.
- Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K. and Zhou, Q. (2005) Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell*, **19**, 535–545.
- Shumyatsky, G.P., Tillib, S.V. and Kramerov, D.A. (1990) B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5' end. *Nucleic Acids Res.*, **18**, 6347–6351.
- Shumyatsky, G., Wright, D. and Reddy, R. (1993) Methylphosphate cap structure increases the stability of 7SK, B2 and U6 small RNAs in *Xenopus* oocytes. *Nucleic Acids Res.*, **21**, 4756–4761.
- Zeghouf, M., Li, J., Butland, G., Borkowska, A., Canadian, V., Richards, D., Beattie, B., Emili, A. and Greenblatt, J.F. (2004) Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes. *J. Proteome Res.*, **3**, 463–468.
- Jeronimo, C., Langelier, M.F., Zeghouf, M., Cojocar, M., Bergeron, D., Baali, D., Forget, D., Mnaimneh, S., Davierwala, A.P., Pootoolal, J. *et al.* (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.
- Li, Q., Cooper, J.J., Altwerger, G.H., Feldkamp, M.D., Shea, M.A. and Price, D.H. (2007) HEXIM1 is a promiscuous double-stranded RNA-binding protein and interacts with RNAs in addition to 7SK in cultured cells. *Nucleic Acids Res.*, **35**, 2503–2512.
- Harrison, G.P. and Lever, A.M. (1992) The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J. Virol.*, **66**, 4144–4153.
- Teplova, M., Yuan, Y.R., Phan, A.T., Malinina, L., Ilin, S., Teplov, A. and Patel, D.J. (2006) Structural basis for recognition and sequestration of UUU(OH) 3' termini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Mol. Cell*, **21**, 75–85.
- Maraia, R.J. and Bayfield, M.A. (2006) The La protein-RNA complex surfaces. *Mol. Cell*, **21**, 149–152.

35. Curry, S. and Conte, M.R. (2006) A terminal affair: 3'-end recognition by the human La protein. *Trends Biochem. Sci.*, **31**, 303–305.
36. Maraia, R.J. (2001) La protein and the trafficking of nascent RNA polymerase III transcripts. *J. Cell Biol.*, **153**, F13–18.
37. Chao, S.H., Fujinaga, K., Marion, J.E., Taube, R., Sausville, E.A., Senderowicz, A.M., Peterlin, B.M. and Price, D.H. (2000) Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. *J. Biol. Chem.*, **275**, 28345–28348.
38. Gupta, S., Busch, R.K., Singh, R. and Reddy, R. (1990) Characterization of U6 small nuclear RNA cap-specific antibodies. Identification of gamma-monomethyl-GTP cap structure in 7SK and several other human small RNAs. *J. Biol. Chem.*, **265**, 19137–19142.
39. Barrandon, C., Bonnet, F., Nguyen, V.T., Labas, V. and Bensaude, O. (2007) The transcription-dependent dissociation of P-TEFb-HEXIM1-7SK RNA relies upon formation of hnRNP-7SK RNA complexes. *Mol. Cell Biol.*, **27**, 6996–7006.
40. Van Herreweghe, E., Egloff, S., Goiffon, I., Jady, B.E., Froment, C., Monsarrat, B. and Kiss, T. (2007) Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. *EMBO J.*, **26**, 3570–3580.
41. Hogg, J.R. and Collins, K. (2007) RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA*, **13**, 868–880.
42. Robb, G.B., Brown, K.M., Khurana, J. and Rana, T.M. (2005) Specific and potent RNAi in the nucleus of human cells. *Nat. Struct. Mol. Biol.*, **12**, 133–137.
43. Stefano, J.E. (1984) Purified lupus antigen La recognizes an oligouridylylate stretch common to the 3' termini of RNA polymerase III transcripts. *Cell*, **36**, 145–154.
44. Guddat, U., Bakken, A.H. and Pieler, T. (1990) Protein-mediated nuclear export of RNA: 5S rRNA containing small RNPs in xenopus oocytes. *Cell*, **60**, 619–628.
45. Boelens, W.C., Palacios, I. and Mattaj, J.W. (1995) Nuclear retention of RNA as a mechanism for localization. *RNA*, **1**, 273–283.
46. Maraia, R., Zasloff, M., Plotz, P. and Adeniyi-Jones, S. (1988) Pathway of B1-Alu expression in microinjected oocytes: Xenopus laevis proteins associated with nuclear precursor and processed cytoplasmic RNAs. *Mol. Cell Biol.*, **8**, 4433–4440.
47. Steitz, J.A., Berg, C., Hendrick, J.P., La Branche-Chabot, H., Metspalu, A., Rinke, J. and Yario, T. (1988) A 5S rRNA/L5 complex is a precursor to ribosome assembly in mammalian cells. *J. Cell Biol.*, **106**, 545–556.
48. Gurney, T., Jr. and Eliceiri, G.L. (1980) Intracellular distribution of low molecular weight RNA species in HeLa cells. *J. Cell Biol.*, **87**, 398–403.
49. Fairley, J.A., Kantidakis, T., Kenneth, N.S., Intine, R.V., Maraia, R.J. and White, R.J. (2005) Human La is found at RNA polymerase III-transcribed genes in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, pp. 18350–18355.
50. Chen, Y., Sinha, K., Perumal, K., Gu, J. and Reddy, R. (1998) Accurate 3' end processing and adenylation of human signal recognition particle RNA and alu RNA in vitro. *J. Biol. Chem.*, **273**, 35023–35031.
51. Chen, Y., Perumal, K. and Reddy, R. (2000) Inhibition of translation of mRNAs containing gamma-monomethylphosphate cap structure in frog oocytes and in mammalian cells. *Gene Expr.*, **9**, 133–143.
52. Sinha, K.M., Gu, J., Chen, Y. and Reddy, R. (1998) Adenylation of small RNAs in human cells. Development of a cell-free system for accurate adenylation on the 3'-end of human signal recognition particle RNA. *J. Biol. Chem.*, **273**, 6853–6859.
53. Intine, R.V., Tenenbaum, S.A., Sakulich, A.L., Keene, J.D. and Maraia, R.J. (2003) Differential phosphorylation and subcellular localization of La RNPs associated with precursor tRNAs and translation-related mRNAs. *Mol. Cell*, **12**, 1301–1307.
54. Garriga, J., Bhattacharya, S., Calbo, J., Marshall, R.M., Truongcao, M., Haines, D.S. and Grana, X. (2003) CDK9 is constitutively expressed throughout the cell cycle, and its steady-state expression is independent of SKP2. *Mol. Cell Biol.*, **23**, 5165–5173.
55. Kiernan, R.E., Emiliani, S., Nakayama, K., Castro, A., Labbe, J.C., Lorca, T., Nakayama, K. and Benkirane, M. (2001) Interaction between cyclin T1 and SCF(SKP2) targets CDK9 for ubiquitination and degradation by the proteasome. *Mol. Cell Biol.*, **21**, 7956–7970.
56. Chiu, Y.-L., Cao, H., Jacque, J.M., Stevenson, M. and Rana, T.M. (2004) Inhibition of human immunodeficiency virus type 1 replication by RNA interference directed against human transcription elongation factor P-TEFb (CDK9/CyclinT1). *J. Virol.*, **78**, 2517–2529.
57. He, N., Pezda, A.C. and Zhou, Q. (2006) Modulation of a P-TEFb functional equilibrium for the global control of cell growth and differentiation. *Mol. Cell Biol.*, **26**, 7068–7076.
58. Rajasekhar, V.K. and Begemann, M. (2007) Concise review: roles of polycomb group proteins in development and disease: a stem cell perspective. *Stem Cells*, **25**, 2498–2510.
59. Santamaria, P. and Randsholt, N.B. (1995) Characterization of a region of the X chromosome of Drosophila including multi sex combs (mxc), a Polycomb group gene which also functions as a tumour suppressor. *Mol. Gen. Genet.*, **246**, 282–290.
60. Mori, Y., Sato, F., Selaru, F.M., Oлару, A., Perry, K., Kimos, M.C., Tamura, G., Matsubara, N., Wang, S., Xu, Y. et al. (2002) Instabilotyping reveals unique mutational spectra in microsatellite-unstable gastric cancers. *Cancer Res.*, **62**, 3641–3645.
61. Saget, O., Forquignon, F., Santamaria, P. and Randsholt, N.B. (1998) Needs and targets for the multi sex combs gene product in Drosophila melanogaster. *Genetics*, **149**, 1823–1838.
62. Remillieux-Leschelle, N., Santamaria, P. and Randsholt, N.B. (2002) Regulation of larval hematopoiesis in Drosophila melanogaster: a role for the multi sex combs gene. *Genetics*, **162**, 1259–1274.

CHAPITRE 6

DISCUSSION

Les protéines cellulaires agissent rarement seules, s'assemblant plutôt en des complexes afin d'exercer leur fonction (Spirin et Mirny, 2003; Rives et Galitski, 2003; Alberts, 1998). Ces complexes protéiques forment des machineries cellulaires spécifiques interagissant entre elles durant les divers processus biologiques, telle la transcription génique, la maturation de l'ARNm, la réplication ou encore la réparation de l'ADN. L'analyse de complexes protéiques et de réseaux d'interaction protéine-protéine ainsi que de leur dynamique de formation en fonction du temps et de l'état cellulaire est d'une importance capitale en recherche biologique. Le but de mon projet de doctorat était de mieux comprendre l'organisation de la machinerie transcriptionnelle de l'ARN Pol II dans la fraction nucléaire soluble de la cellule, avant son association à l'ADN ou suite au processus transcriptionnel. Nous nous étions fixés plusieurs objectifs, notamment de purifier par double chromatographie d'affinité des complexes protéiques solubles interagissant avec la machinerie transcriptionnelle dans les cellules humaines, d'établir des réseaux d'interactions protéine-protéine impliqués en transcription et enfin, d'identifier et de caractériser de nouveaux régulateurs transcriptionnels.

Bien que la machinerie de transcription de l'ARN Pol II soit étudiée depuis des décennies, nos résultats définissent plusieurs protéines additionnelles interagissant avec des composants clés de cette machinerie. Ainsi, le travail de thèse présenté ici a permis de mettre en évidence l'association des protéines XAB1-RPAPs à l'ARN Pol II ainsi que de MEPCE/BCDIN3 et de LARP7 au 7SK, un snARN régulateur du facteur d'élongation transcriptionnel P-TEFb. L'intégration de nos résultats a permis de mieux comprendre le rôle de ces polypeptides dont la fonction était à ce jour inconnue sur le processus transcriptionnel.

6.1 Étude protéomique de la machinerie de l'ARN polymérase II : purification de complexes protéiques et établissement de réseaux d'interactions protéine-protéine

Afin de purifier des complexes protéiques solubles interagissant avec la machinerie transcriptionnelle de l'ARN Pol II dans les cellules humaines, j'ai mis au point dans le laboratoire le système de purification TAP (*tandem affinity purification*), développé originalement pour isoler des complexes protéiques chez la levure (Rigaut et al., 1999; Puig et al., 2001). Grâce à cette technique, nous avons purifié par double chromatographie d'affinité à l'aide de protéines étiquetées (TAP-tag) exprimées à des niveaux physiologiques dans des lignées cellulaires HEK 293, des complexes protéiques contenant des FGT ainsi que l'ARN Pol II (figure 1). Cette stratégie de purification a été choisie pour plusieurs raisons. Tout d'abord, un système d'expression inductible et facilement contrôlable était important afin de minimiser la présence d'interactions protéiques non-spécifiques pouvant résulter de la surexpression des protéines étiquetées. Deuxièmement, la méthode TAP est faite à faible concentration ionique et en absence de détergents, ce qui permet la purification de complexes protéiques fonctionnels et le maintien de certaines interactions faibles ou sensibles aux conditions de purification. Bien que différentes étiquettes ou combinaison d'étiquettes aient été utilisées avec succès pour la purification de protéines à faible échelle (Terpe, 2003), des études protéomiques à large échelle ont surtout employé l'étiquette FLAG ou l'étiquette TAP (Ho et al., 2002; Krogan et al., 2006; Gavin et al., 2006; Gavin et al., 2002). Toutefois, lorsque comparée à la méthode FLAG, la purification par TAP a permis d'obtenir une meilleure efficacité pour la purification de protéines présentes à faible concentration dans un extrait cellulaire (Puig et al., 2001). Finalement, soulignons que cette méthode de purification de protéines étiquetées a permis l'étude de mutants de l'ARN Pol II humaine (travail réalisé par Marie-France Langelier). De plus, les protéines étiquetées avec TAP exprimées à des niveaux physiologiques ont aussi servi à localiser la machinerie transcriptionnelle le long de gènes étudiés à l'aide de la technique de CHIP (figure 1) (travail effectué par Marilena Cojocaru).

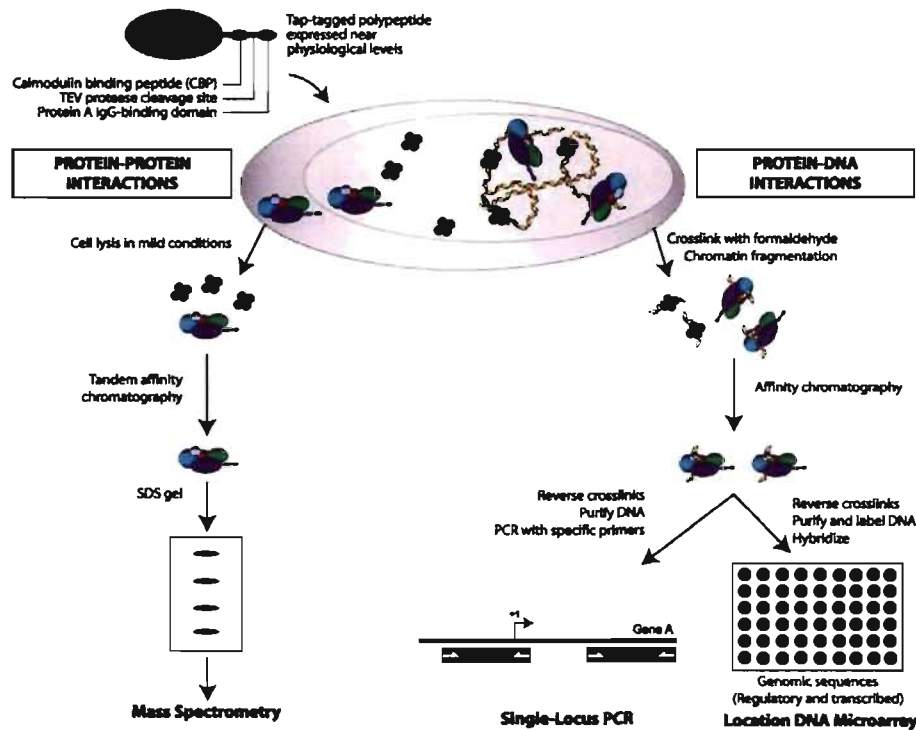


Figure 1 : Procédure générale pour définir des réseaux d'interactions protéine-protéine et protéine-ADN. À l'aide d'un système d'expression inductible, un polypeptide d'intérêt possédant une étiquette TAP est exprimé à des niveaux physiologiques dans les cellules de mammifères. L'étiquette TAP est composée de deux copies du domaine de liaison à l'immunoglobuline G (IgG) de la protéine A et d'un peptide de liaison à la calmoduline (CBP), avec un site de clivage pour la protéase TEV (*tobacco etch virus*) entre ces deux étiquettes d'affinité. À gauche, pour l'analyse de complexes protéiques, les cellules sont lysées dans des conditions douces et la fraction soluble est soumise à deux étapes successives de purification grâce aux billes d'IgG et de calmoduline. L'éluât final est analysé sur des gels de SDS, les bandes excisées et les polypeptides identifiés par spectrométrie de masse. À droite, pour des analyses de complexes protéine-ADN par ChIP, les cellules sont traitées au formaldéhyde, soniquées afin de fragmenter la chromatine et les fragments d'ADN couplés aux protéines sont alors soumis à la chromatographie d'affinité anti-TAP (TAP-ChIP). Les fragments d'ADN enrichis sont analysés par amplification PCR de gènes sélectionnés à l'aide d'amorces spécifiques ou encore marqués et hybridés sur de puces à ADN (ChIP-CHIP). Figure tirée de (Coulombe et al., 2004).

Le couplage entre la purification par chromatographie d'affinité et l'identification de polypeptides par spectrométrie de masse (MS) a grandement contribué à notre connaissance de la composition des complexes protéiques (Gingras et al., 2007). Le séquençage efficace des peptides dérivés par digestion protéolytique est obtenu grâce à

deux stratégies principales d'ionisation des ions peptidiques, à savoir ESI (*electrospray ionization*) et MALDI (*matrix-assisted laser desorption/ionization*) (Domon et Aebersold, 2006; Steen et Mann, 2004). Ces méthodes d'ionisation sont ensuite couplées à divers types de spectromètres de masse en tandem. La MS est présentement la méthode de choix pour le séquençage peptidique (Gingras et al., 2007). En effet, cette méthode est hautement sensible et rapide, permettant l'identification de centaines de peptides présents à de très faibles quantités (de l'ordre du femtomole) par lecture de spectre de masse. Facilement automatisée, la MS est compatible avec des stratégies d'identification protéique à grande échelle. De plus, cette méthode permet la caractérisation de modifications peptidiques, telles les modifications post-traductionnelles. Finalement, la MS peut être adaptée afin de mesurer de façon quantitative l'abondance peptidique. En parallèle, il est important de noter le développement d'outils informatiques permettant de sélectionner l'information à partir des spectres de masse et de donner une validation statistique quant à la précision des séquences déterminées par cette approche (Nesvizhskii et al., 2006).

Dans l'article présenté au chapitre 2 (Jeronimo et al., 2004), nous avons d'abord purifié un complexe protéique comprenant l'ARN Pol II, les FGT TFIIB et TFIIF, Fcp1, et RPAP1, un polypeptide à fonction inconnue. Nous avons montré que la présence de l'étiquette TAP à l'extrémité C-terminal des sous-unités de ce complexe n'altère pas son activité. Le complexe ARN Pol II-TAP est actif *in vitro*. Ainsi, ce complexe est capable de s'associer avec le promoteur majeur tardif de l'adénovirus pour former un complexe de pré-initiation transcriptionnel en compagnie des FGT, d'initier la transcription et de participer dans l'élongation transcriptionnelle. De plus, nous avons été en mesure de démontrer grâce à la technique de CHIP que l'ARN Pol II-TAP peut être recrutée à des promoteurs activement transcrits *in vivo*, indiquant que le complexe ARN Pol II purifié est biologiquement actif. D'autre part, nos résultats ont montré la présence de sous-unités régulatrices associées de manière stable à l'ARN Pol II dans les cellules humaines. Une fonction possible pour Fcp1 dans le complexe ARN Pol II-TAP serait de maintenir le CTD dans un état hypophosphorylé, tel que proposé auparavant chez la levure (Kimura et al., 2002). En effet, l'ARN Pol II est supposée être sous forme soluble afin d'être recrutée

aux promoteurs suite aux signaux d'activation transcriptionnelle (Svejstrup, 2004). Pour supporter cette hypothèse, le CTD de Rpb1 dans le complexe ARN Pol II-TAP est trouvé dans l'état hypophosphorylé. Quant à TFIIB et TFIIF, il fût montré préalablement que ces facteurs régulent le recrutement de l'ARN Pol II au promoteur génique (Conaway et al., 1991; Maldonado et al., 1990; Sopta et al., 1985). De plus, ces deux facteurs peuvent également lier Fcp1 et réguler son activité phosphatase (Chambers et al., 1995). Finalement, des analyses bioinformatiques ont permis de révéler que le polypeptide RPAP1 identifié dans le complexe ARN Pol II-TAP présente des régions d'homologie avec Ydr527wp/RBA50, une protéine essentielle pour la viabilité cellulaire chez *S. cerevisiae*. Ydr527wp/RBA50 a été trouvé en complexe avec des sous-unités de l'ARN Pol II (Giaever et al., 2002; Ito et al., 2001). De plus, en collaboration avec le laboratoire du Dr. Tim Hughes et grâce à des expériences de *silencing* utilisant une collection de souches mutantes dont les gènes sont sous le contrôle du promoteur modulable *Tet*, nous avons montré que chez la levure le *silencing* du gène *YDR527* a le même effet sur l'expression génique globale que celui de la sous-unité Rpb11 de l'ARN Pol II. Ensemble, ces résultats supportent la notion que l'association de RPAP1 à l'ARN Pol II est physiologiquement importante. Toutefois, il nous faudra démontrer le rôle fonctionnel de cette interaction. La fonction proposée de RPAP1 en transcription par l'ARN Pol II sera davantage discutée dans la section suivante (6.2).

