



THESE

Pour obtenir le grade de

Docteur de l'Université de Strasbourg

Ecole Doctorale des sciences de la Vie et de la Santé

Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

Présentée par

Kambiz NADERI

Contribution à la détermination de la structure de la sous-unité RPB11 de l'ARN polymérase II humaine

Soutenue publiquement le 23 septembre 2011

Membres du Jury :

Directeur de Thèse : Dr. Marc VIGNERON, IREBS, Strasbourg

Rapporteur Externe : Pr. Roger G. CLERC, PRBD, Basel

Rapporteur Externe : Pr. Martin TEICHMANN, IECB, Bordeaux

Examineur : Dr. Dominique BURNOUF, IBMC, Strasbourg

Membre Invité : Pr. Claude KEDINGER, IREBS, Strasbourg

Thèse préparée au sein de l'équipe Modifications post-traductionnelles et cancérogenèse
Institut de Recherche de l'Ecole Supérieure de Biotechnologie de Strasbourg
CNRS-UMR 7242

A mes parents

Remerciements

Mes premiers remerciements s'adressent à mon jury, tout d'abord au **Pr. Roger G. Clerc** et au **Pr. Martin Teichmann** d'avoir accepté d'être rapporteurs de ce manuscrit, ainsi qu'au **Dr. Dominique Burnouf** d'être examinateur de mon travail de thèse.

Je souhaite remercier le **Ministère de la Recherche** et l'**Université de Strasbourg** de m'avoir octroyé les financements pour la réalisation de ma thèse.

J'adresse mes sincères remerciements aux **Pr. Claude Kédinger** et **Dr. Jean-Luc Galzi** pour m'avoir accueilli au sein de l'**IREBS**. Je remercie tout particulièrement le **Pr. Claude Kédinger** pour son soutien dans toutes mes démarches malgré ses nombreuses occupations.

Tous mes remerciements au **Dr. Marc Vigneron**, mon directeur de thèse, pour sa gentillesse, sa grande disponibilité, sa culture et ses connaissances scientifiques. Je lui exprime toute ma gratitude pour le soutien, la liberté et la confiance qu'il m'a accordés.

Je remercie sincèrement **Pr. Bruno Chatton** pour son aide précieuse qu'il m'a apportée durant ces quatre années et particulièrement ses qualités humaines.

Je remercie également tous les membres de l'équipe : **Mariel Donzeau** pour ta sympathie et nos discussions fructueuses, **Etienne** pour ton humour, les rires, la musique (je te souhaite bon courage pour ta thèse) et la dernière arrivée **Raquel**. Je te souhaite bonne chance pour ton post-doc.

J'ai également une pensée reconnaissante envers anciens membres de l'équipe **Barbara, Jessica, Ilaria, Adonis, Arnaud** et **Patryk**, qui ont contribué à la bonne ambiance et au merveilleux environnement dans lequel j'ai eu la chance d'évoluer.

Merci **Barbara** pour ta gentillesse, pour me remonter le moral durant cette thèse, pour ton aide lors de mes commandes et pour les pauses café que nous avons partagées.

J'exprime ma reconnaissance aux stagiaires **Audrey** et **Nassim** qui m'ont aidé au cours de ces travaux.

Je tiens à remercier l'ensemble des chercheurs de l'**IREBS** que j'ai eu le plaisir de côtoyer de près ou loin dans cet institut. En particulier, **Pr. Etienne Weiss**, **Dr. Georges Orfanoudakis**, **Dr. Gilles Travé** et **Dr. Renaud Wagner** pour le matériel mis à disposition.

*Je tiens à remercier nos collaborateurs, les **Dr. David Cobessi, Dr. Bruno Klaholz, Dr. Christian D. Muller, Dr. Sankar N. Manicka, Pierre Poussin et Pascal Kessler** pour leurs soutiens et leurs nombreux conseils.*

*Je remercie les **Dr. Patrick Schultz et Dr. Irwin Davidson** d'avoir accepté de faire partie du comité de mi-thèse et d'avoir bien voulu juger de ce travail à mi-parcours.*

*Merci également aux membres des différentes équipes pour leur sympathie : **Anne-Sophie, Audrey, Xavier, Sadek, Anaëlle, Guillaume, Célia, Elise, Thomas, Mike, Christian, Olivier .H, Nadège, Khaled, Karl, Olivier .B, Christelle**, sans lesquels ces années de thèse n'auraient pas eu la même saveur !*

*Merci **Sébastien** pour ton aide précieuse et nos discussions scientifiques et aussi pour les agréables moments que nous avons partagés à Marseille.*

*Je veux aussi dire toute mon amitié et mon affection à **Laurent, Vinod, Sabrina, Laure, Bénédicte, Isabelle, Claire, Mélanie**. Je n'oublie pas les bons moments sympas passés ensemble.*

*Merci au service administratif, aux gestionnaires et au service technique pour leur précieuse aide et leur efficacité : **Pascal Puisoye, Claire Guth, Sylvie Douessin, Brigitte Daul, Monique Maurer, Sabine Krieger, Raphaël Poinignon, Jacqui Probst, Augustine Sado**.*

*Je veux longuement remercier l'ensemble de **ma famille (mon frère et ma soeur)** et **ma belle-famille** particulièrement mon beau frère **Kamran** qui ont toujours été à mon côté, pour leur soutien qui a contribué à la réussite de ce travail, et m'ont permis de garder le moral dans les moments difficiles.*