Après avoir montré que notre stratégie de purification de complexes protéiques couplée à la MS permettait d'isoler des complexes fonctionnels contenant l'ARN Pol II, nous nous sommes intéressés à définir des réseaux régulateurs transcriptionnels. Par le passé, plusieurs méthodes expérimentales ont été développées afin d'identifier des réseaux d'interactions protéine-protéine, incluant la méthode deux-hybrides et l'approche de chromatographie d'affinité couplée à la MS. Ces méthodes ont été utilisées pour caractériser des réseaux d'interactions protéiques, et dans certains cas, pour définir la fonction de protéines sur la base du concept « coupables par association » chez plusieurs organismes modèles, incluant *S. cerevisiae*, *D. melanogaster* et *C. elegans* (Giot et al., 2003; Ho et al., 2002; Uetz et al., 2000; Ito et al., 2001; Gavin et al., 2006; Gavin et al., 2002; Krogan et al., 2006; Butland et al., 2005; Li et al., 2004b). Récemment, ces

Tel que discuté précédemment, notre procédure de purification a été choisie afin de préserver l'intégrité des complexes purifiés. La stratégie de purification par chromatographie d'affinité-MS permet théoriquement d'isoler tous les complexes protéiques contenant le polypeptide étiqueté. Toutefois, cette méthode ne permet pas de déterminer directement l'abondance des complexes purifiés (Gingras et al., 2007). De plus, la technique de chromatographie d'affinité-MS ne génère qu'une liste de protéines dans l'échantillon étudié, sans révéler nécessairement la composition de chacun des complexes protéiques individuellement. L'ensemble des données obtenues pour chaque expérience de chromatographie d'affinité-MS représente donc une moyenne des partenaires de liaison et des complexes protéiques. Advenant le cas où la protéine étiquetée est une composante de multiples complexes alternatifs, une seule analyse de chromatographie d'affinité-MS ne peut pas être utilisée afin de déterminer toute la multiplicité des cibles possibles pour cet appât. De cette façon, la méthode utilisée ne mesure que l'appartenance moyenne d'une protéine dans les cellules d'une population donnée, ne rendant pas compte de l'hétérogénéité de la dynamique de la protéine étudiée au sein d'une même population. Ceci est une limitation significative car les protéines peuvent jouer des rôles très distincts dépendamment de leur appartenance à différents types de complexes. D'autre part, la structure des complexes multi-protéiques ne peut être révélée qu'indirectement par l'approche de chromatographie d'affinité-MS (Gingras et al., 2007). Toutefois, l'analyse de la composition d'un complexe protéique possédant des propriétés biochimiques bien définies peut être utilisée pour révéler directement la composition d'un complexe en particulier. Ainsi, le fractionnement biochimique, tel par exemple le fractionnement selon le poids moléculaire via la filtration sur gel ou les gradients de densité a été couramment utilisé pour la séparation et l'enrichissement de complexes protéiques. Des études montrent qu'en combinant cette technique de fractionnement biochimique à la chromatographie d'affinité, une meilleure caractérisation de complexes protéiques peut être obtenue (Mueller et Jaehning, 2002; Lindstrom et al., 2003).

Le haut niveau de sensibilité de la méthode MS pour l'identification de protéines exige le développement de méthodes bioinformatiques discriminant les interactions

spécifiques de celles qui sont sporadiques (Patil et Nakamura, 2005; Krogan et al., 2006; Gavin et al., 2006). Dans notre cas, nous avons établi une collaboration avec le Dr Mathieu Blanchette (*McGill Centre for Bioinformatics*) pour développer un algorithme capable de sélectionner des interactions protéiques avec un haut niveau de confiance (Jeronimo et al., 2007). Trente-deux protéines étiquetées ont permis de générer un réseau de 805 interactions protéiques validées bioinformatiquement. L'article de Cojocaru *et al.*, 2008 montre que les facteurs étiquetés que nous avons isolés et testés sont actifs dans des réactions de transcription *in vitro*. De plus, les études de ChIP-qPCR ont montré que ces facteurs sont localisés au promoteur et dans la région codante de gènes humains étudiés *in vivo*, indiquant que des complexes transcriptionnels fonctionnels ont été purifiés (Cojocaru et al., 2008). Ce réseau d'interactions révèle l'interconnexion des divers complexes protéiques formant le cœur de la machinerie impliquée dans l'interprétation du génome (figure 3).

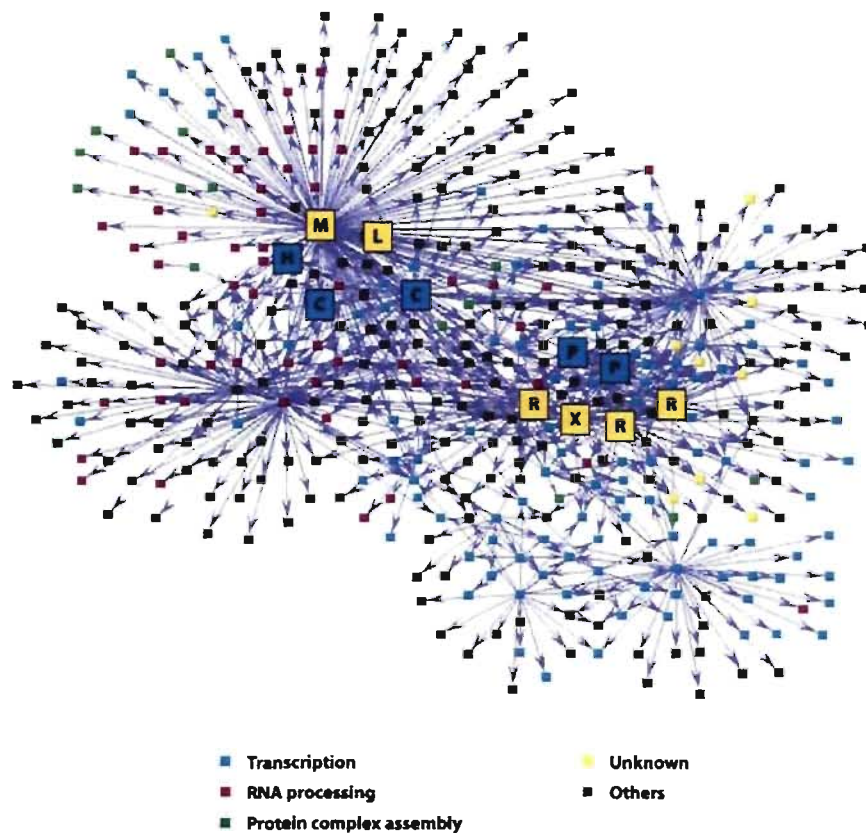


Figure 3 : Carte d'un réseau à haute densité d'interactions protéine-protéine impliquant des facteurs de transcription de l'ARN Pol II et de maturation de l'ARNm. Les protéines étiquetées (appâts) et leurs partenaires d'interaction co-purifiés (cibles) sont représentés par des carrés colorés selon le code indiqué et interconnectés par des flèches. M, MEPCE/BCDIN3; L, LARP7; H, HEXIM1; C, les sous-unités de P-TEFb (CDK9 et CCNT1/Cycline T1); P, les sous-unités de l'ARN polymérase II (Rpb2 and Rpb11); R, RPAP; X, XAB1. Figure modifiée de (Coulombe et al., 2008).

En examinant la topologie du réseau obtenu, il est aisé de voir que plusieurs interactions relient les machineries de transcription (figure 3, carrés bleus) et de maturation de l'ARNm (figure 3, carrés roses), confirmant le couplage bien caractérisé entre ces deux processus biologiques (Proudfoot, 2004; Kornbliht et al., 2004; Maniatis et Reed, 2002; Calvo et Manley, 2003). Par exemple, le FGT TAF10 est lié aux facteurs de maturation de l'ARNm PRPF31 et DHX15 alors que le facteur d'élongation transcriptionnelle TFIIS (TCEA1) a été trouvé en association avec les facteurs de maturation de l'ARNm SF3B1 et SF3B2 (tableau I). D'autre part, le facteur HNRPA1 impliqué dans la maturation de l'ARNm a été trouvé associé aux facteurs CDK9, Cyclin T1 (CCNT1), TFIIA (GTF2A1) et TFIIE (GTF2E2) de la machinerie transcriptionnelle (tableau I).

Au cours de l'analyse du réseau protéique présenté au chapitre 3, nous avons été surpris de constater le nombre significatif de composantes impliquées dans la régulation de la formation de complexes protéiques (tableau I; figure 3, carrés verts). Le terme GO (*Gene Ontology*) de « formation de complexes protéiques » est spécifiquement utilisé ici afin de décrire l'ensemble des étapes depuis la synthèse des polypeptides jusqu'à leur assemblage en des complexes multi-protéiques. Ce terme général inclus par exemple le repliement, l'assemblage et le transport des protéines. Ces partenaires protéiques définissent une nouvelle classe de facteurs régulateurs ciblant directement les machineries de transcription et de maturation de l'ARNm avant leur recrutement à des loci génomiques ou après leur relâchement de la matrice d'ADN, permettant leur recyclage. Nous pensons que ces protéines n'ont pas été identifiées auparavant parce qu'elles ne régulent pas principalement ces machineries lors d'un état transcriptionnel actif, bien que cette possibilité doit encore être testée directement.

Tableau I. Liste des protéines étiquetées et des partenaires d'interaction co-purifiés selon leur GO (*Gene Ontology*)

GO (<i>Gene Ontology</i>)	Protéines
■ Transcription	ACTB, ARID2, ARID3B, C19orf2, CAND1, CCNH, CCNT1, CCNT2, CCT6A, CDC73, CDK7, CDK9, CNOT1, CRSP3, CRSP6, CSDA, CTD1P1, CUTL1, DMAP1, DNM2, ERCC3, ERCC5, GCN5L2, GTF2A1, GTF2A2, GTF2B, GTF2E1, GTF2E2, GTF2F1, GTF2F2, GTF2H1, GTF2H2, GTF2H3, GTF2H4, GTF2I, GTF3C3, GTF3C5, HEXIM1, HEXIM2, ILF2, ILF3, KHDRBS2, KHDRBS3, MED12, MED4, MTA2, PARP1, PCAF, PHF2, PHF8, PNN, POLR1A, POLR1C, POLR1D, POLR2A, POLR2B, POLR2C, POLR2D, POLR2E, POLR2G, POLR2H, POLR2J, POLR3A, POLR3B, POLR3C, POLR3D, POLR3E, POLR3G, POLR3H, PURA, RDBP, RUVBL1, RUVBL2, RXRG, SAFB, SMARCA4, SMARCC1, SMARCC2, SNAPC4, SUPT6H, SUPT7L, TADA1L, TADA3L, TAF1, TAF10, TAF13, TAF1L, TAF2, TAF3, TAF4, TAF4B, TAF5, TAF5L, TAF6, TAF6L, TAF7, TAF7L, TAF9, TAF9L, TBP, TBPL1, TBPL2, TCEA1, TCEA2, THRAP4, TRAP1, TRFP, TRIM28, TRIP12, TRIP13, TRRAP, WHSC2, YEATS4
■ Maturation de l'ARNm	ARF6, CDC2L1, DDX17, DDX48, DHX15, EFTUD2, EXOSC2, EXOSC5, FBL, HNRPA0, HNRPA1, HNRPA2B1, HNRPAB, HNRPC, HNRPD, HNRPDL, HNRPH1, HNRPH3, HNRPK, HNRPM, HNRPR, HNRPU, IGF2BP3, KHDRBS1, KHSRP, LSM2, LSM3, LSM4, LSM7, NCBP1, NHP2L1, NOL5A, PABPC1, PABPC4, PRPF3, PRPF31, PRPF4, PRPF8, RY1, SART1, SART3, SF3B1, SF3B2, SNRPA1, SNRPB, SNRPD1, SNRPD2, SNRPD3, SNRPF, SNW1, SYNCRIP, TARDBP, THOC4, U2AF1, USP39, YBX1
■ Formation de complexes protéiques	ASCC3L1, BAT1, DDX23, DSG1, HCFC1, KPNA3, MNAT1, PPIH, PRPF6, SET, SF3A1, SF3B3, SFRS9, SNRPG, TCOF1
■ Inconnu	BCDN3, C12orf45, C1orf55, C1orf60, C1orf73, C1orf82, C8orf41, FLJ21908, LARP7, LOC116143, MGC14560, NOP17, PDRG1, RPAP1, XAB1

Une fraction significative de protéines codées par le génome humain demeure non-caractérisée et leurs fonctions précises inconnues. En identifiant les partenaires d'interaction protéiques, il nous a été possible de prédire les fonctions présumées de protéines nouvelles non-caractérisées (tableau I ; figure 3, carrés jaunes). Dans ce réseau, environ 2% des 805 protéines identifiées n'a pas de fonction connue. Par la suite, nous avons cherché à savoir quel pouvait être les rôles de certains de ces polypeptides, incluant

XAB1-RPAPs, BCDIN3 et LARP7, dans le processus transcriptionnel de l'ARN Pol II (figure 3, carrés jaunes). Les sections 6.2 et 6.3 traitent de la caractérisation de nouveaux régulateurs transcriptionnels identifiés dans cette étude.

6.2 Complexes protéiques contenant l'ARN Pol II en association avec des régulateurs impliqués dans l'assemblage des protéines : les protéines XAB1-RPAPs

Notre réseau d'interactions révèle que dans le compartiment nucléaire soluble de la cellule un groupe de quatre protéines, notamment RPAP1 (Jeronimo et al., 2004), XAB1, C1ORF82 et FLJ21908, est trouvé associé à l'ARN Pol II. Grâce à des expériences de filtration sur gel, nous avons confirmé que ces protéines co-fractionnent avec l'enzyme. Vu leur association physique à l'ARN Pol II, les polypeptides à fonction inconnue C1ORF82 et FLJ21908 ont été nommés respectivement RPAP2 et RPAP3. Quant à XAB1 (*XPA binding protein 1*), comme son nom l'indique, a été identifiée précédemment en association avec le facteur de réparation XPA (Nitta et al., 2000). Bien qu'une étude a démontré que XAB1 participe dans la régulation transcriptionnelle au niveau des promoteurs méthylés (Lembo et al., 2003), son mécanisme d'action demeure inconnu. Nos résultats montrent que les protéines XAB1-RPAPs sont également associées avec de régulateurs transcriptionnels récemment identifiés. Ainsi, dans cette partie du réseau nous avons pu retrouver le complexe Médiateur en compagnie de son nouveau régulateur positif appelé GRINL1A (Hu et al., 2006) et le complexe Intégrateur, un complexe impliqué dans la maturation des snARN synthétisés par l'ARN Pol II (Baillat et al., 2005). D'autre part, dans ce réseau les protéines XAB1-RPAPs sont placées entre l'ARN Pol II et un groupe de facteurs agissant comme des chaperons moléculaires impliqués dans l'assemblage de complexes protéiques, incluant plusieurs préfoldines (Gstaiger et al., 2003; Vainberg et al., 1998; Zhao et al., 2005) et les deux AAA+ ATPases RUVBL1 et RUVBL2 (Kanemaki et al., 1999; Ikura et al., 2000).

Afin de mieux caractériser ces protéines à fonction inconnue, nous avons effectué une analyse bioinformatique ciblant leurs domaines protéiques. Les résultats obtenus révèlent que ces protéines contiennent elles aussi des motifs caractéristiques de chaperons et de protéines d'échafaudage. RPAP1 possède un domaine ARM, lequel est un élément structural favorisant les interactions protéine-protéine (Huber et al., 1997). Pour ce qui est de RPAP3, cette protéine possède des motifs TPR connus pour être présents chez des cofacteurs de chaperons (Smith, 2004). Enfin, XAB1 appartient à une famille de protéines « *ATP bind 1* » possédant, comme son nom l'indique, un domaine de liaison à l'ATP, lequel pourrait apporter l'énergie nécessaire aux changements conformationnels impliqués dans la formation des complexes protéiques. Il est intéressant de noter que les protéines XAB1-RPAPs sont également associées avec divers autres polypeptides à fonction inconnue, tels MGC14560, LOC116143, NOP17 et PRDG1. La protéine MGC14560 est un autre membre de la famille « *ATP bind 1* » alors que LOC116143 possède un domaine WD40 pouvant jouer un rôle dans la coordination de l'assemblage de complexes multi-protéiques (Smith et al., 1999). D'après les résultats obtenus, il est aisé de constater que l'enzyme ARN Pol II est associée avec un nombre de facteurs connus ou prédits pour réguler la formation de complexes de protéines.

Connaissant les homologues des protéines XAB1-RPAPs, nous avons confirmé qu'un certain nombre d'interactions parmi ces protéines et l'ARN Pol II était également observé chez la levure *S. cerevisiae*. De plus, la suppression de l'expression des homologues de XAB1 et MGC14560, deux gènes essentiels pour la viabilité cellulaire chez la levure, a induit le même effet global sur l'expression génique que celui de la sous-unité Rpb11 de l'ARN Pol II ou l'homologue de RPAP1. Globalement, ces résultats montrent que les protéines XAB1-RPAPs sont physiquement et fonctionnellement associées à l'ARN Pol II, formant une interface entre l'enzyme et des protéines chaperons et d'échafaudage. À la lumière de ceci, nous proposons un rôle pour les protéines XAB1-RPAPs dans l'assemblage de complexes protéiques formant la machinerie de synthèse de l'ARNm. Des expériences sont en cours au laboratoire afin de déterminer la fonction précise de ces nouvelles protéines régulatrices dans le processus transcriptionnel.

6.3 Complexe ribonucléoprotéique 7SK snARN•HEXIM•P-TEFb : rôle des protéines BCDIN3/MEPCE et LARP7

La méthode de chromatographie d'affinité utilisée dans cette étude est également efficace pour l'identification de composants ARN au sein des complexes purifiés, tel que montré chez la levure (Krogan et al., 2004). Les articles présentés aux chapitres 5 et 6 traitent du rôle des protéines à fonction inconnue BCDIN3 et LARP7 trouvées en association avec le facteur d'élongation P-TEFb ainsi que 7SK, le snARN impliqué dans la régulation transcriptionnelle de P-TEFb (Peterlin et Price, 2006; Zhou et Yik, 2006).

6.3.1 Identification de l'enzyme de coiffage de 7SK

Tel que discuté précédemment (voir sections 1.1.3.1 et 1.2.1.3.1 de l'introduction), l'activité de P-TEFb est hautement régulée dans la cellule. Notre réseau d'interactions protéiques a révélé que, dans la fraction nucléaire soluble, en plus de ses régulateurs positif (la protéine Brd4) et négatifs (les protéines HEXIM et le snARN 7SK), P-TEFb est associé avec diverses autres protéines, incluant des facteurs impliqués dans la maturation de l'ARNm, comme par exemple hnRNPA1 et SART3, ainsi que BCDIN3, un polypeptide à fonction inconnue possédant un domaine caractéristique des méthyltransférases (Jeronimo et al., 2007). Nos résultats montrent également que les snARN 7SK et U6, deux transcrits de l'ARN Pol III, font partie du complexe purifié BCDIN3-TAP. Toutefois, U2, un snARN synthétisé par l'ARN Pol II, n'a pas été trouvé associé avec BCDIN3, soulignant la spécificité des composants ARN identifiés par la méthode TAP. Il était connu dans la littérature que les protéines HEXIM lient le 7SK via leurs motifs de liaison à l'ARN riches en arginines (Byers et al., 2005; Yik et al., 2004; Yik et al., 2005). Quant à SART3, ce facteur possède deux motifs de reconnaissance de l'ARN connus pour lier directement le snARN U6 (Bell et al., 2002). Il est intéressant de noter que SART3 interagit avec d'autres composants de la snRNP de U6, suggérant que ce facteur agit en tant qu'interface d'échafaudage pour l'assemblage de la snRNP U4/U6

(Medenbach et al., 2004). Ces résultats ont d'ailleurs permis de proposer un rôle de chaperon d'ARN pour SART3.

Afin de déterminer la fonction de cette nouvelle méthyltransférase, nous nous sommes premièrement intéressés à identifier ses substrats spécifiques. Nous avons alors recherché des substrats parmi les membres de la snRNP de 7SK. Notre analyse bioinformatique a révélé que BCDIN3 possède un domaine méthylase à ARN. Cette observation nous a amené à penser que BCDIN3 pouvait méthyler le snARN 7SK. Pour essayer de vérifier cette hypothèse, une protéine recombinante BCDIN3-His a été produite. Par la suite, afin de montrer que BCDIN3 était impliquée dans la méthylation de 7SK, nous avons mis au point un essai de méthylation de l'ARN *in vitro*. Les résultats obtenus montrent que BCDIN3 méthyle le snARN 7SK. De plus, dans l'essai de méthylation en utilisant un extrait cellulaire déplété pour BCDIN3 endogène, nous avons confirmé que BCDIN3 est une méthyltransférase de 7SK dans les cellules humaines.

Après avoir trouvé cette nouvelle enzyme capable de méthyler le snARN 7SK, nous avons voulu savoir à quel endroit le groupement méthyle était ajouté sur le snARN. D'après la littérature, les transcrits de l'ARN Pol III, tel que 7SK, sont modifiés post-transcriptionnellement par l'addition d'une structure coiffe particulière, soit un monométhyle sur le phosphate gamma à leur extrémité 5' (Shumyatsky et al., 1990; Singh et Reddy, 1989; Gupta et al., 1990). Toutefois, jusqu'au moment de notre étude, l'enzyme de coiffage de ces transcrits n'était pas connue. BCDIN3 est alors devenue un candidat pour être l'enzyme de coiffage méthylphosphate. Afin de tester cette hypothèse, nous avons mis au point deux essais enzymatiques. Dans le premier, le snARN 7SK utilisé dans l'essai de méthylation a été traité avec une phosphatase, enlevant spécifiquement les groupements phosphates à l'extrémité 5' de l'ARN. Dans le second, le 7SK méthylé par BCDIN3 a été déphosphorylé par une pyrophosphatase, une enzyme capable d'enlever les phosphates gamma et beta au niveau de l'extrémité 5' de l'ARN (Shinshi et al., 1976). Les études réalisées ici ont permis de conclure que BCDIN3 est responsable de l'ajout de la coiffe en 5' de 7SK. BCDIN3 a alors été nommée *methylphosphate capping enzyme* (MEPCE). Jusqu'à présent, la structure coiffe de

monométhyle sur le phosphate gamma a été identifié au sein de trois autres petits ARN transcrits par l'ARN Pol III, notamment U6, B2 chez la souris et U3 chez la plante (Shumyatsky et al., 1990; Shimba et al., 1992; Liu et al., 1992). Par ailleurs, nous avons montré que BCDIN3/MEPCE est également capable d'ajouter la coiffe sur les snARN U6 et B2 (données non-montrées). Soulignons que ces résultats étaient attendus vu que les travaux de l'équipe de R. Reddy avaient suggéré la présence d'une enzyme de coiffage commune dans les cellules pour ces snARN (Shimba et Reddy, 1994).