*Je suis immensément reconnaissant envers **mes parents** sans qui je n'aurais pas pu faire de si longues études. Ils m'ont toujours soutenu et encouragé pour aller jusqu'au bout de mes ambitions malgré la distance qui nous sépare. Ils m'ont fait confiance et m'ont donné confiance en moi. Que cette thèse soit le fruit de votre patience et de l'éducation que vous m'avez offert. Les mots sont faibles, j'espère seulement que vous êtes fiers de moi.*

*Enfin, ces remerciements ne seraient pas complets sans mon épouse, **Salomeh**, qui m'a toujours soutenu et encouragé. Salomeh, par sa compréhension, son intelligence, sa patience et son amour a su créer un environnement dans lequel j'ai pu accomplir cette lourde tâche. Elle m'a supporté pendant ces années, et cela n'était pas facile. J'ai bien conscience des sacrifices consentis et je ne l'oublierai jamais.*

Tables des Matières

Listes des Figures.....	8
Listes des tableaux.....	11
Listes des abréviations.....	12
Introduction.....	15
I. Les différentes ARN polymérases.....	17
I.1. Les ARN polymérases monomérique.....	17
I.2. Les ARN polymérases des cellules procaryotes.....	18
I.2.1. L'ARN polymérase bactérienne.....	18
I.2.2. L'ARN polymérase des archéobactéries.....	20
I.3. Les ARN polymérases des cellules eucaryotes.....	23
I.3.1. Les trois ARN polymérases (I, II, III) nucléaires.....	24
I.3.2. L'ARN polymérase mitochondriale.....	25
I.3.3. L'ARN polymérase chloroplastique.....	26
I.3.4. Les ARN polymérases IV et V.....	27
II. Comparaison des ARN polymérases.....	28
II.1. Conservation et spécificité des sous-unités des ARN polymérases.....	28
II.2. Complémentation entre les sous-unités de l'ARNP II humaine et <i>S. cerevisiae</i>	30
III. La transcription par l'ARN polymérase II.....	31
III.1. La transcription des gènes de classe II.....	31
III.1.1. Le promoteur minimal.....	31
III.1.2. Les éléments de régulation.....	33
III.1.2.1. Les séquences régulatrices proximales.....	34
III.1.2.2. Les séquences régulatrices distales.....	35
III.1.3. Les facteurs généraux de la transcription.....	36
III.1.4. Le complexe médiateur.....	38
III.2. Initiation de la transcription.....	40
III.3. Elongation de la transcription.....	41
III.4. Terminaison de la transcription.....	43
IV. ARN polymérase II.....	44
IV.1. Organisation de l'ARN polymérase II.....	44

IV.2. Les sous-unités de l'ARN polymérase II	44
IV.2.1. Les sous-unités RPB1 (220 kDa) et RPB2 (140 kDa)	44
Le domaine Carboxy-terminal de RPB1 (CTD)	47
IV.2.2. Les sous-unités RPB3 (33 kDa) et RPB11 (13,3 kDa)	50
IV.2.3. Les sous-unités RPB4 (16 kDa) et RPB7 (19 kDa)	52
IV.2.4. La sous-unité RPB5 (25 kDa)	54
IV.2.5. La sous-unité RPB6 (14,4 kDa)	55
IV.2.6. La sous-unité RPB8 (17,1 kDa)	56
IV.2.7. La sous-unité RPB9 (14,5 kDa)	57
IV.2.8. La sous-unité RPB10 (7,6 kDa)	59
IV.2.9. La sous-unité RPB12 (7 kDa)	60
IV.3. Structure tridimensionnelle de l'ARN Polymérase II	61
IV.4. L'assemblage de l'ARN polymérase II	65
V. La sous-unité HsRPB11	67
V.1. Organisation du gène et isoformes	67
V.2. L'épissage alternatif de messenger HsRPB11 et l'expression tissu-spécifique	68
V.3. Propriétés des trois protéines HsRPB11	71
VI. Objectif du travail de thèse	74
Résultats et Discussion	75
Chapitre 1 : Etude structurale de la sous-unité HsRPB11a par cristallographie	76
I. Clonage, expression, purification et cristallisation de HsRPB11a	77
I.1. Clonage	77
I.2. Expression chez la cyanobactérie genre Anabaena et chez E.coli BL21 (DE3)	77
I.3. Purification des protéines	79
I.4. Criblage des conditions de cristallisation	82
I.5. Optimisation des conditions de cristallisation	84
I.6. Cryoprotection des cristaux et diffraction par les rayons X	88
I.7. Construction du premier modèle	89
II. Résolution du problème de la phase	89
II.1. Remplacement isomorphe multiple	89
II.2. Analyse par spectrométrie de masse	91
III. Amélioration des conditions de cristallisation	91
IV. Autres approches / nouvelles construction	94

IV.1. cristallisation de la protéine sans étiquette 6His	94
IV.2. Cristallisation de la protéine de fusion MBP-HsRPB11a.....	95
IV.2.1. Expression et purification de la protéine de fusion MBP-HsRPB11a.....	95
IV.2.2. Criblage des conditions de cristallisation	97
IV.2.3. Analyse par diffusion dynamique de la lumière (DLS).....	97
IV.3. Cristallisation de mutants de délétion de HsRPB11a	98
IV.3.1. Les différentes constructions	98
IV.3.2. Expression et purification des protéines	99
IV.3.3. Analyse par spectrométrie de masse.....	101
IV.3.4. Criblage des conditions de cristallisation	102
IV.3.5. Optimisation des conditions de cristallisation.....	102
IV.3.6. Recherche des conditions cryoprotectantes et diffraction par les rayons X.....	103
V. Discussion et perspectives	104
Chapitre 2 : Analyse des interactions de HsRPB11a.....	107
I. Analyse de l'homodimérisation de la sous-unité HsRPB11a	108
I.1. Analyse de l'homodimérisation par double-hybride chez <i>S. cerevisiae</i>	108
I.2. Analyse comparative de l'homodimérisation et de l'hétérodimérisation de HsRPB11a avec HsRPB3	110
I.3. Analyse de l'homodimérisation par FRET dans les cellules humaines	113
I.3.1. Observation des protéines surexprimées au microscope	116
I.3.2. FRET par photoblanchiment (pbFRET).....	118
I.4. Analyse de l'homodimérisation par co-immunoprécipitation	118
II. Discussion et perspectives	120
Chapitre 3 : Recherche de l'isoforme HsRPB11bα	124
I. Stratégie.....	125
I.1. Détection de l'ARNm des différentes isoformes.....	125
I. 2. Analyse des effets de différents siRNAs par qRT-PCR	126
II. Discussion et perspectives	128
Discussion générale	129
ANNEXE	134
Références bibliographiques	161

Listes des Figures

Figure 1 : Modèle de la structure modulaire de la polymérase monomérique.	17
Figure 2 : Structure cristallographique de l'enzyme « core » de l'ARNP de <i>Thermus aquaticus</i>	19
Figure 3 : Comparaison des deux formes de la polymérase bactérienne.	20
Figure 4 : Structure de surface de l'ARNP de <i>Sulfolobus solfataricus</i>	21
Figure 5 : Structure cristallographique de l'ARNP de <i>Sulfolobus solfataricus</i>	22
Figure 6 : Reconnaissance des éléments de promoteur par les facteurs TFIID et TFIIB.	33
Figure 7 : Représentation schématique des éléments régulateurs de l'expression des gènes.	34
Figure 8 : Production de eARN à partir d'enhancer.	35
Figure 9 : Différentes formes de médiateur.	39
Figure 10 : Structure du complexe Pol II/médiateur.	39
Figure 11 : Les modèles d'assemblage du PIC.	41
Figure 12 : Pause de l'ARNP II et sa libération de promoteur proximal.	42
Figure 13 : Structure cristallographique de la sous-unité RPB1.	45
Figure 14 : Structure cristallographique de la sous-unité RPB2.	46
Figure 15 : Schématisation des modifications post-traductionnelles du CTD au cours de la transcription.	48
Figure 16 : Structures de complexes RPB4/7.	53
Figure 17 : Structure cristallographique de la sous-unité RPB5 de <i>S.cerevisiae</i>	54
Figure 18 : Structure de la sous-unité HsRPB6 établie par résonance magnétique nucléaire.	55
Figure 19 : Structure tridimensionnelle de HsRPB8.	57
Figure 20 : Comparaison des structures des parties C-terminales de la sous-unité M (RPB9) et TFIIB d'archéobactérie et TFIIS humain.	58
Figure 21: Structure de la sous-unité RPB10 de l'archéobactérie <i>Methanobacterium thermoautotrophicum</i>	60
Figure 22 : Structure de l'ARNP II complète.	62
Figure 23 : Structure cristallographique de l'ARNP II de <i>S. cerevisiae</i>	63
Figure 24 : Localisation de l' α -amanitine fixée sur l'ARNP II.	64
Figure 25 : Assemblage et transport de l'ARNP II.	66
Figure 26 : Organisation génomique de HsRPB11.	67