6.3.2 Rôle des protéines BCDIN3/MEPCE et LARP7 dans la stabilité du snARN 7SK

Les études précédentes ont montré que la structure coiffe sur le phosphate à la position gamma de 7SK est importante pour protéger l'ARN d'une dégradation exonucléolytique (Shumyatsky et al., 1990). Pour étudier plus à fond le rôle de BCDIN3/MEPCE dans la stabilité du 7SK, des essais d'interférence par l'ARN double brin (siARN) ont été utilisés. Grâce à des Northern blots dirigés contre le snARN 7SK, nous avons été en mesure de démontrer qu'une réduction de l'expression de BCDIN3/MEPCE entraîne une diminution des niveaux cellulaires de 7SK. Ces résultats ont permis de proposer que BCDIN3/MEPCE stabilise le 7SK dans les cellules via l'addition d'une structure coiffe à son extrémité 5'.

Dans le même réseau protéique contenant P-TEFb, nous avons aussi identifié LARP7, un polypeptide à fonction inconnue appartenant à la famille de la protéine La. Comme tous les membres de la famille La, LARP7 possède un domaine LA et des motifs de reconnaissance de l'ARN, nommés RRM (*RNA recognition motifs*). Des études antérieures ont montré que La est une protéine multifonctionnelle connue pour lier l'ARN et jouer des rôles clés dans le métabolisme de l'ARN (Maraia et Bayfield, 2006; Wolin et Cedervall, 2002). L'article présenté au chapitre 5 montre que LARP7 est une composante stable de la ribonucléoprotéine 7SK•HEXIM•MEPCE•P-TEFb. Ceci représente une différence entre LARP7 et la protéine La étant donné que cette dernière n'est pas présente

dans la forme finale mature de la majorité des snRNP (Maraia et Bayfield, 2006; Wolin et Cedervall, 2002). Comme tous les transcrits de l'ARN Pol III, 7SK est lié par la protéine La au niveau de son extrémité 3' riche en uridine (Rinke et Steitz, 1982; Chambers et al., 1983). Toutefois, ces études démontrent que seulement 2-5% du 7SK total est associé avec la protéine La. Ceci expliquerait nos résultats montrant qu'environ 90% de la snRNP 7SK est immunoprécipitée grâce à des anticorps spécifiques de LARP7. À la lumière de ceci, nous proposons que LARP7 remplace la protéine La lors du processus de maturation de la snRNP de 7SK.

Par des expériences de siARN, nous avons pu observer que LARP7 est impliquée dans la stabilité du 7SK. Les résultats présentés ici montrent que LARP7 protège le snARN 7SK de la dégradation, de manière similaire à la protéine La envers d'autres ARN. En effet, il est bien établi que le rôle de la protéine La consiste à protéger l'extrémité 3' des transcrits de l'ARN Pol III d'une dégradation exonucléolytique (Maraia et Bayfield, 2006; Wolin et Cedervall, 2002). Comme LARP7 possède les domaines La et RRM1 de liaison à l'ARN, LARP7 pourrait donc lier le 7SK au niveau de ses extrémités 3' et 5'. Il est intéressant d'indiquer que la liaison de la protéine La aux ARN cibles est un processus hautement régulé et qui peut dépendre par exemple des modifications aux extrémités 3' et 5' de ces ARN. Ainsi, des études ont démontré que la structure coiffe sur le phosphate gamma à l'extrémité 5' de U6 et B2 réduit l'affinité de La pour ces ARN alors que l'adénylation de l'extrémité 3' de ces transcrits permet le relâchement complet de La (Bhattacharya et al., 2002; Perumal et Reddy, 2002). De cette façon, on peut envisager que l'ajout de la coiffe sur 7SK par MEPCE induit le relâchement de la protéine La et l'association subséquente de LARP7. Alternativement, il est possible que la coiffe en 5' de 7SK affecte la façon dont 7SK et LARP7 interagissent, ce qui pourrait avoir un impact sur l'association ou la dissociation de P-TEFb et des protéines HEXIM.

6.3.3 Rôle des protéines BCDIN3/MEPCE et LARP7 dans la régulation de l'activité de P-TEFb

Pour étudier plus en détails les rôles de BCDIN3/MEPCE et LARP7 dans la régulation de l'activité du facteur P-TEFb, nous avons établi une collaboration avec le laboratoire du Dr David Price (*University of Iowa*), un des pionniers de l'étude de la fonction de ce facteur dans le processus transcriptionnel. Rappelons que P-TEFb est présent dans les cellules humaines sous deux formes de complexes : le petit complexe contient P-TEFb libre alors que le deuxième complexe, de grande taille, est catalytiquement inactif et il comprend un P-TEFb associé au snARN 7SK et aux protéines HEXIM (Nguyen et al., 2001; Yang et al., 2001). Nos analyses de sédimentation sur gradient de glycérol des extraits cellulaires traités avec des inhibiteurs de P-TEFb suggèrent une réorganisation de la snRNP de 7SK selon laquelle les protéines LARP7 et BCDIN3/MEPCE ainsi que le snARN 7SK demeurent associés après le relâchement de P-TEFb et des protéines HEXIM.

Par la suite, puisqu'une réduction de l'expression de BCDIN3/MEPCE et LARP7 résulte en une diminution des niveaux cellulaires de 7SK, nous avons voulu déterminer l'effet de ceci sur le ratio des formes cellulaires de P-TEFb et sur l'activité de ce facteur. Pour ce faire, des expériences de sédimentation sur gradient de glycérol ont été effectuées dans les cellules traitées avec siARN pour LARP7 ou BCDIN3/MEPCE. D'après les résultats obtenus, une diminution de l'expression de LARP7 et BCDIN3/MEPCE entraîne une augmentation de la quantité de P-TEFb libre, la forme de P-TEFb actif. Toutefois, l'effet observé pour BCDIN3/MEPCE était moins prononcé. Étant donné que le relâchement de P-TEFb est un processus hautement régulé, nous avons proposé comme hypothèse que l'augmentation de P-TEFb libre induite par la perte des protéines LARP7 et BCDIN3/MEPCE, et par conséquent la perte de 7SK, affecterait la transcription. Auparavant, des études ont démontré que la dégradation de 7SK ou la réduction de l'expression de HEXIM1 augmente la transactivation par Tat du VIH-1 (Yang et al., 2001; Yik et al., 2003), un processus connu pour être dépendant de P-TEFb. Afin de tester cette hypothèse, nous avons collaboré avec le laboratoire du Dr Éric Cohen (Institut

de recherches cliniques de Montréal) qui possède un système de transfection transitoire d'un gène rapporteur lui permettant d'évaluer la transactivation par Tat. Ces expériences nous ont permis de démontrer qu'une diminution de LARP7 induit une augmentation de la transactivation par Tat du VIH-1. Les études réalisées ici ont permis de conclure que LARP7 agit en stabilisant et en maintenant l'intégrité de la snRNP de 7SK. Une diminution de l'expression de LARP7 entraîne une réduction de 7SK et une augmentation de P-TEFb actif.

L'ensemble de ces expériences nous ont permis de proposer un modèle de la participation de MEPCE et LARP7 à la stabilisation du snARN 7SK, contribuant au maintien de P-TEFb dans un état transcriptionnel inactif (figure 4).

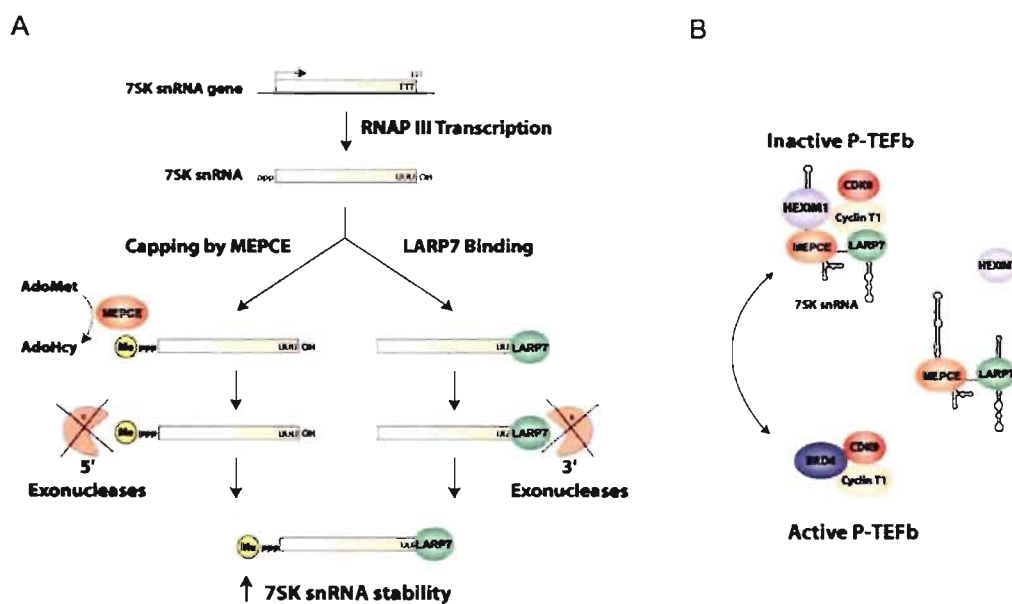


Figure 4 : Modèle de la participation de MEPCE et LARP7 à la stabilisation du snARN 7SK, contribuant au maintien de P-TEFb dans un état transcriptionnel inactif. (A) Le snARN 7SK est un transcrit de 331 nucléotides synthétisé par l'ARN Pol III. L'enzyme de coiffage MEPCE a été montrée pour stabiliser le 7SK via l'addition post-transcriptionnelle d'un groupement méthyle sur le phosphate gamma à son extrémité 5' (Jeronimo et al., 2007). La protéine LARP7 possédant des domaines de liaison à l'ARN LA et RRM, est une composante de la ribonucléoprotéine (snRNP) 7SK•HEXIM•P-TEFb•MEPCE et participe avec MEPCE à la stabilisation de 7SK dans les cellules humaines (Krueger et al., 2008). Étant donné le rôle bien

établi de la protéine La dans la protection de la dégradation exonucléolytique à l'extrémité 3' des transcrits par l'ARN Pol III, nous proposons comme hypothèse que LARP7 lie l'extrémité 3' du 7SK, lui procurant de la stabilité. (B) MEPCE et LARP7 font partie de la snRNP 7SK, formant le complexe P-TEFb transcriptionnellement inactif. Le traitement avec des inhibiteurs de P-TEFb induit une réorganisation de la snRNP 7SK, selon laquelle les protéines MEPCE et LARP7 demeurent associées au 7SK dans un complexe ayant perdu HEXIM et P-TEFb actif.

CONCLUSIONS ET PERSPECTIVES

Le but principal de cette thèse était de mieux comprendre l'organisation de la machinerie transcriptionnelle de l'ARN Pol II dans la fraction nucléaire soluble, avant son association à l'ADN ou après son relâchement suite à la transcription. Le premier objectif était de purifier des complexes protéiques solubles interagissant avec la machinerie de l'ARN Pol II afin d'établir des réseaux d'interactions protéine-protéine impliqués dans le processus transcriptionnel. Grâce à la technique de double chromatographie d'affinité couplée à la MS et à des analyses bioinformatiques, nous avons mis en évidence un complexe protéique fonctionnel *in vitro* et *in vivo* comprenant l'ARN Pol II, TFIIB, TFIIF et Fcp1. Au cours de ces études, nous avons pu étiqueter une trentaine de protéines, ce qui a permis de générer un réseau de plus de 800 interactions protéiques validées bioinformatiquement. De plus, comparativement aux analyses biochimiques réalisées auparavant, cette technique de double chromatographie d'affinité couplée à la MS nous a permis d'identifier plusieurs protéines à fonction inconnue, incluant XAB1-RPAPs, BCDIN3 et LARP7.

L'ensemble des résultats obtenus dans notre étude protéomique a permis d'avoir une vision beaucoup plus claire et complète de la composition et l'organisation de la machinerie transcriptionnelle de l'ARN Pol II dans la fraction soluble des cellules HEK 293. Afin de mieux connaître le réseau d'interactions protéiques de la machinerie de l'ARN Pol II dans ces cellules, il serait intéressant d'augmenter le nombre et la variété des facteurs étiquetés. Ceci nous permettrait d'obtenir un réseau protéique impliquant cette machinerie aussi complet que possible. De plus, en ajoutant des étiquettes à des facteurs jouant un rôle, par exemple, dans le remodelage de la chromatine, la réplication et la réparation de l'ADN, il serait possible de relier le réseau actuel à d'autres processus biologiques. D'autre part, il paraît aussi essentiel d'investiguer la composition et l'organisation de ces machineries moléculaires dans la fraction nucléaire insoluble des cellules, contenant la chromatine et les facteurs liés à l'ADN, en utilisant des techniques de fractionnement biochimique. Dans la poursuite de ce projet, il serait aussi pertinent de déterminer la composition de ces réseaux protéiques dans différentes conditions de

croissance et dans d'autres lignées cellulaires, afin d'investiguer la plasticité des réseaux d'interactions protéiques humains dans diverses conditions. Enfin, il serait intéressant d'identifier les modifications post-traductionnelles des protéines présentes dans les complexes purifiés, et en particulier celles des protéines à fonction inconnue, à l'aide de la MS. Il serait alors possible de déterminer le site précis de la modification, d'identifier l'enzyme catalysant cette modification et d'étudier sa conséquence sur la fonction/régulation de la protéine modifiée.

Le second objectif de cette thèse était d'identifier et de caractériser de nouveaux régulateurs de la machinerie transcriptionnelle de l'ARN Pol II. Nous avons atteint cet objectif en identifiant plusieurs polypeptides à fonction inconnue, dont les XAB1-RPAPs. Nos résultats montrent que les protéines XAB1-RPAPs sont physiquement et fonctionnellement associées à l'ARN Pol II, formant une interface entre l'enzyme et des protéines chaperons et d'échafaudage. Il paraît important de déterminer le rôle précis des facteurs XAB1-RPAPs dans le processus transcriptionnel. Ainsi, il serait particulièrement intéressant d'investiguer leur localisation cellulaire afin de vérifier que ces protéines co-localisent avec l'ARN Pol II. Dans le cas de facteurs nucléaires, il serait aussi pertinent de les localiser sur l'ADN par la technique de CHIP. Puisque les protéines XAB1-RPAPs ont des homologues chez la levure, il serait utile d'utiliser ce modèle biologique pour étudier la fonction de ces protéines. Enfin, il paraît nécessaire d'identifier les partenaires protéiques des facteurs à fonction inconnue associées aux protéines XAB1-RPAPs. Ceci permettrait d'obtenir des informations additionnelles pertinentes sur la composition des complexes protéiques formés par ces protéines.

BCDIN3 et LARP7, deux autres polypeptides à fonction inconnue trouvés en association avec le facteur d'élongation P-TEFb et le snARN 7SK, ont aussi été étudiés. Nous avons montré que BCDIN3 est l'enzyme de coiffage de 7SK. BCDIN3 a été nommée *methylphosphate capping enzyme* (MEPCE). Après avoir montré que MEPCE est l'enzyme de coiffage d'autres transcrits de l'ARN Pol III, notamment du snARN U6 et de B2 chez la souris, il serait intéressant d'investiguer si MEPCE est capable d'ajouter une coiffe sur Alu, un petit ARN humain jouant un rôle identique à celui de l'ARN B2

murin dans la répression de l'ARN Pol II en réponse au choc thermique (Mariner et al., 2008). Éventuellement, la présence de l'ARN Alu dans le complexe MEPCE-TAP pourrait être déterminée. D'autre part, afin de mieux connaître le mécanisme de catalyse de MEPCE, il serait pertinent de déterminer la structure cristallographique de cette nouvelle méthyltransférase à ARN en complexe avec son substrat, le donneur de groupement méthyle (l'AdoMet) ou encore, le produit de la réaction de méthylation (l'AdoHcy).

Quant à LARP7, nous avons montré que cette protéine est une composante stable de la ribonucléoprotéine 7SK•HEXIM•MEPCE•P-TEFb. Il serait essentiel de démontrer que LARP7 est capable de lier le snARN 7SK, et, à quelle extrémité de l'ARN. De plus, il serait intéressant d'investiguer s'il existe un lien entre le coiffage de 7SK et la liaison de cet ARN par LARP7. Aussi, il serait pertinent de vérifier si l'extrémité 5' du 7SK lié par LARP7 dans les cellules est méthylée. D'autre part, nous avons pu déterminer que BCDIN3 et LARP7 jouent un rôle dans la stabilité de 7SK dans les cellules humaines, contribuant ainsi au maintien de P-TEFb dans un état transcriptionnel inactif. Dans la poursuite de ce projet, il serait pertinent d'étudier comment LARP7 et MEPCE régulent la fonction de P-TEFb pendant la différenciation cellulaire, le développement et la progression tumorale.

BIBLIOGRAPHIE

- Ahmad,K. et Henikoff,S. (2002a). Histone H3 variants specify modes of chromatin assembly. *Proc. Natl. Acad. Sci. U. S. A.*
- Ahmad,K. et Henikoff,S. (2002b). The histone variant h3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* 9, 1191-1200.
- Ahn,S.H., Kim,M., et Buratowski,S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell* 13, 67-76.
- Akey,C.W. et Luger,K. (2003). Histone chaperones and nucleosome assembly. *Curr. Opin. Struct. Biol* 13, 6-14.
- Alberts,B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92, 291-294.
- Albright,S.R. et Tjian,R. (2000). TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* 242, 1-13.
- Allen,T.A., Von Kaenel,S., Goodrich,J.A., et Kugel,J.F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat. Struct. Mol Biol* 11, 816-821.
- Archambault,J., Chambers,R.S., Kobor,M.S., Ho,Y., Cartier,M., Bolotin,D., Andrews,B., Kane,C.M., et 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.
- Ares,M., Jr. et Proudfoot,N.J. (2005). The spanish connection: transcription and mRNA processing get even closer. *Cell* 120, 163-166.
- Armache,K.J., Kettenberger,H., et Cramer,P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A* 100, 6964-6968.
- Awrey,D.E., Shimasaki,N., Koth,C., Weilbaecher,R., Olmsted,V., Kazanis,S., Shan,X., Arellano,J., Arrowsmith,C.H., Kane,C.M., et Edwards,A.M. (1998). Yeast transcript elongation factor (TFIIS), structure and function. II: RNA polymerase binding, transcript cleavage, and read-through. *J. Biol Chem.* 273, 22595-22605.
- Awrey,D.E., Weilbaecher,R.G., Hemming,S.A., Orlicky,S.M., Kane,C.M., et Edwards,A.M. (1997). Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J. Biol. Chem.* 272, 14747-14754.

Bachvarova,R. (1988). Small B2 RNAs in mouse oocytes, embryos, and somatic tissues. *Dev. Biol.* 130, 513-523.

Bagby,S., Kim,S., Maldonado,E., Tong,K.I., Reinberg,D., et 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.

Baillat,D., Hakimi,M.A., Naar,A.M., Shilatifard,A., Cooch,N., et Shiekhattar,R. (2005). Integrator, a Multiprotein Mediator of Small Nuclear RNA Processing, Associates with the C-Terminal Repeat of RNA Polymerase II. *Cell* 123, 265-276.

Baker,S.P. et Grant,P.A. (2007). The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. *Oncogene* 26, 5329-5340.

Ballestar,E. et Esteller,M. (2005). Methyl-CpG-binding proteins in cancer: blaming the DNA methylation messenger. *Biochem. Cell Biol.* 83, 374-384.

Bao,Y. et Shen,X. (2007). INO80 subfamily of chromatin remodeling complexes. *Mutat. Res.* 618, 18-29.

Barboric,M., Kohoutek,J., Price,J.P., Blazek,D., Price,D.H., et Peterlin,B.M. (2005). Interplay between 7SK snRNA and oppositely charged regions in HEXIM1 direct the inhibition of P-TEFb. *EMBO J.* 24, 4291-4303.

Barboric,M. et Peterlin,B.M. (2005). A new paradigm in eukaryotic biology: HIV Tat and the control of transcriptional elongation. *PLoS. Biol* 3, e76.

Barilla,D., Lee,B.A., et Proudfoot,N.J. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A* 98, 445-450.

Barrandon,C., Bonnet,F., Nguyen,V.T., Labas,V., et Bensaude,O. (2007). The transcription-dependent dissociation of P-TEFb.HEXIM1.7SK RNA relies upon formation of hnRNP.7SK RNA complexes. *Mol Cell Biol.*

Bartfai,R., Balduf,C., Hilton,T., Rathmann,Y., Hadzhiev,Y., Tora,L., Orban,L., et Muller,F. (2004). TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Curr. Biol* 14, 593-598.

Batsche,E., Yaniv,M., et Muchardt,C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.* 13, 22-29.