Figure 27 : Structure de l'ARN messager et la séquence peptidique de HsRPB11.	69
Figure 28 : Structure de la sous-unité RPB3 et RPB11 de la levure et l'homme.....	71
Figure 29 : Résumé des contacts identifiés entre les sous-unités RPB3 et RPB11 chez l'homme et chez la levure et leurs contributions à la stabilité des hétérodimères.....	73
Figure 30 : Analyse SDS-PAGE de l'expression des protéines chez <i>E.coli</i> BL21(DE3).	78
Figure 31 : Etape de chromatographie d'affinité sur Ni-NTA suivie par SDS-PAGE.	79
Figure 32 : Chromatogramme de gel-filtration.....	80
Figure 33 : L'analyse de la pureté des échantillons obtenus après gel-filtration.	81
Figure 34 : Le méthode de la diffusion de vapeur en goutte suspendue et assise.....	82
Figure 35 : Microcristaux obtenus à l'aide de kits de QIAGEN.	83
Figure 36 : Optimisation de condition de cristallisation par variation de la concentration du PEG.....	84
Figure 37 : Optimisation de condition de cristallisation par variation de la concentration du sel.....	85
Figure 38 : Optimisation de condition de cristallisation par variation de pH.	86
Figure 39 : Optimisation de condition de cristallisation en fonction de la concentration de protéines.....	87
Figure 40 : Modèle approximatif de la structure de HsRPB11a.	89
Figure 41 : l'essai de co-cristallisation des protéines avec des métaux lourds.....	90
Figure 42 : Les échantillons préparés pour être analysés en spectrométrie de masse.....	91
Figure 43 : Analyse de produit de méthylation de HsRPB11a par MALDI-TOF.	93
Figure 44 : Analyse de la digestion des protéines par la protéase TEV.	95
Figure 45 : Purification de protéines de fusion MBP-HsRPB11a.....	96
Figure 46 : Analyse de l'homogénéité structurale de la protéine par DLS.....	98
Figure 47 : Présentation des mutants de délétion de HsRPB11a.....	99
Figure 48 : Chromatogramme de gel-filtration et analyse des fractions par électrophorèse.	100
Figure 49 : Analyse des protéines purifiée en gel SDS-PAGE	101
Figure 50 : Nouvelle forme cristalline obtenue à partir des mutants de délétion.	103
Figure 51: Étude de l'interaction protéine-protéine dans la levure par double-hybride.....	110
Figure 52 : Analyse de l'interaction proteine-proteine par double-hybride dans la levure.	112
Figure 53 : Analyse de l'expression de la protéine de fusion par western blot.....	114
Figure 54 : Couple de FRET eGFP/TMR.	115
Figure 55 : Localisation subcellulaire des protéines de fusion.....	117

Figure 56 : Analyse de l'homodimérisation par co-immunoprécipitation.....	119
Figure 57 : Comparaison des interactions HsRPB11a/HsRPB11a et ScRPB3/ScRPB11.	122
Figure 58 : Positions des oligonucléotides utilisés sur l'ARN messager des isoforme HsRPB11.	125
Figure 59 : Identification des ARN messagers correspondant aux deux isoformes de HsRPB11 par RT-PCR.	126
Figure 60 : Analyse de la fonctionnalité des siRNA dirigés contre le gène POLR2j1 par qRT-PCR.....	127

Listes des tableaux

Tableau 1 : Comparaison des sous-unités de l'ARNP II, IV et V.	27
Tableau 2 : Classification des sous-unités des ARNP.	28
Tableau 3 : Comparaison des résultats de complémentation Homme-levure.	30
Tableau 4 : Liste des séquences consensus et positions des éléments de promoteur.	33
Tableau 5 : Composition et propriétés des facteurs généraux de transcription.	37
Tableau 6 : Présentation des conditions obtenues lors de premier essai de cristallisation.	83
Tableau 7 : Présentation des nouvelles conditions de cristallisation	94
Tableau 8 : Masses des polypeptides mesurées par MALDI-TOF.	101
Tableau 9 : Conditions obtenues lors des premiers essais des cristallisation des mutants de délétion.	102
Tableau 10 : Liste des couples d'oligonucléotides utilisés en RT-PCR.	125
Tableau 11 : Liste des séquences des siRNA utilisés	126

Listes des abréviations

3D	Tridimensionnel
Å	Angström
ADN	Acide Désoxyribonucléique
ADNc	Acide Désoxyribonucléique Complémentaire
ADNmt	ADN mitochondrial
ARC	Activator Recruited Cofactor
ARN	Acide Ribonucléique
ARNm	Acide Ribonucléique Messenger
ARNr	Acide Ribonucléique Ribosomal
ARNt	Acide Ribonucléique de Transfert
ARNP	ARN polymérase ADN dépendante
ARNP I	ARN polymérase I ADN dépendante
ARNP II	ARN polymérase II ADN dépendante
ARNP III	ARN polymérase III ADN dépendante
ATF	Activating Transcription Factor
ATG	ATG Codon
ATP	Adenosine Tri-Phosphate
BARD1	BRCA1-Associated RING Domain protein 1
BRE	TFIIB Responsive Element
BRCA1	Breast Cancer 1
CARM1	Coactivator-Associated arginine Methyltransferas
CCP4	Collaborative Computational Project No.4
CDK7	Cycline Dependent Kinase 7
CHC	Carcinome Hépatocellulaire
ChIP	Chromatin ImmunoPrecipitation
CPF	Cleavage/Polyadenylation Factor
CRSP	Cofactor Required for SP1 activation
CTD	Carboxy Terminal Domain
CTF	CAAT box Transcription Factor
Da	Dalton
DCE	Downstream Core Element
DLS	Diffusion Light Scattering
DO	Densité Optique
DPE	Downstream Promoter Element
DRIP	vitamin D Receptor Interacting Proteins
DRB	5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole
DSIF	DRB Sensitivity Inducing Factor
DST	Downstream element
DTT	Dithiothréitol
eARN	Enhancer Acide Ribonucléique
<i>E. coli</i>	<i>Escherichia coli</i>
eGFP	Enhanced Green Fluorescent Protein
ELL	Elevenineteen Lysine-rich in Leukemia
GC	Guanine-Cytosine
GTF	General Transcription Factors
GTP	Guanosine Tri-Phosphate
HATs	Histone Acetyl Transferase
HDAC	Histone Deacetylase
hMed	Human Mediator
hPMS2L	Human Post-Meiotic Segregation 2 Like
HPRT	Hypoxanthine-guanine PhosphoRibosyl Transferase