Bell,M., Schreiner,S., Damianov,A., Reddy,R., et Bindereif,A. (2002). p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.* 21, 2724-2735.

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

- Belotserkovskaya,R. et Reinberg,D. (2004). Facts about FACT and transcript elongation through chromatin. *Curr. Opin. Genet. Dev.* *14*, 139-146.
- Bengal,E., Flores,O., Krauskopf,A., Reinberg,D., et 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.
- Bentley,D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol* *17*, 251-256.
- Berger,S.L. (2007). The complex language of chromatin regulation during transcription. *Nature* *447*, 407-412.
- Bernstein,B.E., Kamal,M., Lindblad-Toh,K., Bekiranov,S., Bailey,D.K., Huebert,D.J., McMahon,S., Karlsson,E.K., Kulbokas,E.J., III, Gingeras,T.R., Schreiber,S.L., et Lander,E.S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* *120*, 169-181.
- Bhattacharya,R., Perumal,K., Sinha,K., Maraia,R., et Reddy,R. (2002). Methylphosphate cap structure in small RNAs reduces the affinity of RNAs to La protein. *Gene Expr.* *10*, 243-253.
- Bhaumik,S.R., Smith,E., et Shilatifard,A. (2007). Covalent modifications of histones during development and disease pathogenesis. *Nat. Struct. Mol. Biol.* *14*, 1008-1016.
- Bird,A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* *16*, 6-21.
- Birse,C.E., Minvielle-Sebastia,L., Lee,B.A., Keller,W., et Proudfoot,N.J. (1998). Coupling termination of transcription to messenger RNA maturation in yeast. *Science* *280*, 298-301.
- Bjorklund,S. et Gustafsson,C.M. (2005). The yeast Mediator complex and its regulation. *Trends Biochem. Sci.* *30*, 240-244.
- Black,D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* *72*, 291-336.
- Blackwood,E.M. et Kadonaga,J.T. (1998). Going the distance: a current view of enhancer action. *Science* *281*, 61-63.
- Blazek,D., Barboric,M., Kohoutek,J., Oven,I., et Peterlin,B.M. (2005). Oligomerization of HEXIM1 via 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. *Nucleic Acids Res.* *33*, 7000-7010.
- Blencowe,B.J. (2006). Alternative splicing: new insights from global analyses. *Cell* *126*, 37-47.

- Boeger,H., Bushnell,D.A., Davis,R., Griesenbeck,J., Lorch,Y., Strattan,J.S., Westover,K.D., et Kornberg,R.D. (2005). Structural basis of eukaryotic gene transcription. *FEBS Letters* 579, 899-903.
- Bortvin,A. et Winston,F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science* 272, 1473-1476.
- Boube,M., Joulia,L., Cribbs,D.L., et Bourbon,H.M. (2002). Evidence for a Mediator of RNA Polymerase II Transcriptional Regulation Conserved from Yeast to Man. *Cell* 110, 143-151.
- Boyer,T.G., Martin,M.E., Lees,E., Ricciardi,R.P., et Berk,A.J. (1999). Mammalian Srb/Mediator complex is targeted by adenovirus E1A protein. *Nature* 399, 276-279.
- Breathnach,R. et Chambon,P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50, 349-383.
- Brehm,A., Langst,G., Kehle,J., Clapier,C.R., Imhof,A., Eberharter,A., Muller,J., et Becker,P.B. (2000). dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. *EMBO J.* 19, 4332-4341.
- Brehm,A., Tufteland,K.R., Aasland,R., et Becker,P.B. (2004). The many colours of chromodomains. *Bioessays* 26, 133-140.
- Bres,V., Gomes,N., Pickle,L., et Jones,K.A. (2005). A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. *Genes Dev.* 19, 1211-1226.
- Brower,C.S., Sato,S., Tomomori-Sato,C., Kamura,T., Pause,A., Stearman,R., Klausner,R.D., Malik,S., Lane,W.S., Sorokina,I., Roeder,R.G., Conaway,J.W., et Conaway,R.C. (2002). Mammalian mediator subunit mMED8 is an Elongin BC-interacting protein that can assemble with Cul2 and Rbx1 to reconstitute a ubiquitin ligase. *Proc. Natl. Acad. Sci. U. S. A* 99, 10353-10358.
- Bruno,M., Flaus,A., Stockdale,C., Rencurel,C., Ferreira,H., et Owen-Hughes,T. (2003). Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities. *Mol. Cell* 12, 1599-1606.
- Buratowski,S. (2003). The CTD code. *Nat. Struct. Biol.* 10, 679-680.
- Buratowski,S. (2005). Connections between mRNA 3' end processing and transcription termination. *Curr. Opin. Cell Biol* 17, 257-261.
- Buratowski,S., Hahn,S., Guarente,L., et Sharp,P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 56, 549-561.

- Buratowski,S., Sopta,M., Greenblatt,J., et Sharp,P.A. (1991). RNA polymerase II-associated proteins are required for a DNA conformation change in the transcription initiation complex. *Proc. Natl. Acad. Sci. U. S. A* 88, 7509-7513.
- Burke,T.W. et 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.
- Burke,T.W. et Kadonaga,J.T. (1997). The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev.* 11, 3020-3031.
- Burley,S.K. et Roeder,R.G. (1996). Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* 65, 769-799.
- Bushnell,D.A., Bamdad,C., et Kornberg,R.D. (1996). A minimal set of RNA polymerase II transcription protein interactions. *J. Biol. Chem.* 271, 20170-20174.
- Bushnell,D.A. et Kornberg,R.D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc. Natl. Acad. Sci. U. S. A* 100, 6969-6973.
- Bushnell,D.A., Westover,K.D., Davis,R.E., et Kornberg,R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Ångstroms. *Science* 303, 983-988.
- Butland,G., Peregrin-Alvarez,J.M., Li,J., Yang,W., Yang,X., Canadien,V., Starostine,A., Richards,D., Beattie,B., Krogan,N., Davey,M., Parkinson,J., Greenblatt,J., et Emili,A. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 433, 531-537.
- Butler,J.E. et Kadonaga,J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* 16, 2583-2592.
- Byers,S.A., Price,J.P., Cooper,J.J., Li,Q., et Price,D.H. (2005). HEXIM2, a HEXIM1-related protein, regulates positive transcription elongation factor b through association with 7SK. *J. Biol Chem.* 280, 16360-16367.
- Cairns,B.R. (2007). Chromatin remodeling: insights and intrigue from single-molecule studies. *Nat. Struct. Mol. Biol.* 14, 989-996.
- Calvo,O. et Manley,J.L. (2005). The transcriptional coactivator PC4/Sub1 has multiple functions in RNA polymerase II transcription. *EMBO J.* 24, 1009-1020.
- Calvo,O. et Manley,J.L. (2003). Strange bedfellows: polyadenylation factors at the promoter. *Genes and Development* 17, 1321-1327.

- Capelson, M. et Corces, V.G. (2004). Boundary elements and nuclear organization. *Biol. Cell* *96*, 617-629.
- Carlson, M. (1997). Genetics of transcriptional regulation in yeast: connections to the RNA polymerase II CTD. *Annu. Rev. Cell Dev. Biol.* *13*, 1-23.
- Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., et Workman, J.L. (2005). Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell* *123*, 581-592.
- Chambers, J.C., Kurilla, M.G., et Keene, J.D. (1983). Association between the 7 S RNA and the lupus La protein varies among cell types. *J. Biol. Chem.* *258*, 11438-11441.
- Chambers, R.S., Wang, B.Q., Burton, Z.F., et 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.
- Chang, W.H. et Kornberg, R.D. (2000). Electron crystal structure of the transcription factor and DNA repair complex, core TFIIF. *Cell* *102*, 609-613.
- Chen, H.T. et Hahn, S. (2003). Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol. Cell* *12*, 437-447.
- Chen, J., Larochelle, S., Li, X., et Suter, B. (2003). Xpd/Ercc2 regulates CAK activity and mitotic progression. *Nature* *424*, 228-232.
- Chen, L. et Widom, J. (2005). Mechanism of transcriptional silencing in yeast. *Cell* *120*, 37-48.
- Chen, R., Yang, Z., et Zhou, Q. (2004). Phosphorylated positive transcription elongation factor b (P-TEFb) is tagged for inhibition through association with 7SK snRNA. *J. Biol. Chem.* *279*, 4153-4160.
- Chen, X., Ruggiero, C., et Li, S. (2007). Yeast Rpb9 plays an important role in ubiquitylation and degradation of Rpb1 in response to UV-induced DNA damage. *Mol. Cell Biol.* *27*, 4617-4625.
- Chiu, Y.L., Ho, C.K., Saha, N., Schwer, B., Shuman, S., et Rana, T.M. (2002). Tat stimulates cotranscriptional capping of HIV mRNA. *Mol Cell* *10*, 585-597.
- Cho, H., Kim, T.K., Mancebo, H., Lane, W.S., Flores, O., et Reinberg, D. (1999). A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev.* *13*, 1540-1552.
- Choder, M. (2004). Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. *Trends Biochem. Sci.* *29*, 674-681.

- Choder,M. et Young,R.A. (1993). A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. *Mol. Cell Biol.* *13*, 6984-6991.
- Chung,W.H., Craighead,J.L., Chang,W.H., Ezeokonkwo,C., Bareket-Samish,A., Kornberg,R.D., et Asturias,F.J. (2003). RNA Polymerase II/TFIIF Structure and Conserved Organization of the Initiation Complex. *Mol. Cell* *12*, 1003-1013.
- Claessens,F. et Gewirth,D.T. (2004). DNA recognition by nuclear receptors. *Essays Biochem.* *40*, 59-72.
- Coin,F., De Santis,L.P., Nardo,T., Zlobinskaya,O., Stefanini,M., et Egly,J.M. (2006). p8/TTD-A as a repair-specific TFIIH subunit. *Mol Cell* *21*, 215-226.
- Coin,F. et Egly,J.M. (1998). Ten years of TFIIH. *Cold Spring Harb. Symp. Quant. Biol.* *63*, 105-110.
- Coin,F., Oksenysh,V., et Egly,J.M. (2007). Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol. Cell* *26*, 245-256.
- Cojocar,M., Jeronimo,C., Forget,D., Bouchard,A., Bergeron,D., Cote,P., Poirier,G.G., Greenblatt,J., et Coulombe,B. (2008). Genomic location of the human RNA polymerase II general machinery: evidence for a role of TFIIF and Rpb7 at both early and late stages of transcription. *Biochem. J.* *409*, 139-147.
- Colgan,D.F. et Manley,J.L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* *11*, 2755-2766.
- Conaway,J.W. et Conaway,R.C. (1999). Transcription elongation and human disease. *Annu. Rev. Biochem.* *68*, 301-319.
- Conaway,R.C., Garrett,K.P., Hanley,J.P., et Conaway,J.W. (1991). Mechanism of promoter selection by RNA polymerase II: mammalian transcription factors alpha and beta gamma promote entry of polymerase into the preinitiation complex. *Proc. Natl. Acad. Sci. U. S. A* *88*, 6205-6209.
- Conaway,R.C., Sato,S., Tomomori-Sato,C., Yao,T., et Conaway,J.W. (2005). The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci.* *30*, 250-255.
- Connelly,S. et Manley,J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* *2*, 440-452.
- Cook,P.R. (2003). Nongenic transcription, gene regulation and action at a distance. *J. Cell Sci.* *116*, 4483-4491.

- Cooper,S.J., Trinklein,N.D., Anton,E.D., Nguyen,L., et Myers,R.M. (2006). Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* *16*, 1-10.
- Corden,J., Wasylyk,B., Buchwalder,A., Sassone-Corsi,P., Kedinger,C., et Chambon,P. (1980). Promoter sequences of eukaryotic protein-coding genes. *Science* *209*, 1406-1414.
- Cortes,P., Flores,O., et Reinberg,D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II: purification and analysis of transcription factor IIA and identification of transcription factor IIJ. *Mol. Cell Biol.* *12*, 413-421.
- Cosma,M.P., Panizza,S., et Nasmyth,K. (2001). Cdk1 triggers association of RNA polymerase to cell cycle promoters only after recruitment of the mediator by SBF. *Mol. Cell* *7*, 1213-1220.
- Costanzi,C. et Pehrson,J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* *393*, 599-601.
- Cougot,N., van Dijk,E., Babajko,S., et Seraphin,B. (2004). 'Cap-tabolism'. *Trends Biochem. Sci.* *29*, 436-444.
- Coulombe,B. et Burton,Z.F. (1999). DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. *Microbiol. Mol. Biol. Rev.* *63*, 457-478.
- Coulombe,B., Jeronimo,C., Langelier,M.F., Cojocaru,M., et Bergeron,D. (2004). Interaction networks of the molecular machines that decode, replicate and maintain the integrity of the human genome. *Mol. Cell Proteomics.* *3*, 851-856.
- Coulombe,B., Blanchette,M., et Jeronimo,C. (2008). Steps towards a repertoire of comprehensive maps of human protein interaction networks: the Human Proteotheque Initiative (HuPI). *Biochemistry and Cell Biology* *86*, 149-156.
- Courey,A.J. et Jia,S. (2001). Transcriptional repression: the long and the short of it. *Genes Dev.* *15*, 2786-2796.
- Cramer,P. (2006). Deciphering the RNA polymerase II structure: a personal perspective. *Nat. Struct. Mol Biol* *13*, 1042-1044.
- Cramer,P. (2004). RNA polymerase II structure: from core to functional complexes. *Curr. Opin. Genet. Dev.* *14*, 218-226.
- Cramer,P., Bushnell,D.A., et Kornberg,R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* *292*, 1863-1876.
- Cui,Y. et Denis,C.L. (2003). In vivo evidence that defects in the transcriptional elongation factors RPB2, TFIIS, and SPT5 enhance upstream poly(A) site utilization. *Mol. Cell Biol.* *23*, 7887-7901.

- Dahmus, M.E. (1996). Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *J. Biol. Chem.* *271*, 19009-19012.
- Dalal, Y., Furuyama, T., Vermaak, D., et Henikoff, S. (2007). Structure, dynamics, and evolution of centromeric nucleosomes. *Proc. Natl. Acad. Sci. U. S. A* *104*, 15974-15981.
- Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M., et Reed, R. (2006). Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev.* *20*, 1100-1109.
- Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., et Reed, R. (2007). SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. *Mol Cell* *26*, 867-881.
- Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., et Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* *319*, 1097-1113.
- Davidson, I. (2003). The genetics of TBP and TBP-related factors. *Trends Biochem. Sci.* *28*, 391-398.
- Davidson, I., Kobi, D., Fadloun, A., et Mengus, G. (2005). New insights into TAFs as regulators of cell cycle and signaling pathways. *Cell Cycle* *4*, 1486-1490.
- Davie, J.K. et Kane, C.M. (2000). Genetic interactions between TFIIS and the Swi-Snf chromatin-remodeling complex. *Mol. Cell Biol.* *20*, 5960-5973.
- de Boer, J. et Hoeijmakers, J.H. (2000). Nucleotide excision repair and human syndromes. *Carcinogenesis* *21*, 453-460.
- De Koning, L., Corpet, A., Haber, J.E., et Almouzni, G. (2007). Histone chaperones: an escort network regulating histone traffic. *Nat. Struct. Mol. Biol.* *14*, 997-1007.
- de la, M.M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., et Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* *12*, 525-532.
- de, V.H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H., et Keller, W. (2000). Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.* *19*, 5895-5904.
- Dean, A. (2006). On a chromosome far, far away: LCRs and gene expression. *Trends Genet.* *22*, 38-45.
- Deato, M.D. et Tjian, R. (2007). Switching of the core transcription machinery during myogenesis. *Genes Dev.* *21*, 2137-2149.
- DeJong, J. et Roeder, R.G. (1993). A single cDNA, hTFIIA/alpha, encodes both the p35 and p19 subunits of human TFIIA. *Genes Dev.* *7*, 2220-2234.

- Demény, M.A., Soutoglou, E., Nagy, Z., Scheer, E., Janoshazi, A., Richardot, M., Argentini, M., Kessler, P., et Tora, L. (2007). Identification of a small TAF complex and its role in the assembly of TAF-containing complexes. *PLoS ONE* 2, e316.
- Deng, W. et Roberts, S.G. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev.* 19, 2418-2423.
- Deng, W. et Roberts, S.G. (2007). TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma* 116, 417-429.
- Denslow, S.A. et Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. *Oncogene* 26, 5433-5438.
- Deuring, R., Fanti, L., Armstrong, J.A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S.L., Berloco, M., Tsukiyama, T., Wu, C., Pimpinelli, S., et Tamkun, J.W. (2000). The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol. Cell* 5, 355-365.
- Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., et Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399, 491-496.
- Ding, B., Ruggiero, C., Chen, X., et Li, S. (2007). Tfb5 is partially dispensable for Rad26 mediated transcription coupled nucleotide excision repair in yeast. *DNA Repair (Amst)* 6, 1661-1669.
- Domon, B. et Aebersold, R. (2006). Mass spectrometry and protein analysis. *Science* 312, 212-217.
- Douziech, M., Coin, F., Chipoulet, J.M., Arai, Y., Ohkuma, Y., Egly, J.M., et 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.
- Drapkin, R., Reardon, J.T., Ansari, A., Huang, J.C., Zawel, L., Ahn, K., Sancar, A., et Reinberg, D. (1994). Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature* 368, 769-772.
- Dreyfus, M. et Regnier, P. (2002). The Poly(A) Tail of mRNAs. Bodyguard in Eukaryotes, Scavenger in Bacteria. *Cell* 111, 611-613.
- Dubaele, S., De Santis, L.P., Bienstock, R.J., Keriél, A., Stefanini, M., Van Houten, B., et Egly, J.M. (2003). Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Mol. Cell* 11, 1635-1646.
- Dunn, R.K. et Kingston, R.E. (2007). Gene regulation in the postgenomic era: technology takes the wheel. *Mol. Cell* 28, 708-714.

- Dvir,A., Conaway,J.W., et Conaway,R.C. (2001). Mechanism of transcription initiation and promoter escape by RNA polymerase II. *Curr. Opin. Genet. Dev.* 11, 209-214.
- Dye,M.J. et Proudfoot,N.J. (2001). Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* 105, 669-681.
- Edwards,A.M., Kane,C.M., Young,R.A., et Kornberg,R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J. Biol. Chem.* 266, 71-75.
- Egloff,S., Van Herreweghe,E., et Kiss,T. (2006). Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. *Mol Cell Biol* 26, 630-642.
- Elsby,L.M. et Roberts,S.G. (2004). The role of TFIIB conformation in transcriptional regulation. *Biochem. Soc. Trans.* 32, 1098-1099.
- Endoh,M., Zhu,W., Hasegawa,J., Watanabe,H., Kim,D.K., Aida,M., Inukai,N., Narita,T., Yamada,T., Furuya,A., Sato,H., Yamaguchi,Y., Mandal,S.S., Reinberg,D., Wada,T., et Handa,H. (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Mol. Cell Biol.* 24, 3324-3336.
- Espinoza,C.A., Allen,T.A., Hieb,A.R., Kugel,J.F., et Goodrich,J.A. (2004). B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat. Struct. Mol Biol* 11, 822-829.
- Esteller,M. (2007). Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum. Mol. Genet.* 16 *Spec No 1*, R50-R59.
- Ewing,R.M., Chu,P., Elisma,F., Li,H., Taylor,P., Climie,S., McBroom-Cerajewski,L., Robinson,M.D., O'Connor,L., Li,M., Taylor,R., Dharsee,M., Ho,Y., Heilbut,A., Moore,L., Zhang,S., Ornatsky,O., Bukhman,Y.V., Ethier,M., Sheng,Y., Vasilescu,J., Abu-Farha,M., Lambert,J.P., Duewel,H.S., Stewart,I.I., Kuehl,B., Hogue,K., Colwill,K., Gladwish,K., Muskat,B., Kinach,R., Adams,S.L., Moran,M.F., Morin,G.B., Topaloglou,T., et Figeys,D. (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol. Syst. Biol.* 3, 89.
- Exinger,F. et Lacroute,F. (1992). 6-Azauracil Inhibition of Gtp Biosynthesis in *Saccharomyces-Cerevisiae*. *Current Genetics* 22, 9-11.
- Faast,R., Thonglairoam,V., Schulz,T.C., Beall,J., Wells,J.R., Taylor,H., Matthaei,K., Rathjen,P.D., Tremethick,D.J., et Lyons,I. (2001). Histone variant H2A.Z is required for early mammalian development. *Curr. Biol.* 11, 1183-1187.
- Fairley,J.A., Evans,R., Hawkes,N.A., et 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.

- Fan,X., Chou,D.M., et Struhl,K. (2006). Activator-specific recruitment of Mediator in vivo. *Nat. Struct. Mol. Biol.* *13*, 117-120.
- Fang,S.M. et 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.
- Farago,M., Nahari,T., Hammel,C., Cole,C.N., et Choder,M. (2003). Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. *Mol Biol Cell* *14*, 2744-2755.
- Fatica,A., Morlando,M., et Bozzoni,I. (2000). Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. *EMBO J.* *19*, 6218-6229.
- Feaver,W.J., Svejstrup,J.Q., Henry,N.L., et Kornberg,R.D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIF/TFIIFK. *Cell* *79*, 1103-1109.
- Feng,Q., Wang,H., Ng,H.H., Erdjument-Bromage,H., Tempst,P., Struhl,K., et Zhang,Y. (2002). Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* *12*, 1052-1058.
- Feng,Q. et Zhang,Y. (2003). The NuRD complex: linking histone modification to nucleosome remodeling. *Curr. Top. Microbiol. Immunol.* *274*, 269-290.
- Findly,R.C. et Pederson,T. (1981). Regulated transcription of the genes for actin and heat-shock proteins in cultured *Drosophila* cells. *J. Cell Biol.* *88*, 323-328.
- Finkelstein,A., Kostrub,C.F., Li,J., Chavez,D.P., Wang,B.Q., Fang,S.M., Greenblatt,J., et Burton,Z.F. (1992). A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II. *Nature* *355*, 464-467.
- Fish,R.N. et Kane,C.M. (2002). Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta* *1577*, 287-307.
- Flanagan,J.F., Mi,L.Z., Chruszcz,M., Cymborowski,M., Clines,K.L., Kim,Y., Minor,W., Rastinejad,F., et Khorasanizadeh,S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* *438*, 1181-1185.
- Flores,O., Lu,H., Killeen,M., Greenblatt,J., Burton,Z.F., et 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.
- Fondell,J.D., Ge,H., et Roeder,R.G. (1996). Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc. Natl. Acad. Sci. U. S. A* *93*, 8329-8333.
- Fondell,J.D., Guermah,M., Malik,S., et Roeder,R.G. (1999). Thyroid hormone receptor-associated proteins and general positive cofactors mediate thyroid hormone receptor

function in the absence of the TATA box-binding protein-associated factors of TFIID. *Proc. Natl. Acad. Sci. U. S. A* *96*, 1959-1964.

Fong, Y.W. et Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature* *414*, 929-933.

Forget, D., Robert, F., Grondin, G., Burton, Z.F., Greenblatt, J., et Coulombe, B. (1997). RAP74 induces promoter contacts by RNA polymerase II upstream and downstream of a DNA bend centered on the TATA box. *Proc. Natl. Acad. Sci. U. S. A* *94*, 7150-7155.

Fornace, A.J., Jr. et Mitchell, J.B. (1986). Induction of B2 RNA polymerase III transcription by heat shock: enrichment for heat shock induced sequences in rodent cells by hybridization subtraction. *Nucleic Acids Res.* *14*, 5793-5811.

Fu, T.J., Peng, J., Lee, G., Price, D.H., et Flores, O. (1999). Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. *J. Biol. Chem.* *274*, 34527-34530.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., et Peterlin, B.M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol. Cell Biol.* *24*, 787-795.

Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A., et Proudfoot, N.J. (2002). Promoter proximal splice sites enhance transcription. *Genes Dev.* *16*, 2792-2799.

Furter-Graves, E.M., Hall, B.D., et Furter, R. (1994). Role of a small RNA pol II subunit in TATA to transcription start site spacing. *Nucleic Acids Res.* *22*, 4932-4936.

Gal-Yam, E.N., Saito, Y., Egger, G., et Jones, P.A. (2007). Cancer Epigenetics: Modifications, Screening, and Therapy. *Annu. Rev. Med.*

Gandhi, T.K., Zhong, J., Mathivanan, S., Karthick, L., Chandrika, K.N., Mohan, S.S., Sharma, S., Pinkert, S., Nagaraju, S., Periaswamy, B., Mishra, G., Nandakumar, K., Shen, B., Deshpande, N., Nayak, R., Sarker, M., Boeke, J.D., Parmigiani, G., Schultz, J., Bader, J.S., et Pandey, A. (2006). Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat. Genet.* *38*, 285-293.

Gangaraju, V.K. et Bartholomew, B. (2007). Mechanisms of ATP dependent chromatin remodeling. *Mutat. Res.* *618*, 3-17.

Gangloff, Y.G., Romier, C., Thuault, S., Werten, S., et Davidson, I. (2001). The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem. Sci.* *26*, 250-257.

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

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

Gavin,A.C., Aloy,P., Grandi,P., Krause,R., Boesche,M., Marzioch,M., Rau,C., Jensen,L.J., Bastuck,S., Dumpelfeld,B., Edelmann,A., Heurtier,M.A., Hoffman,V., Hoefert,C., Klein,K., Hudak,M., Michon,A.M., Schelder,M., Schirle,M., Remor,M., Rudi,T., Hooper,S., Bauer,A., Bouwmeester,T., Casari,G., Drewes,G., Neubauer,G., Rick,J.M., Kuster,B., Bork,P., Russell,R.B., et Superti-Furga,G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440, 631-636.

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., Edelmann,A., Querfurth,E., Rybin,V., Drewes,G., Raida,M., Bouwmeester,T., Bork,P., Seraphin,B., Kuster,B., Neubauer,G., et Superti-Furga,G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes
1. *Nature* 415, 141-147.

Ge,H. et Roeder,R.G. (1994). Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell* 78, 513-523.

Geiger,J.H., Hahn,S., Lee,S., et Sigler,P.B. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* 272, 830-836.

Gerber,M., Ma,J., Dean,K., Eissenberg,J.C., et Shilatifard,A. (2001). Drosophila ELL is associated with actively elongating RNA polymerase II on transcriptionally active sites in vivo. *EMBO J.* 20, 6104-6114.

Gerber,M., Tenney,K., Conaway,J.W., Conaway,R.C., Eissenberg,J.C., et Shilatifard,A. (2005). Regulation of heat shock gene expression by RNA polymerase II elongation factor, Elongin A. *J. Biol Chem.* 280, 4017-4020.

Gershenson,N.I. et Ioshikhes,I.P. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics.* 21, 1295-1300.

Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Veronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., Andre,B., Arkin,A.P., Astromoff,A., El Bakkoury,M., Bangham,R., Benito,R., Brachat,S., Campanaro,S., Curtiss,M., Davis,K., Deutschbauer,A., Entian,K.D., Flaherty,P., Foury,F., Garfinkel,D.J., Gerstein,M., Gotte,D., Guldener,U., Hegemann,J.H., Hempel,S., Herman,Z., Jaramillo,D.F., Kelly,D.E., Kelly,S.L., Kotter,P., LaBonte,D., Lamb,D.C., Lan,N., Liang,H., Liao,H., Liu,L., Luo,C., Lussier,M., Mao,R., Menard,P., Ooi,S.L., Revuelta,J.L., Roberts,C.J., Rose,M., Ross-Macdonald,P., Scherens,B., Schimmack,G., Shafer,B., Shoemaker,D.D., Sookhai-Mahadeo,S., Storms,R.K., Strathern,J.N., Valle,G., Voet,M., Volckaert,G., Wang,C.Y., Ward,T.R., Wilhelmy,J., Winzeler,E.A., Yang,Y., Yen,G., Youngman,E., Yu,K., Bussey,H., Boeke,J.D., Snyder,M., Philippsen,P., Davis,R.W., et Johnston,M.

(2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387-391.

Giglia-Mari,G., Coin,F., Ranish,J.A., Hoogstraten,D., Theil,A., Wijgers,N., Jaspers,N.G., Raams,A., Argentini,M., van der Spek,P.J., Botta,E., Stefanini,M., Egly,J.M., Aebersold,R., Hoeijmakers,J.H., et Vermeulen,W. (2004). A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat. Genet.* 36, 714-719.

Giglia-Mari,G., Miquel,C., Theil,A.F., Mari,P.O., Hoogstraten,D., Ng,J.M., Dinant,C., Hoeijmakers,J.H., et Vermeulen,W. (2006). Dynamic interaction of TTDA with TFIIH is stabilized by nucleotide excision repair in living cells. *PLoS. Biol.* 4, e156.

Gilmour,D.S. et Lis,J.T. (1985). In vivo interactions of RNA polymerase II with genes of *Drosophila melanogaster*. *Mol. Cell Biol.* 5, 2009-2018.

Gingras,A.C., Gstaiger,M., Raught,B., et Aebersold,R. (2007). Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* 8, 645-654.

Giot,L., Bader,J.S., Brouwer,C., Chaudhuri,A., Kuang,B., Li,Y., Hao,Y.L., Ooi,C.E., Godwin,B., Vitols,E., Vijayadamar,G., Pochart,P., Machineni,H., Welsh,M., Kong,Y., Zerhusen,B., Malcolm,R., Varrone,Z., Collis,A., Minto,M., Burgess,S., McDaniel,L., Stimpson,E., Spriggs,F., Williams,J., Neurath,K., Ioime,N., Agee,M., Voss,E., Furtak,K., Renzulli,R., Aanensen,N., Carrolla,S., Bickelhaupt,E., Lazovatsky,Y., DaSilva,A., Zhong,J., Stanyon,C.A., Finley,R.L., Jr., White,K.P., Braverman,M., Jarvie,T., Gold,S., Leach,M., Knight,J., Shimkets,R.A., McKenna,M.P., Chant,J., et Rothberg,J.M. (2003). A protein interaction map of *Drosophila melanogaster*. *Science* 302, 1727-1736.

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

Goldberg,A.D., Allis,C.D., et Bernstein,E. (2007). Epigenetics: a landscape takes shape. *Cell* 128, 635-638.

Goldmark,J.P., Fazzio,T.G., Estep,P.W., Church,G.M., et Tsukiyama,T. (2000). The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. *Cell* 103, 423-433.

Goll,M.G. et Bestor,T.H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481-514.

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

Green,M.R. (2005). Eukaryotic transcription activation: right on target. *Mol Cell* 18, 399-402.

Grewal,S.I. et Elgin,S.C. (2007). Transcription and RNA interference in the formation of heterochromatin. *Nature* 447, 399-406.

Groft,C.M., Uljon,S.N., Wang,R., et Werner,M.H. (1998). Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. *Proc. Natl. Acad. Sci. U. S. A* 95, 9117-9122.

Gromak,N., West,S., et Proudfoot,N.J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol* 26, 3986-3996.

Gross,P. et Oelgeschlager,T. (2006). Core promoter-selective RNA polymerase II transcription. *Biochem. Soc. Symp.* 225-236.

Groth,A., Rocha,W., Verreault,A., et Almouzni,G. (2007). Chromatin challenges during DNA replication and repair. *Cell* 128, 721-733.

Grune,T., Brzeski,J., Eberharter,A., Clapier,C.R., Corona,D.F., Becker,P.B., et Muller,C.W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol. Cell* 12, 449-460.

Gstaiger,M., Luke,B., Hess,D., Oakeley,E.J., Wirbelauer,C., Blondel,M., Vigneron,M., Peter,M., et Krek,W. (2003). Control of nutrient-sensitive transcription programs by the unconventional prefoldin URI. *Science* 302, 1208-1212.

Gu,M. et Lima,C.D. (2005). Processing the message: structural insights into capping and decapping mRNA. *Curr. Opin. Struct. Biol.* 15, 99-106.

Gu,W., Malik,S., Ito,M., Yuan,C.X., Fondell,J.D., Zhang,X., Martinez,E., Qin,J., et Roeder,R.G. (1999). A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. *Mol. Cell* 3, 97-108.

Guarente,L. (1988). UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals. *Cell* 52, 303-305.

Gupta,S., Busch,R.K., Singh,R., et Reddy,R. (1990). Characterization of U6 small nuclear RNA cap-specific antibodies. Identification of gamma-monomethyl-GTP cap structure in 7SK and several other human small RNAs. *J. Biol. Chem.* 265, 19137-19142.

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

Hahn,S. et Roberts,S. (2000). The zinc ribbon domains of the general transcription factors TFIIB and Brf: conserved functional surfaces but different roles in transcription initiation. *Genes Dev.* 14, 719-730.

Hampsey,M. et Reinberg,D. (2003). Tails of Intrigue. Phosphorylation of RNA Polymerase II Mediates Histone Methylation. *Cell* 113, 429-432.

- Hannon,G.J., Rivas,F.V., Murchison,E.P., et Steitz,J.A. (2006). The expanding universe of noncoding RNAs. *Cold Spring Harb. Symp. Quant. Biol* 71, 551-564.
- Harris,M.B., Mostecky,J., et Rothman,P.B. (2005). Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. *J. Biol. Chem.* 280, 13114-13121.
- Hartzog,G.A. et Tamkun,J.W. (2007). A new role for histone tail modifications in transcription elongation. *Genes Dev.* 21, 3209-3213.
- Hartzog,G.A., Wada,T., Handa,H., et Winston,F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes Dev.* 12, 357-369.
- Hassan,A.H., Prochasson,P., Neely,K.E., Galasinski,S.C., Chandy,M., Carrozza,M.J., et Workman,J.L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* 111, 369-379.
- Hausmann,S. et Shuman,S. (2002). Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. *J. Biol. Chem.* 277, 21213-21220.
- Hawkes,N.A., Evans,R., et Roberts,S.G. (2000). The conformation of the transcription factor TFIIB modulates the response to transcriptional activators in vivo. *Curr. Biol.* 10, 273-276.
- Hawkes,N.A., Otero,G., Winkler,G.S., Marshall,N., Dahmus,M.E., Krappmann,D., Scheidreit,C., Thomas,C.L., Schiavo,G., Erdjument-Bromage,H., Tempst,P., et Svejstrup,J.Q. (2002). Purification and characterization of the human elongator complex. *J. Biol. Chem.* 277, 3047-3052.
- Hemming,S.A., Jansma,D.B., Macgregor,P.F., Goryachev,A., Friesen,J.D., et Edwards,A.M. (2000). RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. *J. Biol. Chem.* 275, 35506-35511.
- Hengartner,C.J., Myer,V.E., Liao,S.M., Wilson,C.J., Koh,S.S., et Young,R.A. (1998). Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol. Cell* 2, 43-53.
- Henikoff,S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat. Rev. Genet.* 9, 15-26.
- Henikoff,S., Furuyama,T., et Ahmad,K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. *Trends Genet.* 20, 320-326.
- Hicks,M.J., Yang,C.R., Kotlajich,M.V., et Hertel,K.J. (2006). Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. *PLoS. Biol.* 4, e147.

- Hirose, Y. et Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395, 93-96.
- Hirose, Y. et Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev.* 14, 1415-1429.
- Ho, C.K. et Shuman, S. (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol. Cell* 3, 405-411.
- 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., Jespersen, 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., Figgeys, D., et Tyers, M. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180-183.
- Hochheimer, A. et Tjian, R. (2003). Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. *Genes Dev.* 17, 1309-1320.
- Hochheimer, A., Zhou, S., Zheng, S., Holmes, M.C., et Tjian, R. (2002). TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* 420, 439-445.
- Hogan, C. et Varga-Weisz, P. (2007). The regulation of ATP-dependent nucleosome remodelling factors. *Mutat. Res.* 618, 41-51.
- Hoiby, T., Mitsiou, D.J., Zhou, H.Q., Erdjument-Bromage, H., Tempst, P., et Stunnenberg, H.G. (2004). Cleavage and proteasome-mediated degradation of the basal transcription factor TFIIA. *EMBO J.* 23, 3083-3091.
- Hoiby, T., Zhou, H., Mitsiou, D.J., et Stunnenberg, H.G. (2007). A facelift for the general transcription factor TFIIA. *Biochim. Biophys. Acta* 1769, 429-436.
- Holmes, M.C. et Tjian, R. (2000). Promoter-selective properties of the TBP-related factor TRF1. *Science* 288, 867-870.
- Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., et Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717-728.
- Hu, X., Malik, S., Negroiu, C.C., Hubbard, K., Velalar, C.N., Hampton, B., Grosu, D., Catalano, J., Roeder, R.G., et Gnatt, A. (2006). A Mediator-responsive form of metazoan RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A* 103, 9506-9511.

- Huber, A.H., Nelson, W.J., et Weis, W.I. (1997). Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* 90, 871-882.
- Hull, M.W., McKune, K., et Woychik, N.A. (1995). RNA polymerase II subunit RPB9 is required for accurate start site selection. *Genes Dev.* 9, 481-490.
- Huttenhofer, A., Schattner, P., et Polacek, N. (2005). Non-coding RNAs: hope or hype? *Trends Genet.* 21, 289-297.
- Hyland, E.M., Cosgrove, M.S., Molina, H., Wang, D., Pandey, A., Cottee, R.J., et Boeke, J.D. (2005). Insights into the role of histone H3 and histone H4 core modifiable residues in *Saccharomyces cerevisiae*. *Mol Cell Biol* 25, 10060-10070.
- Ikura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., et Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* 102, 463-473.
- Imbalzano, A.N., Zaret, K.S., et Kingston, R.E. (1994). Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *J. Biol Chem.* 269, 8280-8286.
- Itin, P.H., Sarasin, A., et Pittelkow, M.R. (2001). Trichothiodystrophy: update on the sulfur-deficient brittle hair syndromes. *J. Am. Acad. Dermatol.* 44, 891-920.
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., et Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A* 98, 4569-4574.
- Izban, M.G. et Luse, D.S. (1992). Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J. Biol. Chem.* 267, 13647-13655.
- Jallow, Z., Jacobi, U.G., Weeks, D.L., Dawid, I.B., et Veenstra, G.J. (2004). Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in *Xenopus*. *Proc. Natl. Acad. Sci. U. S. A* 101, 13525-13530.
- Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., et Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol Cell* 19, 523-534.
- Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., et Smale, S.T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell Biol.* 14, 116-127.
- Jenuwein, T. et Allis, C.D. (2001). Translating the histone code. *Science* 293, 1074-1080.
- Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J., Chabot, B., Poirier, G.G., Hughes, T.R.,

- Blanchette,M., Price,D.H., et Coulombe,B. (2007). Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol Cell* 27, 262-274.
- Jeronimo,C., Langelier,M.F., Zeghouf,M., Cojocaru,M., Bergeron,D., Baali,D., Forget,D., Mnaimneh,S., Davierwala,A.P., Pootoolal,J., Chandy,M., Canadien,V., Beattie,B.K., Richards,D.P., Workman,J.L., Hughes,T.R., Greenblatt,J., et Coulombe,B. (2004). RPAP1, a novel human RNA polymerase II-associated protein affinity purified with recombinant wild-type and mutated polymerase subunits. *Mol Cell Biol* 24, 7043-7058.
- Jiang,Y.W., Veschambre,P., Erdjument-Bromage,H., Tempst,P., Conaway,J.W., Conaway,R.C., et Kornberg,R.D. (1998). Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc. Natl. Acad. Sci. U. S. A* 95, 8538-8543.
- Jin,J., Cai,Y., Li,B., Conaway,R.C., Workman,J.L., Conaway,J.W., et Kusch,T. (2005). In and out: histone variant exchange in chromatin. *Trends Biochem. Sci.* 30, 680-687.
- Jones,J.C., Phatnani,H.P., Haystead,T.A., MacDonald,J.A., Alam,S.M., et Greenleaf,A.L. (2004). C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats. *J. Biol Chem.* 279, 24957-24964.
- Jones,P.A. et Baylin,S.B. (2007). The epigenomics of cancer. *Cell* 128, 683-692.
- Joshi,A.A. et Struhl,K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell* 20, 971-978.
- Jurica,M.S. et Moore,M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* 12, 5-14.
- Juven-Gershon,T., Hsu,J.Y., et Kadonaga,J.T. (2006). Perspectives on the RNA polymerase II core promoter. *Biochem. Soc. Trans.* 34, 1047-1050.
- Kadonaga,J.T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* 116, 247-257.
- Kamakaka,R.T. et Biggins,S. (2005). Histone variants: deviants? *Genes Dev.* 19, 295-310.
- Kampa,D., Cheng,J., Kapranov,P., Yamanaka,M., Brubaker,S., Cawley,S., Drenkow,J., Piccolboni,A., Bekiranov,S., Helt,G., Tammanna,H., et Gingeras,T.R. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* 14, 331-342.
- Kanemaki,M., Kurokawa,Y., Matsu-ura,T., Makino,Y., Masani,A., Okazaki,K., Morishita,T., et Tamura,T.A. (1999). TIP49b, a new RuvB-like DNA helicase, is

included in a complex together with another RuvB-like DNA helicase, TIP49a. *J. Biol. Chem.* *274*, 22437-22444.

Kang, J.J., Auble, D.T., Ranish, J.A., et Hahn, S. (1995). Analysis of the yeast transcription factor TFIIA: distinct functional regions and a polymerase II-specific role in basal and activated transcription. *Mol. Cell Biol.* *15*, 1234-1243.

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

Kephart, D.D., Wang, B.Q., Burton, Z.F., et Price, D.H. (1994). Functional analysis of Drosophila factor 5 (TFIIF), a general transcription factor. *J. Biol. Chem.* *269*, 13536-13543.

Kerwitz, Y., Kuhn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., Schwarz, E., et Wahle, E. (2003). Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. *EMBO J.* *22*, 3705-3714.

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

Kim, B., Nesvizhskii, A.I., Rani, P.G., Hahn, S., Aebersold, R., et Ranish, J.A. (2007). The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc. Natl. Acad. Sci. U. S. A* *104*, 16068-16073.

Kim, J.H., Lane, W.S., et Reinberg, D. (2002). Human Elongator facilitates RNA polymerase II transcription through chromatin. *Proc. Natl. Acad. Sci. U. S. A* *99*, 1241-1246.

Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., et Buratowski, S. (2004a). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J.* *23*, 354-364.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedeau, E., Greenblatt, J.F., et Buratowski, S. (2004b). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* *432*, 517-522.

Kim, T.K., Ebright, R.H., et Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIIH. *Science* *288*, 1418-1422.

Kim, T.K., Lagrange, T., Wang, Y.H., Griffith, J.D., Reinberg, D., et 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.