HsRBP	Homo Sapiens RNA Polymerase B
HSP	Heat-Shock Protein
Inr	Initiator element
IPTG	IsoPropyl β -D-1-ThioGalactopyranoside
kDa	kiloDalton
LEU	Leucine
MAD	Multiwavelength Anomalous Diffraction
MBP	Maltose-Binding-Protein
miARN	Micro Acide Ribonucléique
MIR	Remplacement Isomorphe Multiple
MPD	Methyl PropaneDiol
NAT	Negative regulator of Activated Transcription
Ndr1	Non-race-specific Disease Resistance
NEB	Nuclear Encoded RNA Polymerase
NELF	Negativ Elongation Factor
NES	Nuclear Export Signal
Ni-NTA	Nickel-NitriloTriacetic Acid
NLS	Nuclear Localization Signal
NTD	N-Terminal Domain
NTP	Nucléosides Tri-Phosphates
OB	Oligonucleotide/Oligosaccharide Binding
Pb	Paire de Base
PC2	Positive Cofactor 2
PEG	Poly Ethylène Glycol
PEP	Plastid Encoded RNA Polymerase
<i>P. falciparum</i>	<i>Plasmodium falciparum</i>
PfRBP	<i>Plasmodium falciparum</i> RNA Polymerase B
PIC	Pre-Initiation Complex
POLR2J	Polymerase (RNA) II (DNA directed) polypeptide J
RPA	RNA Polymerase A sous-unité
RPB	RNA Polymerase B sous-unité
RPC	RNA Polymerase C sous-unité
RT-PCR	Reverse Transcription Polymerase Chain Reaction
ScRPB	<i>Saccharomyces Cerevisiae</i> RNA Polymerase B
<i>S. cerevisiae</i>	<i>Saccharomyces Cerevisiae</i>
SDS-PAGE	Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
Ser	Serine
siRNA	Short Interfering RNA
SMCC	SRB/MED-containing Cofactor Complex
snRNA	Small Nuclear Acide Ribonucléique
snoRNA	Small Nucleolar Acide Ribonucléique
<i>S. pombe</i>	<i>Saccharomyces pombe</i>
SRB	Suppressors of RNA polymerase B
<i>S.solfataricus</i>	<i>Sulfolobus Solfataricus</i>
Sp1	Specificity Protein 1
TAF	TBP Associated Factors
TAP	Tandem Affinity Purification
TBP	TATA Binding Protein
TCR	Transcription Coupled Repair
TEV	Tobacco Etch Virus
TFII	Transcription Factor of RNA polymerase II

TMR	TetraMethylRhodamine
<i>Th. aquaticus</i>	<i>Thermus aquaticus</i>
TRAP	Thyroid hormone Receptor Associated Proteins
TRP	Tryptophan
Tris	Tris hydroxyméthyl-aminométhane
TSS	Transcription Start Site
UAS	Upstream Activating Sequence
URS	Upstream Repressing Sequence
VIH-1	Virus d'Immunodéficience Humain de type I

Introduction

Introduction générale

La première étape de l'expression des gènes est la transcription qui permet l'expression de l'information génétique contenue dans l'ADN. Chez les eucaryotes et les procaryotes, la transcription est un mécanisme déterminant à tous les stades de la croissance cellulaire : de la division à la mort. Elle est essentielle à l'expression des gènes et sa régulation permet le contrôle et la modulation de la concentration intracellulaire des ARN et des différentes protéines nécessaires à la survie cellulaire. Sa régulation s'effectue par le biais de nombreux facteurs agissant au travers d'interactions protéine-protéine et protéine-ADN. La transcription dépend d'une activité enzymatique médiée par une ARN polymérase ADN dépendante (ARNP) qui catalyse de manière processive la polymérisation de ribonucléotides triphosphates pour synthétiser une chaîne d'ARN complémentaire à la matrice d'ADN.

Cette activité enzymatique a été mise en évidence pour la première fois en 1959 dans un extrait de foie de rat (Weiss and Gladstone, 1959). Cette nouvelle activité est alors recherchée et détectée dans divers organismes vivants (Huang et al., 1960; Hurwitz, 1959). La démonstration que chez les bactéries une enzyme unique composée de trois polypeptides (α , β , β') est responsable de la transcription de tous les gènes en ARN date de 1962 (Chamberlin and Berg, 1962). Il a fallu attendre les années 1970 pour que Roeder et Rutter soient capables de séparer trois activités de transcription nucléaires du foie de Rat en fonction de leur comportement chromatographique (Roeder and Rutter, 1970). Dans le même temps, C. Keding et ses collaborateurs distinguaient ces trois enzymes à partir de thymus de veau sur la base de leur sensibilité à l' α -amanitine (Keding et al., 1970). Depuis, ces trois formes d'ARNP (ARNP I, ARNP II et ARNP III) nucléaires ont été bien caractérisées chez la levure *Saccharomyces cerevisiae*, et identifiées dans les cellules animales et végétales (Sentenac, 1985). Ces enzymes se distinguent par leur composition sous-unitaire, leur localisation cellulaire, leur sensibilité à une toxine, l' α -amanitine, ainsi que par les classes de gènes qu'elles transcrivent et les facteurs de transcription qui leur sont associés.

I. Les différentes ARN polymérases

I.1. Les ARN polymérases monomérique

Chez les virus, la transcription est souvent accomplie par une chaîne polypeptidique unique d'un poids moléculaire de l'ordre de 100 kDa illustrée par la polymérase du bactériophage T7. L'étude structurale des polymérases monomériques comme la polymérase du bactériophage T7, la transcriptase reverse du virus VIH-1 (figure 1), l'ADN polymérase de *Th. aquaticus* et l'ADN polymérase I d'*E. coli* a permis de remarquer une forte homologie structurale en dépit d'une faible homologie de séquence et d'une différence de substrat (Hsiou et al., 1996; Kim et al., 1995; Ollis et al., 1985; Sousa et al., 1993).

Les différentes structures obtenues pour l'ARNP du bactériophage T7 ont montré une forte similitude avec le fragment de Klenow de l'ADN polymérase I d'*E. coli*. Trois domaines topologiques ont pu être distingués dans les deux cas : les doigts, la paume et le pouce qui s'organisent autour d'un sillon central en une structure décrite comme « main droite incurvée » (figure 1) (Cheetham and Steitz, 2000b; Sousa, 1996).

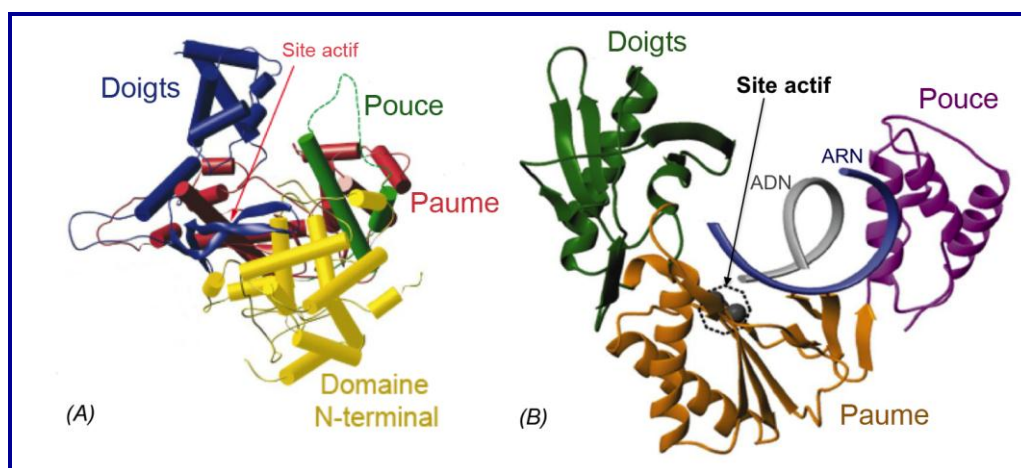


Figure 1 : Modèle de la structure modulaire de la polymérase monomérique.

(A) : Structure de la T7 RNA polymérase. (B) Structure de la transcriptase reverse du virus VIH-1. Les trois domaines (les doigts, la paume et le pouce) sont représentés (Cheetham and Steitz, 2000a; Singh and Modak, 1998).

Les doigts, constitués majoritairement d'hélices α , sont essentiellement impliqués dans la stabilisation de l'interaction avec la matrice. Par ailleurs, un motif présent au niveau du domaine des doigts, appelé boucle de spécificité, est impliqué dans la reconnaissance du promoteur (Raskin et al., 1992; Rong et al., 1998). Le sillon, destiné à recevoir la matrice, contient le cœur catalytique de l'enzyme. Le site est constitué de résidus acides fortement conservés et rapprochés par le repliement 3D de la chaîne polypeptidique. La comparaison des structures d'une ARNP virale, du fragment de Klenow et de la transcriptase reverse fait apparaître une ressemblance sans doute imputable à la similitude de leur activités catalytiques (Moras, 1993). Finalement, le pouce confère la processivité à l'enzyme (Bonner et al., 1994).

I.2. Les ARN polymérases des cellules procaryotes

I.2.1. L'ARN polymérase bactérienne

Tous les ARNs d'une cellule procaryotique sont synthétisés par une même ARNP. L'ARNP d'*E.coli* est une enzyme hétéro-multimérique de 450 kDa. Elle existe sous deux formes : l'enzyme « core » et l'« holo-enzyme ». L'enzyme « core » est constituée de cinq sous-unités : β' et β , deux sous-unité α identiques et une petite sous-unité ω (figure 2). Elle présente une conformation tridimensionnelle dite en « pince de crabe » née de l'assemblage des deux grandes sous unités β' et β qui portent les résidus catalytiques (Borukhov and Nudler, 2008). La sous-unité α est impliquée dans l'assemblage de l'enzyme par sa partie N-terminale. En effet, son homodimérisation est réalisé grâce à cette partie N-terminale qui permet d'assembler les deux sous-unités β' et β (Zhang and Darst, 1998). La partie C-terminale de la sous-unité α est un domaine de fixation à l'ADN, qui joue un rôle dans l'interaction avec certains promoteurs (Gourse et al., 2000). La petite sous-unité ω n'a pas de rôle direct dans la transcription, mais semble fonctionner comme une chaperonne dans le repliement de la sous unité β' et son association avec le complexe intermédiaire de l'assemblage $\beta\alpha^1\alpha^2$ (Minakhin et al., 2001).