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

- Kimura,M., Suzuki,H., et Ishihama,A. (2002). Formation of a Carboxy-Terminal Domain Phosphatase (Fcp1)/TFIIF/RNA Polymerase II (pol II) Complex in *Schizosaccharomyces pombe* Involves Direct Interaction between Fcp1 and the Rpb4 Subunit of pol II. *Mol. Cell Biol.* *22*, 1577-1588.
- Klose,R.J. et Bird,A.P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* *31*, 89-97.
- Klose,R.J. et Zhang,Y. (2007). Regulation of histone methylation by demethylination and demethylation. *Nat. Rev. Mol. Cell Biol.* *8*, 307-318.
- Knoepfler,P.S. et Eisenman,R.N. (1999). Sin meets NuRD and other tails of repression. *Cell* *99*, 447-450.
- Kobor,M. et Greenblatt,J. (2002). Regulation of transcription elongation by phosphorylation. *Biochim. Biophys. Acta* *1577*, 261.
- Kobor,M.S., Archambault,J., Lester,W., Holstege,F.C., Gileadi,O., Jansma,D.B., Jennings,E.G., Kouyoumdjian,F., Davidson,A.R., Young,R.A., et Greenblatt,J. (1999). An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in *S. cerevisiae*. *Mol. Cell* *4*, 55-62.
- Kobor,M.S., Simon,L.D., Omichinski,J., Zhong,G., Archambault,J., et Greenblatt,J. (2000). A motif shared by TFIIF and TFIIB mediates their interaction with the RNA polymerase II carboxy-terminal domain phosphatase Fcp1p in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* *20*, 7438-7449.
- Kobor,M.S., Venkatasubrahmanyam,S., Meneghini,M.D., Gin,J.W., Jennings,J.L., Link,A.J., Madhani,H.D., et Rine,J. (2004). A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS. Biol* *2*, E131.
- Koh,S.S., Ansari,A.Z., Ptashne,M., et Young,R.A. (1998). An activator target in the RNA polymerase II holoenzyme. *Mol. Cell* *1*, 895-904.
- Kokubo,T., Swanson,M.J., Nishikawa,J.I., Hinnebusch,A.G., et Nakatani,Y. (1998). The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol. Cell Biol.* *18*, 1003-1012.
- Komarnitsky,P., Cho,E.J., et Buratowski,S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* *14*, 2452-2460.
- Kornberg,R.D. (2005). Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* *30*, 235-239.
- Kornberg,R.D. (2007). The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci. U. S. A* *104*, 12955-12961.

- Kornberg,R.D. et Lorch,Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98, 285-294.
- Kornblihtt,A.R., de la,M.M., Fededa,J.P., Munoz,M.J., et Nogues,G. (2004). Multiple links between transcription and splicing. *RNA*. 10, 1489-1498.
- Kouzarides,T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.
- Kraemer,K.H. (2004). From proteomics to disease. *Nat. Genet.* 36, 677-678.
- Kraemer,S.M., Ranallo,R.T., Ogg,R.C., et Stargell,L.A. (2001). TFIIA interacts with TFIID via association with TATA-binding protein and TAF40. *Mol. Cell Biol.* 21, 1737-1746.
- Kramerov,D.A., Tillib,S.V., Shumyatsky,G.P., et Georgiev,G.P. (1990). The most abundant nascent poly(A) + RNAs are transcribed by RNA polymerase III in murine tumor cells. *Nucleic Acids Res.* 18, 4499-4506.
- Kramerov,D.A. et Vassetzky,N.S. (2005). Short retroposons in eukaryotic genomes. *Int. Rev. Cytol.* 247, 165-221.
- Krishnamurthy,S., He,X., Reyes-Reyes,M., Moore,C., et Hampsey,M. (2004). Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell* 14, 387-394.
- Krogan,N.J., Cagney,G., Yu,H., Zhong,G., Guo,X., Ignatchenko,A., Li,J., Pu,S., Datta,N., Tikuisis,A.P., Punna,T., Peregrin-Alvarez,J.M., Shales,M., Zhang,X., Davey,M., Robinson,M.D., Paccanaro,A., Bray,J.E., Sheung,A., Beattie,B., Richards,D.P., Canadien,V., Lalev,A., Mena,F., Wong,P., Starostine,A., Canete,M.M., Vlasblom,J., Wu,S., Orsi,C., Collins,S.R., Chandran,S., Haw,R., Rilstone,J.J., Gandi,K., Thompson,N.J., Musso,G., St Onge,P., Ghanny,S., Lam,M.H., Butland,G., Altaf-Ul,A.M., Kanaya,S., Shilatifard,A., O'Shea,E., Weissman,J.S., Ingles,C.J., Hughes,T.R., Parkinson,J., Gerstein,M., Wodak,S.J., Emili,A., et Greenblatt,J.F. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440, 637-643.
- Krogan,N.J., Dover,J., Wood,A., Schneider,J., Heidt,J., Boateng,M.A., Dean,K., Ryan,O.W., Golshani,A., Johnston,M., Greenblatt,J.F., et Shilatifard,A. (2003a). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol. Cell* 11, 721-729.
- Krogan,N.J., Keogh,M.C., Datta,N., Sawa,C., Ryan,O.W., Ding,H., Haw,R.A., Pootoolal,J., Tong,A., Canadien,V., Richards,D.P., Wu,X., Emili,A., Hughes,T.R., Buratowski,S., et Greenblatt,J.F. (2003b). A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Mol. Cell* 12, 1565-1576.
- Krogan,N.J., Kim,M., Ahn,S.H., Zhong,G., Kobor,M.S., Cagney,G., Emili,A., Shilatifard,A., Buratowski,S., et Greenblatt,J.F. (2002). RNA polymerase II elongation

factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell Biol.* 22, 6979-6992.

Krogan, N.J., Peng, W.T., Cagney, G., Robinson, M.D., Haw, R., Zhong, G., Guo, X., Zhang, X., Canadien, V., Richards, D.P., Beattie, B.K., Lalev, A., Zhang, W., Davierwala, A.P., Mnaimneh, S., Starostine, A., Tikuisis, A.P., Grigull, J., Datta, N., Bray, J.E., Hughes, T.R., Emili, A., et Greenblatt, J.F. (2004). High-definition macromolecular composition of yeast RNA-processing complexes. *Mol. Cell* 13, 225-239.

Krueger, B.J., Jeronimo, C., Roy, B.B., Bouchard, A., Barrandon, C., Byers, S.A., Searcey, C.E., Cooper, J.J., Bensaude, O., Cohen, E.A., Coulombe, B., et Price, D.H. (2008). LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. *Nucleic Acids Res.*

Kugel, J.F. et Goodrich, J.A. (2002). Translocation after synthesis of a four-nucleotide RNA commits RNA polymerase II to promoter escape. *Mol. Cell Biol.* 22, 762-773.

Kulaeva, O.I., Gaykalova, D.A., et Studitsky, V.M. (2007). Transcription through chromatin by RNA polymerase II: histone displacement and exchange. *Mutat. Res.* 618, 116-129.

Kuldell, N.H. et Buratowski, S. (1997). Genetic analysis of the large subunit of yeast transcription factor IIE reveals two regions with distinct functions. *Mol. Cell Biol.* 17, 5288-5298.

Kuras, L. et Struhl, K. (1999). Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. *Nature* 399, 609-613.

Kwek, K.Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N.J., et Akoulitchev, A. (2002). U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat. Struct. Biol.* 9, 800-805.

Lacadie, S.A., Tardiff, D.F., Kadener, S., et Rosbash, M. (2006). In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. *Genes Dev.* 20, 2055-2066.

Lacoste, N., Utley, R.T., Hunter, J.M., Poirier, G.G., et Cote, J. (2002). Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. *J. Biol. Chem.* 277, 30421-30424.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., et Ebricht, 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.

Landick, R. (2006). A long time in the making--the Nobel Prize for RNA polymerase. *Cell* 127, 1087-1090.

- Langelier, M.F., Trinh, V., et Coulombe, B. (2002). Gros plan sur l'ARN polymérase II. *Medecine Sciences* 18, 211-216.
- Langst, G. et Becker, P.B. (2004). Nucleosome remodeling: one mechanism, many phenomena? *Biochim. Biophys. Acta* 1677, 58-63.
- Larochelle, S., Pandur, J., Fisher, R.P., Salz, H.K., et Suter, B. (1998). Cdk7 is essential for mitosis and for in vivo Cdk-activating kinase activity. *Genes Dev.* 12, 370-381.
- Latham, J.A. et Dent, S.Y. (2007). Cross-regulation of histone modifications. *Nat. Struct. Mol. Biol.* 14, 1017-1024.
- Laybourn, P.J. et Dahmus, M.E. (1989). Transcription-dependent structural changes in the C-terminal domain of mammalian RNA polymerase subunit IIa/o. *J. Biol. Chem.* 264, 6693-6698.
- Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., et Lewis, B.A. (2005). Functional Characterization of Core Promoter Elements: the Downstream Core Element Is Recognized by TAF1. *Mol Cell Biol* 25, 9674-9686.
- Lee, K.K. et Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. *Nat. Rev. Mol. Cell Biol.* 8, 284-295.
- Lee, T.I. et Young, R.A. (1998). Regulation of gene expression by TBP-associated proteins. *Genes Dev.* 12, 1398-1408.
- Lee, T.I. et Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* 34, 77-137.
- Lei, L., Ren, D., Finkelstein, A., et 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.
- Lembo, F., Pero, R., Angrisano, T., Vitiello, C., Iuliano, R., Bruni, C.B., et Chiariotti, L. (2003). MBDin, a novel MBD2-interacting protein, relieves MBD2 repression potential and reactivates transcription from methylated promoters. *Mol Cell Biol* 23, 1656-1665.
- Lemon, B. et Tjian, R. (2000). Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.* 14, 2551-2569.
- Leurent, C., Sanders, S.L., Demeny, M.A., Garbett, K.A., Ruhlmann, C., Weil, P.A., Tora, L., et Schultz, P. (2004). Mapping key functional sites within yeast TFIID. *EMBO J.* 23, 719-727.
- Leuther, K.K., Bushnell, D.A., et Kornberg, R.D. (1996). Two-dimensional crystallography of TFIIB- and IIE-RNA polymerase II complexes: implications for start site selection and initiation complex formation. *Cell* 85, 773-779.

- Lewis, B.A., Kim, T.K., et Orkin, S.H. (2000). A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proc. Natl. Acad. Sci. U. S. A* 97, 7172-7177.
- Li, B., Carey, M., et Workman, J.L. (2007). The role of chromatin during transcription. *Cell* 128, 707-719.
- Li, E., Bestor, T.H., et Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915-926.
- Li, L., He, S., Sun, J.M., et Davie, J.R. (2004a). Gene regulation by Sp1 and Sp3. *Biochem. Cell Biol.* 82, 460-471.
- Li, Q., Peterson, K.R., Fang, X., et Stamatoyannopoulos, G. (2002). Locus control regions. *Blood* 100, 3077-3086.
- Li, Q., Price, J.P., Byers, S.A., Cheng, D., Peng, J., et Price, D.H. (2005). Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. *J. Biol Chem.* 280, 28819-28826.
- Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., Goldberg, D.S., Li, N., Martinez, M., Rual, J.F., Lamesch, P., Xu, L., Tewari, M., Wong, S.L., Zhang, L.V., Berriz, G.F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H.W., Elewa, A., Baumgartner, B., Rose, D.J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S.E., Saxton, W.M., Strome, S., Van Den, H.S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K.C., Harper, J.W., Cusick, M.E., Roth, F.P., Hill, D.E., et Vidal, M. (2004b). A map of the interactome network of the metazoan *C. elegans*. *Science* 303, 540-543.
- Li, S., Ding, B., Chen, R., Ruggiero, C., et Chen, X. (2006). Evidence that the transcription elongation function of Rpb9 is involved in transcription-coupled DNA repair in *Saccharomyces cerevisiae*. *Mol Cell Biol* 26, 9430-9441.
- Li, S. et Smerdon, M.J. (2002). Rpb4 and Rpb9 mediate subpathways of transcription-coupled DNA repair in *Saccharomyces cerevisiae*. *EMBO J.* 21, 5921-5929.
- Li, T., Spearow, J., Rubin, C.M., et Schmid, C.W. (1999a). Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. *Gene* 239, 367-372.
- Li, X.Y. et Green, M.R. (1998). The HIV-1 Tat cellular coactivator Tat-SF1 is a general transcription elongation factor. *Genes Dev.* 12, 2992-2996.
- Li, X.Y., Virbasius, A., Zhu, X., et Green, M.R. (1999b). Enhancement of TBP binding by activators and general transcription factors. *Nature* 399, 605-609.

- Liao,S.M., Zhang,J., Jeffery,D.A., Koleske,A.J., Thompson,C.M., Chao,D.M., Viljoen,M., van Vuuren,H.J., et Young,R.A. (1995). A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature* 374, 193-196.
- Licatalosi,D.D., Geiger,G., Minet,M., Schroeder,S., Cilli,K., McNeil,J.B., et Bentley,D.L. (2002). Functional Interaction of Yeast Pre-mRNA 3' End Processing Factors with RNA Polymerase II. *Mol. Cell* 9, 1101-1111.
- Lieberman,P.M. et Berk,A.J. (1994). A mechanism for TAFs in transcriptional activation: activation domain enhancement of TFIID-TFIIA--promoter DNA complex formation. *Genes Dev.* 8, 995-1006.
- Lim,C.Y., Santoso,B., Boulay,T., Dong,E., Ohler,U., et Kadonaga,J.T. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev.* 18, 1606-1617.
- Lin,P.S., Dubois,M.F., et Dahmus,M.E. (2002). TFIIF-associating Carboxyl-terminal Domain Phosphatase Dephosphorylates Phosphoserines 2 and 5 of RNA Polymerase II. *J. Biol. Chem.* 277, 45949-45956.
- Lindstrom,D.L., Squazzo,S.L., Muster,N., Burckin,T.A., Wachter,K.C., Emigh,C.A., McCleery,J.A., Yates,J.R., III, et Hartzog,G.A. (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell Biol.* 23, 1368-1378.
- Lis,J.T., Mason,P., Peng,J., Price,D.H., et Werner,J. (2000). P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev.* 14, 792-803.
- Listerman,I., Sapra,A.K., et Neugebauer,K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol Biol* 13, 815-822.
- Liu,C.L., Kaplan,T., Kim,M., Buratowski,S., Schreiber,S.L., Friedman,N., et Rando,O.J. (2005). Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol.* 3, e328.
- Liu,M.H., Busch,R.K., Buckley,B., et Reddy,R. (1992). Characterization of antibodies against methyl-pppN cap structure: plant U3 small nucleolar RNA is recognized by these antibodies. *Nucleic Acids Res.* 20, 4299-4304.
- Liu,W.M., Chu,W.M., Choudary,P.V., et Schmid,C.W. (1995). Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.* 23, 1758-1765.
- Lo,K. et Smale,S.T. (1996). Generality of a functional initiator consensus sequence. *Gene* 182, 13-22.

- Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr., et Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta major-globin gene. *Proc. Natl. Acad. Sci. U. S. A* *84*, 8306-8310.
- Lonard, D.M. et O'Malley, B.W. (2005). Expanding functional diversity of the coactivators. *Trends Biochem. Sci.* *30*, 126-132.
- Loppin, B., Bonnefoy, E., Anselme, C., Laurencon, A., Karr, T.L., et Couble, P. (2005). The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. *Nature* *437*, 1386-1390.
- Lorch, Y., Maier-Davis, B., et Kornberg, R.D. (2006). Chromatin remodeling by nucleosome disassembly in vitro. *Proc. Natl. Acad. Sci. U. S. A* *103*, 3090-3093.
- Lotan, R., Bar-On VG, Harel-Sharvit, L., Duek, L., Melamed, D., et Choder, M. (2005). The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes Dev.* *19*, 3004-3016.
- Lu, H., Flores, O., Weinmann, R., et 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.
- Lu, H., Zawel, L., Fisher, L., Egly, J.M., et Reinberg, D. (1992). Human general transcription factor IIIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* *358*, 641-645.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., et Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* *389*, 251-260.
- Luo, W., Johnson, A.W., et Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev.* *20*, 954-965.
- Lusser, A. et Kadonaga, J.T. (2003). Chromatin remodeling by ATP-dependent molecular machines. *Bioessays* *25*, 1192-1200.
- Lykke-Andersen, S. et Jensen, T.H. (2007). Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* *89*, 1177-1182.
- Ma, D., Watanabe, H., Mermelstein, F., Admon, A., Oguri, K., Sun, X., Wada, T., Imai, T., Shiroya, T., Reinberg, D., et . (1993). Isolation of a cDNA encoding the largest subunit of TFIIA reveals functions important for activated transcription. *Genes Dev.* *7*, 2246-2257.
- Maile, T., Kwoczyński, S., Katzenberger, R.J., Wassarman, D.A., et Sauer, F. (2004). TAF1 activates transcription by phosphorylation of serine 33 in histone H2B. *Science* *304*, 1010-1014.

- Malagon,F., Tong,A.H., Shafer,B.K., et Strathern,J.N. (2004). Genetic interactions of DST1 in *Saccharomyces cerevisiae* suggest a role of TFIIS in the initiation-elongation transition. *Genetics* 166, 1215-1227.
- Maldonado,E., Ha,I., Cortes,P., Weis,L., et Reinberg,D. (1990). Factors involved in specific transcription by mammalian RNA polymerase II: role of transcription factors IIA, IID, and IIB during formation of a transcription-competent complex. *Mol. Cell Biol.* 10, 6335-6347.
- Maldonado,E., Shiekhatar,R., Sheldon,M., Cho,H., Drapkin,R., Rickert,P., Lees,E., Anderson,C.W., Linn,S., et Reinberg,D. (1996). A human RNA polymerase II complex associated with SRB and DNA-repair proteins. *Nature* 381, 86-89.
- Malik,H.S. et Henikoff,S. (2003). Phylogenomics of the nucleosome. *Nat. Struct. Biol.* 10, 882-891.
- Malik,S., Gu,W., Wu,W., Qin,J., et Roeder,R.G. (2000). The USA-derived transcriptional coactivator PC2 is a submodule of TRAP/SMCC and acts synergistically with other PCs. *Mol. Cell* 5, 753-760.
- Malik,S. et Roeder,R.G. (2000). Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. *Trends Biochem. Sci.* 25, 277-283.
- Malik,S. et Roeder,R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem. Sci.* 30, 256-263.
- Mandal,S.S., Chu,C., Wada,T., Handa,H., Shatkin,A.J., et Reinberg,D. (2004). Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A* 101, 7572-7577.
- Maniatis,T. et Reed,R. (2002). An extensive network of coupling among gene expression machines. *Nature* 416, 499-506.
- Manley,J.L. (2002). Nuclear coupling: RNA processing reaches back to transcription. *Nat. Struct. Biol.* 9, 790-791.
- Maraia,R.J. et Bayfield,M.A. (2006). The La protein-RNA complex surfaces. *Mol. Cell* 21, 149-152.
- Marfella,C.G. et Imbalzano,A.N. (2007). The Chd family of chromatin remodelers. *Mutat. Res.* 618, 30-40.
- Mariner,P.D., Walters,R.D., Espinoza,C.A., Drullinger,L.F., Wagner,S.D., Kugel,J.F., et Goodrich,J.A. (2008). Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol. Cell* 29, 499-509.

- Marshall,N.F., Peng,J., Xie,Z., et Price,D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* *271*, 27176-27183.
- Marshall,N.F. et Price,D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* *270*, 12335-12338.
- Martens,J.A. et Winston,F. (2002). Evidence that Swi/Snf directly represses transcription in *S. cerevisiae*. *Genes Dev.* *16*, 2231-2236.
- Martens,J.A. et Winston,F. (2003). Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr. Opin. Genet. Dev.* *13*, 136-142.
- Martinez,E., Kundu,T.K., Fu,J., et Roeder,R.G. (1998). A human SPT3-TAFII31-GCN5-L acetylase complex distinct from transcription factor IID. *J. Biol. Chem.* *273*, 23781-23785.
- Mason,P.B. et Struhl,K. (2003). The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Mol. Cell Biol.* *23*, 8323-8333.
- Maston,G.A., Evans,S.K., et Green,M.R. (2006). Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* *7*, 29-59.
- Matlin,A.J., Clark,F., et Smith,C.W. (2005). Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* *6*, 386-398.
- Mattick,J.S. et Makunin,I.V. (2006). Non-coding RNA. *Hum. Mol. Genet.* *15 Spec No 1*, R17-R29.
- Maxon,M.E., Goodrich,J.A., et Tjian,R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes Dev.* *8*, 515-524.
- Maxon,M.E. et Tjian,R. (1994). Transcriptional activity of transcription factor IIE is dependent on zinc binding. *Proc. Natl. Acad. Sci. U. S. A* *91*, 9529-9533.
- McCracken,S., Fong,N., Rosonina,E., Yankulov,K., Brothers,G., Siderovski,D., Hessel,A., Foster,S., Shuman,S., et Bentley,D.L. (1997). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.* *11*, 3306-3318.
- McCracken,S. et Greenblatt,J. (1991). Related RNA polymerase-binding regions in human RAP30/74 and *Escherichia coli* sigma 70. *Science* *253*, 900-902.
- Medenbach,J., Schreiner,S., Liu,S., Luhrmann,R., et Bindereif,A. (2004). Human U4/U6 snRNP recycling factor p110: mutational analysis reveals the function of the tetratricopeptide repeat domain in recycling. *Mol. Cell Biol.* *24*, 7392-7401.