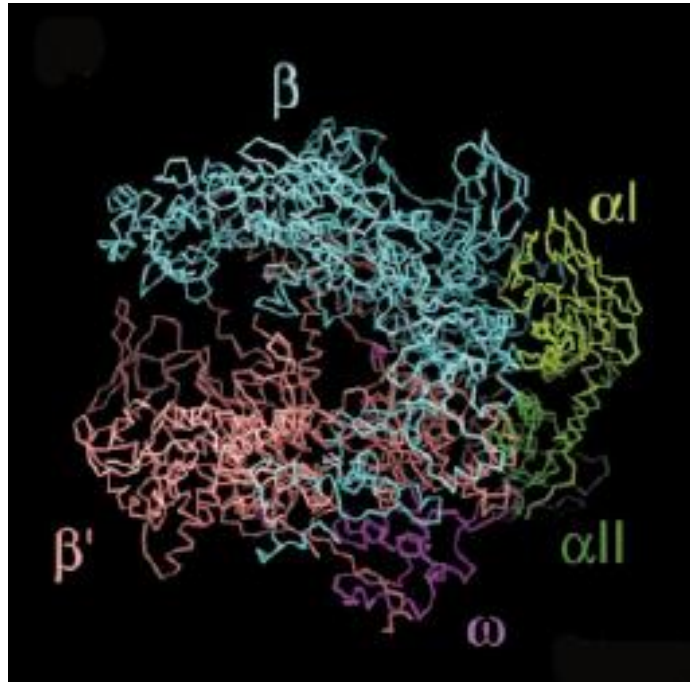
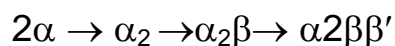


Figure 2 : Structure cristallographique de l'enzyme « core » de l'ARNP de *Thermus aquaticus*. Les sous-unités (β' , β , α^I , α^{II} , ω) sont colorées différemment (Minakhin et al., 2001).

Ce complexe enzymatique n'est pas fonctionnel pour initier la transcription ; en effet, s'il présente une certaine affinité pour les molécules d'ADN, il n'y distingue pas les séquences spécifiques de la plupart des promoteurs. L'association réversible d'une sixième sous-unité σ ou facteur σ de 70 kDa sur l'enzyme « core » constitue une « holo-enzyme » dont l'affinité générale pour l'ADN est réduite au profit d'une forte spécificité pour les promoteurs (Burgess et al., 1969; Gruber and Gross, 2003). Le premier facteur σ bactérien a été découvert en 1969 comme la sous-unité de l'ARNP d'*E.coli* essentielle pour la sélection du promoteur (Burgess et al., 1969). La plupart des bactéries possèdent plusieurs facteurs σ permettent de reconnaître différents promoteurs (Ades, 2008; Campbell et al., 2008a; Chilcott and Hughes, 2000; Klauck et al., 2007). La figure 3 montre une comparaison des structures de 2 types de complexes : l'enzyme « core » et l'« holo-enzyme ». De plus, il est établi que la séquence d'assemblage de la polymérase bactérienne suit le schéma (Ishihama, 1981; Zillig et al., 1976):



Cependant, une différence majeure se situe au sommet de la zone du pouce.

Ce domaine flexible intervient dans la processivité de l'enzyme, car l'enzyme « core » n'exhibe pas de sillon ouvert accessible à l'ADN (Figure 3). Ce domaine se referme sur le sillon et le transforme en une voûte fermée ou une fente à travers laquelle passerait l'ADN ; ce changement de conformation permet à l'enzyme de ceinturer l'ADN et de stabiliser sa forme « core » lors de la phase d'élongation (Polyakov et al., 1995). Comme chez les polymérases monomériques, ce domaine est essentiel pour la processivité de l'enzyme.

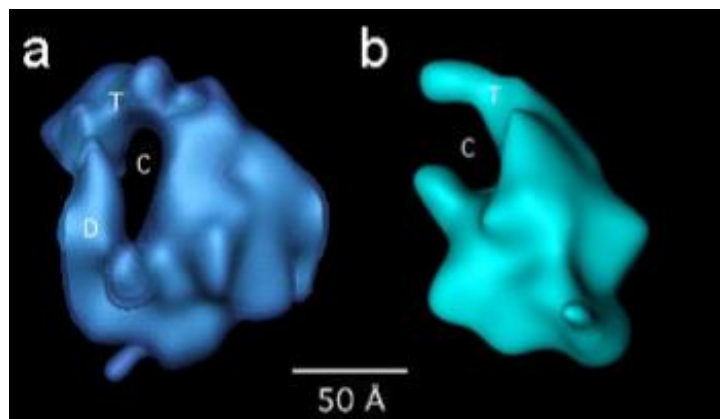


Figure 3 : Comparaison des deux formes de la polymérase bactérienne.

La polymérase d'*E. coli* est en configuration «core-enzyme » en (a) et « holo-enzyme » en (b). T : domaine du pouce; C : sillon et D : densité additionnelle (Polyakov et al., 1995).