- Meinhart,A., Kamenski,T., Hoepfner,S., Baumli,S., et Cramer,P. (2005). A structural perspective of CTD function. *Genes Dev.* *19*, 1401-1415.
- Meredith,G.D., Chang,W.H., Li,Y., Bushnell,D.A., Darst,S.A., et Kornberg,R.D. (1996). The C-terminal domain revealed in the structure of RNA polymerase II. *J. Mol. Biol.* *258*, 413-419.
- Michels,A.A., Fraldi,A., Li,Q., Adamson,T.E., Bonnet,F., Nguyen,V.T., Sedore,S.C., Price,J.P., Price,D.H., Lania,L., et Bensaude,O. (2004). Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J.* *23*, 2608-2619.
- Michels,A.A., Nguyen,V.T., Fraldi,A., Labas,V., Edwards,M., Bonnet,F., Lania,L., et Bensaude,O. (2003). MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol Cell Biol* *23*, 4859-4869.
- Millar,C.B. et Grunstein,M. (2006). Genome-wide patterns of histone modifications in yeast. *Nat. Rev. Mol. Cell Biol.* *7*, 657-666.
- Millar,C.B., Xu,F., Zhang,K., et Grunstein,M. (2006). Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. *Genes Dev.* *20*, 711-722.
- Mitsiou,D.J. et Stunnenberg,H.G. (2003). p300 is involved in formation of the TBP-TFIIA-containing basal transcription complex, TAC. *EMBO J.* *22*, 4501-4511.
- Mitsuzawa,H., Kanda,E., et Ishihama,A. (2003). Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res.* *31*, 4696-4701.
- Mizuguchi,G., Shen,X., Landry,J., Wu,W.H., Sen,S., et Wu,C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* *303*, 343-348.
- Mizzen,C.A., Yang,X.J., Kokubo,T., Brownell,J.E., Bannister,A.J., Owen-Hughes,T., Workman,J., Wang,L., Berger,S.L., Kouzarides,T., Nakatani,Y., et Allis,C.D. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell* *87*, 1261-1270.
- Molina,C. et Grotewold,E. (2005). Genome wide analysis of Arabidopsis core promoters. *BMC. Genomics* *6*, 25.
- Moore,P.A., Ozer,J., Salunek,M., Jan,G., Zerby,D., Campbell,S., et Lieberman,P.M. (1999). A human TATA binding protein-related protein with altered DNA binding specificity inhibits transcription from multiple promoters and activators. *Mol. Cell Biol.* *19*, 7610-7620.
- Moreland,R.J., Hanas,J.S., Conaway,J.W., et Conaway,R.C. (1998). Mechanism of action of RNA polymerase II elongation factor Elongin. Maximal stimulation of elongation

requires conversion of the early elongation complex to an Elongin-activable form. *J. Biol. Chem.* *273*, 26610-26617.

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

Morlando,M., Greco,P., Dichtl,B., Fatica,A., Keller,W., et Bozzoni,I. (2002). Functional analysis of yeast snoRNA and snRNA 3'-end formation mediated by uncoupling of cleavage and polyadenylation. *Mol. Cell Biol.* *22*, 1379-1389.

Mousson,F., Ochsenbein,F., et Mann,C. (2007). The histone chaperone Asf1 at the crossroads of chromatin and DNA checkpoint pathways. *Chromosoma* *116*, 79-93.

Mueller,C.L. et Jaehning,J.A. (2002). Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol. Cell Biol.* *22*, 1971-1980.

Muller,F., Demeny,M.A., et Tora,L. (2007). New Problems in RNA Polymerase II Transcription Initiation: Matching the Diversity of Core Promoters with a Variety of Promoter Recognition Factors. *J. Biol Chem.* *282*, 14685-14689.

Murphy,S., Di Liegro,C., et Melli,M. (1987). The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* *51*, 81-87.

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

Naar,A.M., Beurang,P.A., Zhou,S., Abraham,S., Solomon,W., et Tjian,R. (1999). Composite co-activator ACR mediates chromatin-directed transcriptional activation. *Nature* *398*, 828-832.

Nagy,Z. et Tora,L. (2007). Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* *26*, 5341-5357.

Nesser,N.K., Peterson,D.O., et Hawley,D.K. (2006). RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. *Proc. Natl. Acad. Sci. U. S. A* *103*, 3268-3273.

Nesvizhskii,A.I., Roos,F.F., Grossmann,J., Vogelzang,M., Eddes,J.S., Grisse,W., Baginsky,S., et Aebersold,R. (2006). Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data: toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides. *Mol. Cell Proteomics.* *5*, 652-670.

Neugebauer,K.M. (2002). On the importance of being co-transcriptional. *J. Cell Sci.* *115*, 3865-3871.

Ng,H.H., Feng,Q., Wang,H., Erdjument-Bromage,H., Tempst,P., Zhang,Y., et Struhl,K. (2002). Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev.* *16*, 1518-1527.

Ng,H.H., Robert,F., Young,R.A., et Struhl,K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell* *11*, 709-719.

Ng,K., Pullirsch,D., Leeb,M., et Wutz,A. (2007). Xist and the order of silencing. *EMBO Rep.* *8*, 34-39.

Nguyen,V.T., Kiss,T., Michels,A.A., et Bensaude,O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* *414*, 322-325.

Ni,Z., Schwartz,B.E., Werner,J., Suarez,J.R., et Lis,J.T. (2004). Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol. Cell* *13*, 55-65.

Nikolov,D.B. et Burley,S.K. (1997). RNA polymerase II transcription initiation: a structural view. *Proc. Natl. Acad. Sci. U. S. A* *94*, 15-22.

Nikolov,D.B., Chen,H., Halay,E.D., Usheva,A.A., Hisatake,K., Lee,D.K., Roeder,R.G., et Burley,S.K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* *377*, 119-128.

Nitta,M., Saijo,M., Kodo,N., Matsuda,T., Nakatsu,Y., Tamai,H., et Tanaka,K. (2000). A novel cytoplasmic GTPase XAB1 interacts with DNA repair protein XPA. *Nucleic Acids Res.* *28*, 4212-4218.

Nogues,G., Munoz,M.J., et Kornblihtt,A.R. (2003). Influence of polymerase II processivity on alternative splicing depends on splice site strength. *J. Biol. Chem.* *278*, 52166-52171.

Nudler,E., Mustaev,A., Lukhtanov,E., et Goldfarb,A. (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* *89*, 33-41.

O'Brien,T. et Lis,J.T. (1993). Rapid changes in *Drosophila* transcription after an instantaneous heat shock. *Mol. Cell Biol.* *13*, 3456-3463.

O'Gorman,W., Thomas,B., Kwek,K.Y., Furger,A., et Akoulitchev,A. (2005). Analysis of U1 small nuclear RNA interaction with cyclin H. *J. Biol. Chem.* *280*, 36920-36925.

Oelgeschlager,T., Chiang,C.M., et Roeder,R.G. (1996). Topology and reorganization of a human TFIID-promoter complex. *Nature* *382*, 735-738.

- 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.
- Ohkuma, Y., Hashimoto, S., Wang, C.K., Horikoshi, M., et 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.
- Ohkuma, Y. et Roeder, R.G. (1994). Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature* *368*, 160-163.
- Ohkuma, Y., Sumimoto, H., Hoffmann, A., Shimasaki, S., Horikoshi, M., et Roeder, R.G. (1991). Structural motifs and potential sigma homologies in the large subunit of human general transcription factor TFIIE. *Nature* *354*, 398-401.
- Ohkuma, Y., Sumimoto, H., Horikoshi, M., et Roeder, R.G. (1990). Factors involved in specific transcription by mammalian RNA polymerase II: purification and characterization of general transcription factor TFIIE. *Proc. Natl. Acad. Sci. U. S. A* *87*, 9163-9167.
- Ohler, U. (2006). Identification of core promoter modules in *Drosophila* and their application in accurate transcription start site prediction. *Nucleic Acids Res.* *34*, 5943-5950.
- Ohler, U., Liao, G.C., Niemann, H., et Rubin, G.M. (2002). Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol* *3*, RESEARCH0087.
- Okamoto, T., Yamamoto, S., Watanabe, Y., Ohta, T., Hanaoka, F., Roeder, R.G., et Ohkuma, Y. (1998). Analysis of the role of TFIIE in transcriptional regulation through structure-function studies of the TFIIEbeta subunit. *J. Biol. Chem.* *273*, 19866-19876.
- Okano, M., Bell, D.W., Haber, D.A., et Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* *99*, 247-257.
- Okuda, M., Watanabe, Y., Okamura, H., Hanaoka, F., Ohkuma, Y., et Nishimura, Y. (2000). Structure of the central core domain of TFIIEbeta with a novel double-stranded DNA-binding surface. *EMBO J.* *19*, 1346-1356.
- Olins, D.E. et Olins, A.L. (2003). Chromatin history: our view from the bridge. *Nat. Rev. Mol Cell Biol* *4*, 809-814.
- Olmsted, V.K., Awrey, D.E., Koth, C., Shan, X., Morin, P.E., Kazanis, S., Edwards, A.M., et Arrowsmith, C.H. (1998). Yeast transcript elongation factor (TFIIS), structure and function. I: NMR structural analysis of the minimal transcriptionally active region. *J. Biol. Chem.* *273*, 22589-22594.

- Orlicky,S.M., Tran,P.T., Sayre,M.H., et Edwards,A.M. (2001). Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J. Biol. Chem.* *276*, 10097-10102.
- Orphanides,G., Lagrange,T., et Reinberg,D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev.* *10*, 2657-2683.
- Orphanides,G., LeRoy,G., Chang,C.H., Luse,D.S., et Reinberg,D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* *92*, 105-116.
- Orphanides,G., Wu,W.H., Lane,W.S., Hampsey,M., et Reinberg,D. (1999). The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* *400*, 284-288.
- Ossipow,V., Tassan,J.P., Nigg,E.A., et Schibler,U. (1995). A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell* *83*, 137-146.
- Otero,G., Fellows,J., Li,Y., de Bizemont,T., Dirac,A.M., Gustafsson,C.M., Erdjument-Bromage,H., Tempst,P., et Svejstrup,J.Q. (1999). Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol. Cell* *3*, 109-118.
- Ozer,J., Moore,P.A., Bolden,A.H., Lee,A., Rosen,C.A., et 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.
- Pabo,C.O. et Sauer,R.T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* *61*, 1053-1095.
- Park,J.M., Werner,J., Kim,J.M., Lis,J.T., et Kim,Y.J. (2001). Mediator, not holoenzyme, is directly recruited to the heat shock promoter by HSF upon heat shock. *Mol. Cell* *8*, 9-19.
- Parvin,J.D., Timmers,H.T., et Sharp,P.A. (1992). Promoter specificity of basal transcription factors. *Cell* *68*, 1135-1144.
- Patel,A.A. et Steitz,J.A. (2003). Splicing double: insights from the second spliceosome. *Nat. Rev. Mol. Cell Biol.* *4*, 960-970.
- Patikoglou,G.A., Kim,J.L., Sun,L., Yang,S.H., Kodadek,T., et Burley,S.K. (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* *13*, 3217-3230.
- Patil,A. et Nakamura,H. (2005). Filtering high-throughput protein-protein interaction data using a combination of genomic features. *BMC. Bioinformatics.* *6*, 100.

- Pei, Y., Hausmann, S., Ho, C.K., Schwer, B., et Shuman, S. (2001). The length, phosphorylation state, and primary structure of the RNA polymerase II carboxyl-terminal domain dictate interactions with mRNA capping enzymes. *J. Biol. Chem.* *276*, 28075-28082.
- Pei, Y., Schwer, B., et Shuman, S. (2003). Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. *J. Biol. Chem.* *278*, 7180-7188.
- Pei, Y. et Shuman, S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* *277*, 19639-19648.
- Peng, J., Zhu, Y., Milton, J.T., et Price, D.H. (1998). Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev.* *12*, 755-762.
- Persengiev, S.P., Zhu, X., Dixit, B.L., Maston, G.A., Kittler, E.L., et Green, M.R. (2003). TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. *Proc. Natl. Acad. Sci. U. S. A* *100*, 14887-14891.
- Perumal, K. et Reddy, R. (2002). The 3' end formation in small RNAs. *Gene Expr.* *10*, 59-78.
- Peterlin, B.M. et Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* *23*, 297-305.
- Pham, A.D. et Sauer, F. (2000). Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in *Drosophila*. *Science* *289*, 2357-2360.
- Phatnani, H.P. et Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* *20*, 2922-2936.
- Pokholok, D.K., Hannett, N.M., et Young, R.A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol Cell* *9*, 799-809.
- Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D.K., et Young, R.A. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* *122*, 517-527.
- Polo, S.E. et Almouzni, G. (2006). Chromatin assembly: a basic recipe with various flavours. *Curr. Opin. Genet. Dev.* *16*, 104-111.
- Price, D.H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell Biol.* *20*, 2629-2634.
- Privalsky, M.L. (2004). The role of corepressors in transcriptional regulation by nuclear hormone receptors. *Annu. Rev. Physiol.* *66*, 315-360.

- Proudfoot,N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr. Opin. Cell Biol* *16*, 272-278.
- Proudfoot,N. et O'Sullivan,J. (2002). Polyadenylation: a tail of two complexes. *Curr. Biol.* *12*, R855-R857.
- Proudfoot,N.J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.* *14*, 105-110.
- Proudfoot,N.J., Furger,A., et Dye,M.J. (2002). Integrating mRNA processing with transcription. *Cell* *108*, 501-512.
- Ptashne,M. (1988). How eukaryotic transcriptional activators work. *Nature* *335*, 683-689.
- Ptashne,M. et Gann,A. (1997). Transcriptional activation by recruitment. *Nature* *386*, 569-577.
- Pugh,B.F. (2000). Control of gene expression through regulation of the TATA-binding protein. *Gene* *255*, 1-14.
- Puig,O., Caspary,F., Rigaut,G., Rutz,B., Bouveret,E., Bragado-Nilsson,E., Wilm,M., et Seraphin,B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* *24*, 218-229.
- Qiu,H., Hu,C., Wong,C.M., et Hinnebusch,A.G. (2006). The Spt4p subunit of yeast DSIF stimulates association of the Paf1 complex with elongating RNA polymerase II. *Mol Cell Biol* *26*, 3135-3148.
- Rabenstein,M.D., Zhou,S., Lis,J.T., et Tjian,R. (1999). TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. *Proc. Natl. Acad. Sci. U. S. A* *96*, 4791-4796.
- Raisner,R.M., Hartley,P.D., Meneghini,M.D., Bao,M.Z., Liu,C.L., Schreiber,S.L., Rando,O.J., et Madhani,H.D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* *123*, 233-248.
- Raisner,R.M. et Madhani,H.D. (2006). Patterning chromatin: form and function for H2A.Z variant nucleosomes. *Curr. Opin. Genet. Dev.* *16*, 119-124.
- Rando,O.J. (2007). Global patterns of histone modifications. *Curr. Opin. Genet. Dev.* *17*, 94-99.
- Ranish,J.A. et Hahn,S. (1991). The yeast general transcription factor TFIIA is composed of two polypeptide subunits. *J. Biol. Chem.* *266*, 19320-19327.
- Ranish,J.A., Hahn,S., Lu,Y., Yi,E.C., Li,X.J., Eng,J., et Aebersold,R. (2004). Identification of TFB5, a new component of general transcription and DNA repair factor IIIH. *Nat. Genet.* *36*, 707-713.

- Ranish, J.A., Yi, E.C., Leslie, D.M., Purvine, S.O., Goodlett, D.R., Eng, J., et Aebersold, R. (2003). The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* *33*, 349-355.
- Reese, J.C. (2003). Basal transcription factors. *Curr. Opin. Genet. Dev.* *13*, 114-118.
- Reina, J.H. et Hernandez, N. (2007). On a roll for new TRF targets. *Genes Dev.* *21*, 2855-2860.
- Reinberg, D., Horikoshi, M., et Roeder, R.G. (1987). Factors involved in specific transcription in mammalian RNA polymerase II. Functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream of the initiation site. *J. Biol. Chem.* *262*, 3322-3330.
- Richmond, T.J. et Davey, C.A. (2003). The structure of DNA in the nucleosome core. *Nature* *423*, 145-150.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., et Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* *17*, 1030-1032.
- Rinke, J. et Steitz, J.A. (1982). Precursor molecules of both human 5S ribosomal RNA and transfer RNAs are bound by a cellular protein reactive with anti-La lupus antibodies. *Cell* *29*, 149-159.
- Rives, A.W. et Galitski, T. (2003). Modular organization of cellular networks. *Proc. Natl. Acad. Sci. U. S. A.*
- Robert, F., Douziech, M., Forget, D., Egly, J.M., Greenblatt, J., Burton, Z.F., et Coulombe, B. (1998). Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. *Mol. Cell* *2*, 341-351.
- Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Letters* *579*, 909-915.
- Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., et Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* *273*, 5858-5868.
- Rosonina, E., Kaneko, S., et Manley, J.L. (2006). Terminating the transcript: breaking up is hard to do. *Genes Dev.* *20*, 1050-1056.
- Rosonina, E. et Manley, J.L. (2005). From Transcription to mRNA: PAF Provides a New Path. *Mol Cell* *20*, 167-168.
- Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D.S., Zhang, L.V., Wong, S.L., Franklin, G., Li, S.,

Albala,J.S., Lim,J., Fraughton,C., Llamosas,E., Cevik,S., Bex,C., Lamesch,P., Sikorski,R.S., Vandenhaute,J., Zoghbi,H.Y., Smolyar,A., Bosak,S., Sequerra,R., Doucette-Stamm,L., Cusick,M.E., Hill,D.E., Roth,F.P., et Vidal,M. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437, 1173-1178.

Rudin,C.M. et Thompson,C.B. (2001). Transcriptional activation of short interspersed elements by DNA-damaging agents. *Genes Chromosomes. Cancer* 30, 64-71.

Runner,V.M., Podolny,V., et Buratowski,S. (2008). The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Mol. Cell Biol.* 28, 1883-1891.

Ryu,S. et Tjian,R. (1999). Purification of transcription cofactor complex CRSP. *Proc. Natl. Acad. Sci. U. S. A* 96, 7137-7142.

Ryu,S., Zhou,S., Ladurner,A.G., et Tjian,R. (1999). The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* 397, 446-450.

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

Saha,A., Wittmeyer,J., et Cairns,B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.* 7, 437-447.

Saha,A., Wittmeyer,J., et Cairns,B.R. (2002). Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev.* 16, 2120-2134.

Sampath,V. et Sadhale,P. (2005). Rpb4 and Rpb7: a sub-complex integral to multi-subunit RNA polymerases performs a multitude of functions. *IUBMB. Life* 57, 93-102.

Sandelin,A., Carninci,P., Lenhard,B., Ponjavic,J., Hayashizaki,Y., et Hume,D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.* 8, 424-436.

Sarma,K. et Reinberg,D. (2005). Histone variants meet their match. *Nat. Rev. Mol. Cell Biol.* 6, 139-149.

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., et Conaway,R.C. (2004). A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Mol Cell* 14, 685-691.

Saunders,A., Core,L.J., et Lis,J.T. (2006). Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol.* 7, 557-567.

- Saunders,A., Werner,J., Andrusis,E.D., Nakayama,T., Hirose,S., Reinberg,D., et Lis,J.T. (2003). Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. *Science* 301, 1094-1096.
- Sawadogo,M. et Roeder,R.G. (1985). Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proc. Natl. Acad. Sci. U. S. A* 82, 4394-4398.
- Schaeffer,L., Moncollin,V., Roy,R., Staub,A., Mezzina,M., Sarasin,A., Weeda,G., Hoeijmakers,J.H., et Egly,J.M. (1994). The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. *EMBO J.* 13, 2388-2392.
- Schaeffer,L., Roy,R., Humbert,S., Moncollin,V., Vermeulen,W., Hoeijmakers,J.H., Chambon,P., et Egly,J.M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* 260, 58-63.
- Schroeder,S.C., Schwer,B., Shuman,S., et Bentley,D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev.* 14, 2435-2440.
- Schug,J., Schuller,W.P., Kappen,C., Salbaum,J.M., Bucan,M., et Stoeckert,C.J., Jr. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* 6, R33.
- Schultz,P., Fribourg,S., Poterszman,A., Mallouh,V., Moras,D., et Egly,J.M. (2000). Molecular structure of human TFIIH. *Cell* 102, 599-607.
- Schwabish,M.A. et Struhl,K. (2004). Evidence for Eviction and Rapid Deposition of Histones upon Transcriptional Elongation by RNA Polymerase II. *Mol Cell Biol* 24, 10111-10117.
- Selleck,W., Howley,R., Fang,Q., Podolny,V., Fried,M.G., Buratowski,S., et Tan,S. (2001). A histone fold TAF octamer within the yeast TFIID transcriptional coactivator. *Nat. Struct. Biol.* 8, 695-700.
- Shahbazian,M.D. et Grunstein,M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annu. Rev. Biochem.* 76, 75-100.
- Shatkin,A.J. et Manley,J.L. (2000). The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol* 7, 838-842.
- Shi,Y. et Whetstine,J.R. (2007). Dynamic regulation of histone lysine methylation by demethylases. *Mol Cell* 25, 1-14.
- Shilatifard,A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu. Rev. Biochem.* 75, 243-269.
- Shilatifard,A., Conaway,R.C., et Conaway,J.W. (2003). The RNA polymerase II elongation complex. *Annu. Rev. Biochem.* 72, 693-715.

- Shim,E.Y., Walker,A.K., Shi,Y., et Blackwell,T.K. (2002). CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the *C. elegans* embryo. *Genes Dev.* *16*, 2135-2146.
- Shimba,S., Buckley,B., Reddy,R., Kiss,T., et Filipowicz,W. (1992). Cap structure of U3 small nucleolar RNA in animal and plant cells is different. gamma-Monomethyl phosphate cap structure in plant RNA. *J. Biol. Chem.* *267*, 13772-13777.
- Shimba,S. et Reddy,R. (1994). Purification of human U6 small nuclear RNA capping enzyme. Evidence for a common capping enzyme for gamma-monomethyl-capped small RNAs. *J. Biol. Chem.* *269*, 12419-12423.
- Shinshi,H., Miwa,M., et Sugimura,T. (1976). Enzyme cleaving the 5'-terminal methylated blocked structure of messenger RNA. *FEBS Lett.* *65*, 254-257.
- Shuman,S. (2001). Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog. Nucleic Acid Res. Mol. Biol.* *66*, 1-40.
- Shumyatsky,G.P., Tillib,S.V., et Kramerov,D.A. (1990). B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5' end. *Nucleic Acids Res.* *18*, 6347-6351.
- Shykind,B.M., Kim,J., et Sharp,P.A. (1995). Activation of the TFIID-TFIIA complex with HMG-2. *Genes Dev.* *9*, 1354-1365.
- Sims,R.J., III, Belotserkovskaya,R., et Reinberg,D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* *18*, 2437-2468.
- Singh,B.N. et Hampsey,M. (2007). A transcription-independent role for TFIIB in gene looping. *Mol. Cell* *27*, 806-816.
- Singh,K., Carey,M., Saragosti,S., et Botchan,M. (1985). Expression of enhanced levels of small RNA polymerase III transcripts encoded by the B2 repeats in simian virus 40-transformed mouse cells. *Nature* *314*, 553-556.
- Singh,R. et Reddy,R. (1989). Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc. Natl. Acad. Sci. U. S. A* *86*, 8280-8283.
- Smale,S.T. et Baltimore,D. (1989). The "initiator" as a transcription control element. *Cell* *57*, 103-113.
- Smale,S.T. et Kadonaga,J.T. (2003). The RNA polymerase II core promoter. *Annu. Rev. Biochem.* *72*, 449-479.
- Smith,D.F. (2004). Tetratricopeptide repeat cochaperones in steroid receptor complexes. *Cell Stress. Chaperones.* *9*, 109-121.

- Smith,T.F., Gaitatzes,C., Saxena,K., et Neer,E.J. (1999). The WD repeat: a common architecture for diverse functions. *Trends Biochem. Sci.* *24*, 181-185.
- Sonna,L.A., Gaffin,S.L., Pratt,R.E., Cullivan,M.L., Angel,K.C., et Lilly,C.M. (2002). Effect of acute heat shock on gene expression by human peripheral blood mononuclear cells. *J. Appl. Physiol* *92*, 2208-2220.
- Sopta,M., Carthew,R.W., et Greenblatt,J. (1985). Isolation of three proteins that bind to mammalian RNA polymerase II. *J. Biol. Chem.* *260*, 10353-10360.
- Spiegelman,B.M. et Heinrich,R. (2004). Biological control through regulated transcriptional coactivators. *Cell* *119*, 157-167.
- Spirin,V. et Mirny,L.A. (2003). Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. U. S. A* *100*, 12123-12128.
- Srinivasan,L. et Atchison,M.L. (2004). YY1 DNA binding and PcG recruitment requires CtBP. *Genes Dev.* *18*, 2596-2601.
- Steen,H. et Mann,M. (2004). The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol Cell Biol* *5*, 699-711.
- Stefanini,M., Vermeulen,W., Weeda,G., Giliani,S., Nardo,T., Mezzina,M., Sarasin,A., Harper,J.I., Arlett,C.F., Hoeijmakers,J.H., et . (1993). A new nucleotide-excision-repair gene associated with the disorder trichothiodystrophy. *Am. J. Hum. Genet.* *53*, 817-821.
- Steitz,T.A. (1998). A mechanism for all polymerases. *Nature* *391*, 231-232.
- Stelzl,U., Worm,U., Lalowski,M., Haenig,C., Brembeck,F.H., Goehler,H., Stroedicke,M., Zenkner,M., Schoenherr,A., Koeppen,S., Timm,J., Mintzlaff,S., Abraham,C., Bock,N., Kietzmann,S., Goedde,A., Toksoz,E., Droege,A., Krobitsch,S., Korn,B., Birchmeier,W., Lehrach,H., et Wanker,E.E. (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell* *122*, 957-968.
- Storz,G., Altuvia,S., et Wassarman,K.M. (2005). An abundance of RNA regulators. *Annu. Rev. Biochem.* *74*, 199-217.
- Strahl,B.D. et Allis,C.D. (2000). The language of covalent histone modifications. *Nature* *403*, 41-45.
- Struhl,K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* *12*, 599-606.
- Struhl,K. (1996). Chromatin structure and RNA polymerase II connection: implications for transcription. *Cell* *84*, 179-182.
- Struhl,K. (2005). Transcriptional activation: mediator can act after preinitiation complex formation. *Mol Cell* *17*, 752-754.

- Sullivan,K.F. (2001). A solid foundation: functional specialization of centromeric chromatin. *Curr. Opin. Genet. Dev.* *11*, 182-188.
- Sumimoto,H., Ohkuma,Y., Sinn,E., Kato,H., Shimasaki,S., Horikoshi,M., et Roeder,R.G. (1991). Conserved sequence motifs in the small subunit of human general transcription factor TFIIE. *Nature* *354*, 401-404.
- Sun,X., Ma,D., Sheldon,M., Yeung,K., et Reinberg,D. (1994). Reconstitution of human TFIIA activity from recombinant polypeptides: a role in TFIID-mediated transcription. *Genes Dev.* *8*, 2336-2348.
- Sun,X., Zhang,Y., Cho,H., Rickert,P., Lees,E., Lane,W., et Reinberg,D. (1998). NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. *Mol. Cell* *2*, 213-222.
- Sun,Z.W., Tessmer,A., et Hampsey,M. (1996). Functional interaction between TFIIB and the Rpb9 (Ssu73) subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* *24*, 2560-2566.
- Sune,C., Hayashi,T., Liu,Y., Lane,W.S., Young,R.A., et Garcia-Blanco,M.A. (1997). CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is involved in Tat-activated human immunodeficiency virus type 1 transcription. *Mol. Cell Biol.* *17*, 6029-6039.
- Svejstrup,J.Q. (2007). Elongator complex: how many roles does it play? *Curr. Opin. Cell Biol.* *19*, 331-336.
- Svejstrup,J.Q. (2004). The RNA polymerase II transcription cycle: cycling through chromatin. *Biochim. Biophys. Acta* *1677*, 64-73.
- Svejstrup,J.Q., Conaway,R.C., et Conaway,J.W. (2006). RNA polymerase II: a "nobel" enzyme demystified. *Mol Cell* *24*, 637-642.
- Svejstrup,J.Q., Li,Y., Fellows,J., Gnatt,A., Bjorklund,S., et Kornberg,R.D. (1997). Evidence for a mediator cycle at the initiation of transcription. *Proc. Natl. Acad. Sci. U. S. A* *94*, 6075-6078.
- Tagami,H., Ray-Gallet,D., Almouzni,G., et Nakatani,Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* *116*, 51-61.
- Takagi,Y., Komori,H., Chang,W.H., Hudmon,A., Erdjument-Bromage,H., Tempst,P., et Kornberg,R.D. (2003). Revised Subunit Structure of Yeast Transcription Factor IIIH (TFIIH) and Reconciliation with Human TFIIH. *J. Biol. Chem.* *278*, 43897-43900.
- Tan,S., Aso,T., Conaway,R.C., et Conaway,J.W. (1994). Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. *J. Biol. Chem.* *269*, 25684-25691.

Tan,S., Hunziker,Y., Sargent,D.F., et Richmond,T.J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* 381, 127-151.

Taverna,S.D., Li,H., Ruthenburg,A.J., Allis,C.D., et Patel,D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* 14, 1025-1040.

Terpe,K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 60, 523-533.

Thompson,C.M. et Young,R.A. (1995). General requirement for RNA polymerase II holoenzymes in vivo. *Proc. Natl. Acad. Sci. U. S. A* 92, 4587-4590.

Tirode,F., Busso,D., Coin,F., et Egly,J.M. (1999). Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol. Cell* 3, 87-95.

Tolhuis,B., Palstra,R.J., Splinter,E., Grosveld,F., et de Laat,W. (2002). Looping and Interaction between Hypersensitive Sites in the Active beta-globin Locus. *Mol. Cell* 10, 1453-1465.

Tora,L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev.* 16, 673-675.

Tremethick,D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* 128, 651-654.

Trinh,V., Langelier,M.F., Archambault,J., et Coulombe,B. (2006). Structural perspective on mutations affecting the function of multisubunit RNA polymerases. *Microbiol. Mol. Biol. Rev.* 70, 12-36.

Turner,B.M. (2002). Cellular memory and the histone code. *Cell* 111, 285-291.

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., et Rothberg,J.M. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627.

Ujvari,A. et Luse,D.S. (2006). RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat. Struct. Mol Biol* 13, 49-54.

Ujvari,A. et Luse,D.S. (2004). Newly Initiated RNA encounters a factor involved in splicing immediately upon emerging from within RNA polymerase II. *J. Biol Chem.* 279, 49773-49779.

Ullu,E., Esposito,V., et Melli,M. (1982). Evolutionary conservation of the human 7 S RNA sequences. *J. Mol. Biol.* 161, 195-201.

Vainberg,I.E., Lewis,S.A., Rommelaere,H., Ampe,C., Vandekerckhove,J., Klein,H.L., et Cowan,N.J. (1998). Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* 93, 863-873.

Van Dyke,M.W., Roeder,R.G., et Sawadogo,M. (1988). Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science* 241, 1335-1338.

Van Herreweghe,E., Egloff,S., Goiffon,I., Jady,B.E., Froment,C., Monsarrat,B., et Kiss,T. (2007). Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. *EMBO J.* 26, 3570-3580.

van Leeuwen,F., Gafken,P.R., et Gottschling,D.E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* 109, 745-756.

Van,M., V, Wery,M., Werner,M., Vandenhoute,J., et Thuriaux,P. (2002). The Rpb9 Subunit of RNA Polymerase II Binds Transcription Factor TFIIE and Interferes with the SAGA and Elongator Histone Acetyltransferases. *J. Biol. Chem.* 277, 10220-10225.

Vermeulen,W., de,B.J., Citterio,E., van Gool,A.J., van der Horst,G.T., Jaspers,N.G., de Laat,W.L., Sijbers,A.M., van der Spek,P.J., Sugasawa,K., Weeda,G., Winkler,G.S., Bootsma,D., Egly,J.M., et Hoeijmakers,J.H. (1997). Mammalian nucleotide excision repair and syndromes. *Biochem. Soc. Trans.* 25, 309-315.

Vilar,J.M. et Saiz,L. (2005). DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr. Opin. Genet. Dev.* 15, 136-144.

Wada,T., Takagi,T., Yamaguchi,Y., Ferdous,A., Imai,T., Hirose,S., Sugimoto,S., Yano,K., Hartzog,G.A., Winston,F., Buratowski,S., et Handa,H. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* 12, 343-356.

Wallace,J.A. et Felsenfeld,G. (2007). We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* 17, 400-407.

Wang,B.Q. et Burton,Z.F. (1995). Functional domains of human RAP74 including a masked polymerase binding domain. *J. Biol. Chem.* 270, 27035-27044.

Wassarman,D.A. et Steitz,J.A. (1991). Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol. Cell Biol.* 11, 3432-3445.

Wassarman,K.M. (2007). 6S RNA: a regulator of transcription. *Mol. Microbiol.* 65, 1425-1431.

Watt,F. et Molloy,P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* 2, 1136-1143.

- Wei,P., Garber,M.E., Fang,S.M., Fischer,W.H., et Jones,K.A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* 92, 451-462.
- Weiner,A.M. (2002). SINEs and LINEs: the art of biting the hand that feeds you. *Curr. Opin. Cell Biol.* 14, 343-350.
- Weiss,M.A. et Narayana,N. (1998). RNA recognition by arginine-rich peptide motifs. *Biopolymers* 48, 167-180.
- Wen,Y. et Shatkin,A.J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.* 13, 1774-1779.
- Wery,M., Shematorova,E., Van Driessche,B., Vandenhaute,J., Thuriaux,P., et Van,M., V (2004). Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIS. *EMBO J.* 23, 4232-4242.
- West,S., Gromak,N., et Proudfoot,N.J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432, 522-525.
- West,S. et Proudfoot,N.J. (2008). Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. *Nucleic Acids Res.* 36, 905-914.
- West,S., Zaret,K., et Proudfoot,N.J. (2006). Transcriptional termination sequences in the mouse serum albumin gene. *RNA.* 12, 655-665.
- Westover,K.D., Bushnell,D.A., et Kornberg,R.D. (2004a). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* 303, 1014-1016.
- Westover,K.D., Bushnell,D.A., et Kornberg,R.D. (2004b). Structural Basis of Transcription; Nucleotide Selection by Rotation in the RNA Polymerase II Active Center. *Cell* 119, 481-489.
- White,R.J., Stott,D., et Rigby,P.W. (1989). Regulation of RNA polymerase III transcription in response to F9 embryonal carcinoma stem cell differentiation. *Cell* 59, 1081-1092.
- Whitehouse,I., Stockdale,C., Flaus,A., Szczelkun,M.D., et Owen-Hughes,T. (2003). Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. *Mol Cell Biol* 23, 1935-1945.
- Whitelaw,E. et Proudfoot,N. (1986). Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *EMBO J.* 5, 2915-2922.

- Will,C.L. et Luhrmann,R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* *13*, 290-301.
- Wind,M. et Reines,D. (2000). Transcription elongation factor SII. *Bioessays* *22*, 327-336.
- Winston,F. et Allis,C.D. (1999). The bromodomain: a chromatin-targeting module? *Nat. Struct. Biol.* *6*, 601-604.
- Wittschieben,B.O., Otero,G., de Bizemont,T., Fellows,J., Erdjument-Bromage,H., Ohba,R., Li,Y., Allis,C.D., Tempst,P., et Svejstrup,J.Q. (1999). A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol. Cell* *4*, 123-128.
- Wolin,S.L. et Cedervall,T. (2002). The La protein. *Annu. Rev. Biochem.* *71*, 375-403.
- Wood,A., Schneider,J., Dover,J., Johnston,M., et Shilatifard,A. (2003). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J. Biol. Chem.* *278*, 34739-34742.
- Woodcock,C.L. (2006). Chromatin architecture. *Curr. Opin. Struct. Biol.* *16*, 213-220.
- Woodcock,C.L., Skoultchi,A.I., et Fan,Y. (2006). Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome. Res.* *14*, 17-25.
- Woychik,N.A. et Hampsey,M. (2002). The RNA Polymerase II Machinery. Structure Illuminates Function. *Cell* *108*, 453-463.
- Wright,K.J., Marr,M.T., et Tjian,R. (2006). TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc. Natl. Acad. Sci. U. S. A* *103*, 12347-12352.
- Wutz,A. et Gribnau,J. (2007). X inactivation Xplained. *Curr. Opin. Genet. Dev.* *17*, 387-393.
- Yamada,T., Yamaguchi,Y., Inukai,N., Okamoto,S., Mura,T., et Handa,H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell* *21*, 227-237.
- Yamaguchi,Y., Takagi,T., Wada,T., Yano,K., Furuya,A., Sugimoto,S., Hasegawa,J., et Handa,H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* *97*, 41-51.
- Yan,Q., Moreland,R.J., Conaway,J.W., et Conaway,R.C. (1999). Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J. Biol. Chem.* *274*, 35668-35675.

- Yang,X., Zaurin,R., Beato,M., et Peterson,C.L. (2007). Swi3p controls SWI/SNF assembly and ATP-dependent H2A-H2B displacement. *Nat. Struct. Mol. Biol.* *14*, 540-547.
- Yang,Z., Yik,J.H., Chen,R., He,N., Jang,M.K., Ozato,K., et Zhou,Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* *19*, 535-545.
- Yang,Z., Zhu,Q., Luo,K., et Zhou,Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* *414*, 317-322.
- Yeo,M., Lin,P.S., Dahmus,M.E., et Gill,G.N. (2003). A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. *J. Biol Chem.* *278*, 26078-26085.
- Yik,J.H., Chen,R., Nishimura,R., Jennings,J.L., Link,A.J., et Zhou,Q. (2003). Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol. Cell* *12*, 971-982.
- Yik,J.H., Chen,R., Pezda,A.C., Samford,C.S., et Zhou,Q. (2004). A human immunodeficiency virus type 1 Tat-like arginine-rich RNA-binding domain is essential for HEXIM1 to inhibit RNA polymerase II transcription through 7SK snRNA-mediated inactivation of P-TEFb. *Mol. Cell Biol.* *24*, 5094-5105.
- Yik,J.H., Chen,R., Pezda,A.C., et Zhou,Q. (2005). Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive positive transcription elongation factor b complexes for control of transcription. *J. Biol Chem.* *280*, 16368-16376.
- Yokomori,K., Verrijzer,C.P., et Tjian,R. (1998). An interplay between TATA box-binding protein and transcription factors IIE and IIA modulates DNA binding and transcription. *Proc. Natl. Acad. Sci. U. S. A* *95*, 6722-6727.
- Yokomori,K., Zeidler,M.P., Chen,J.L., Verrijzer,C.P., Mlodzik,M., et Tjian,R. (1994). Drosophila TFIIA directs cooperative DNA binding with TBP and mediates transcriptional activation. *Genes Dev.* *8*, 2313-2323.
- Yuan,G.C., Liu,Y.J., Dion,M.F., Slack,M.D., Wu,L.F., Altschuler,S.J., et Rando,O.J. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* *309*, 626-630.
- Yudkovsky,N., Ranish,J.A., et Hahn,S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature* *408*, 225-229.
- Zaret,K.S. et Sherman,F. (1982). DNA sequence required for efficient transcription termination in yeast. *Cell* *28*, 563-573.

- Zawel,L., Kumar,K.P., et Reinberg,D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev.* 9, 1479-1490.
- Zeng,L. et Zhou,M.M. (2002). Bromodomain: an acetyl-lysine binding domain. *FEBS Lett.* 513, 124-128.
- Zhang,H., Roberts,D.N., et Cairns,B.R. (2005a). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* 123, 219-231.
- Zhang,J. et Corden,J.L. (1991). Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *J. Biol Chem.* 266, 2297-2302.
- Zhang,R., Poustovoitov,M.V., Ye,X., Santos,H.A., Chen,W., Daganzo,S.M., Erzberger,J.P., Serebriiskii,I.G., Canutescu,A.A., Dunbrack,R.L., Pehrson,J.R., Berger,J.M., Kaufman,P.D., et Adams,P.D. (2005b). Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. *Dev. Cell* 8, 19-30.
- Zhang,Z., Fu,J., et Gilmour,D.S. (2005c). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev.* 19, 1572-1580.
- Zhao,H., Wang,Q., Zhang,H., Liu,Q., Du,X., Richter,M., et Greene,M.I. (2005). UXT is a novel centrosomal protein essential for cell viability. *Mol. Biol. Cell* 16, 5857-5865.
- Zhao,J., Hyman,L., et Moore,C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol Biol Rev.* 63, 405-445.
- Zheng,L., Hoeflich,K.P., Elsby,L.M., Ghosh,M., Roberts,S.G., et Ikura,M. (2004). FRET evidence for a conformational change in TFIIB upon TBP-DNA binding. *Eur. J. Biochem.* 271, 792-800.
- Zhou,H., Spicuglia,S., Hsieh,J.J., Mitsiou,D.J., Hoiby,T., Veenstra,G.J., Korsmeyer,S.J., et Stunnenberg,H.G. (2006). Uncleaved TFIIA is a substrate for taspase 1 and active in transcription. *Mol Cell Biol* 26, 2728-2735.
- Zhou,Q., Chen,D., Pierstorff,E., et Luo,K. (1998). Transcription elongation factor P-TEFb mediates Tat activation of HIV-1 transcription at multiple stages. *EMBO J.* 17, 3681-3691.
- Zhou,Q. et Sharp,P.A. (1996). Tat-SF1: cofactor for stimulation of transcriptional elongation by HIV-1 Tat. *Science* 274, 605-610.

- Zhou,Q. et Yik,J.H. (2006). The Yin and Yang of P-TEFb Regulation: Implications for Human Immunodeficiency Virus Gene Expression and Global Control of Cell Growth and Differentiation. *Microbiol. Mol. Biol. Rev.* 70, 646-659.
- Zhou,Y., Kou,H., et Wang,Z. (2007). Tfb5 interacts with Tfb2 and facilitates nucleotide excision repair in yeast. *Nucleic Acids Res.* 35, 861-871.
- Zhu,W., Zeng,Q., Colangelo,C.M., Lewis,M., Summers,M.F., et Scott,R.A. (1996). The N-terminal domain of TFIIB from *Pyrococcus furiosus* forms a zinc ribbon. *Nat. Struct. Biol.* 3, 122-124.
- Zieve,G., Benecke,B.J., et Penman,S. (1977). Synthesis of two classes of small RNA species in vivo and in vitro. *Biochemistry* 16, 4520-4525.
- Zorio,D.A. et Bentley,D.L. (2004). The link between mRNA processing and transcription: communication works both ways. *Exp. Cell Res.* 296, 91-97.
- Zurita,M. et Merino,C. (2003). The transcriptional complexity of the TFIID complex. *Trends Genet.* 19, 578-584.