I.2.2. L'ARN polymérase des archéobactéries

Les archéobactéries n'utilisent qu'une ARNP qui comporte entre 10 et 14 sous-unités selon l'espèce (Langer et al., 1995). L'ARNP archéobactérienne est un complexe multiprotéique d'environ 370 kDa. Contrairement aux bactéries pour lesquelles l'ARNP s'associe à différents facteurs σ , l'initiation de la transcription chez les eucaryotes et les archéobactéries nécessite plusieurs facteurs généraux de la transcription qui ne sont pas associés de façon permanente à l'ARNP archéobactérienne (Murakami et al., 2002a; Murakami et al., 2002b; Soppa, 1999a; Soppa, 1999b; Soppa, 1999c). L'initiation requiert deux facteurs généraux de la transcription : TBP (*TATA-box Binding Protein*) et TFB (*Transcription Factor B*) (Hausner et al., 1996). TFB se lie à TBP et à l'ADN au niveau d'un élément appelé BRE (*TFB Responsive Element*). Il joue un rôle important dans le recrutement de l'ARNP (Littlefield et al., 1999; Qureshi et al., 1995).

Les sous-unités de l'enzyme archéobactérienne sont désignées par des lettres et peuvent être classées en 3 groupes fonctionnels: (1) celles impliquées dans la catalyse (A'/A'', B'/B'' ou B); (2) celles impliquées dans l'assemblage du complexe (L, N, D et P) et enfin (3) celles impliquées dans des fonctions auxiliaires (F, E, H et K) (figure 4). Les polypeptides de type A et B sont homologues des sous-unités eubactériennes β' et β tandis que l'homodimère α_2 est homologue de L, N, D et P et enfin la sous-unité ω est homologue de K (Werner, 2007).

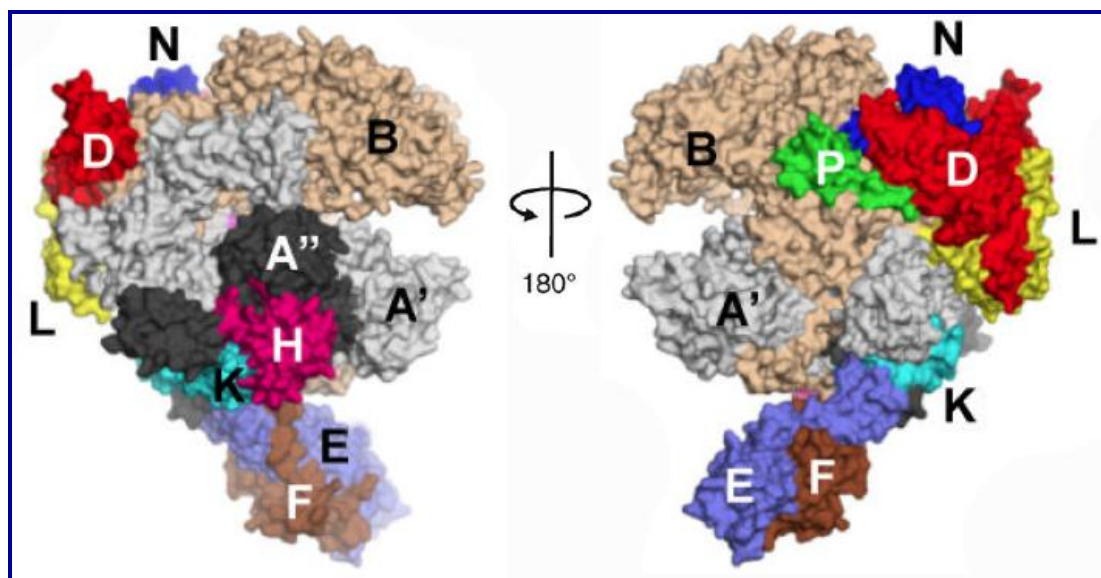


Figure 4 : Structure de surface de l'ARNP de *Sulfolobus solfataricus*.
Chaque sous-unité est notée par une couleur unique et étiquetée (Hirata et al., 2008).

La structure de l'ARNP de l'archéobactérie thermophile *Pyrococcus furiosus* a été déterminée par cryomicroscopie électronique à une résolution de 13 Å (Kusser et al., 2008). Ces résultats ont été confirmés et affinés par la publication de la structure cristallographique de l'ARNP de *Sulfolobus solfataricus* à 3,4 Å et celle de son sous-complexe D/L à une résolution de 1,76 Å (Hirata et al., 2008) (figure 5). Dans cette structure, les trois grandes sous-unités A', A'' et B forment un large corps ellipsoïde en forme de pince de crabe. Elles constituent la partie essentielle de l'enzyme et portent tous les domaines et motifs structuraux nécessaires à la polymérisation de l'ARN. Elles portent notamment le domaine responsable de recrutement de l'enzyme sur l'ADN et constituent la pince flexible qui enserre l'ADN ainsi que le site actif. L'efficacité de l'assemblage des grandes sous-unités requiert les quatre petits

polypeptides D-L-N-P qui forment un sous-complexe (Werner et al., 2000).

Les sous-unités E et F constituent un sous-ensemble stable qui forme une queue qui dépasse du corps de l'enzyme (figure 5). Ce complexe E-F interagit probablement avec le transcrit naissant, car il émerge du canal de sortie de l'ARN qui contient des motifs de fixation de l'ARN, et est impliqué dans la séparation des brins d'ADN au cours de l'initiation de la transcription (Grohmann and Werner, 2011; Meka et al., 2005; Naji et al., 2007).

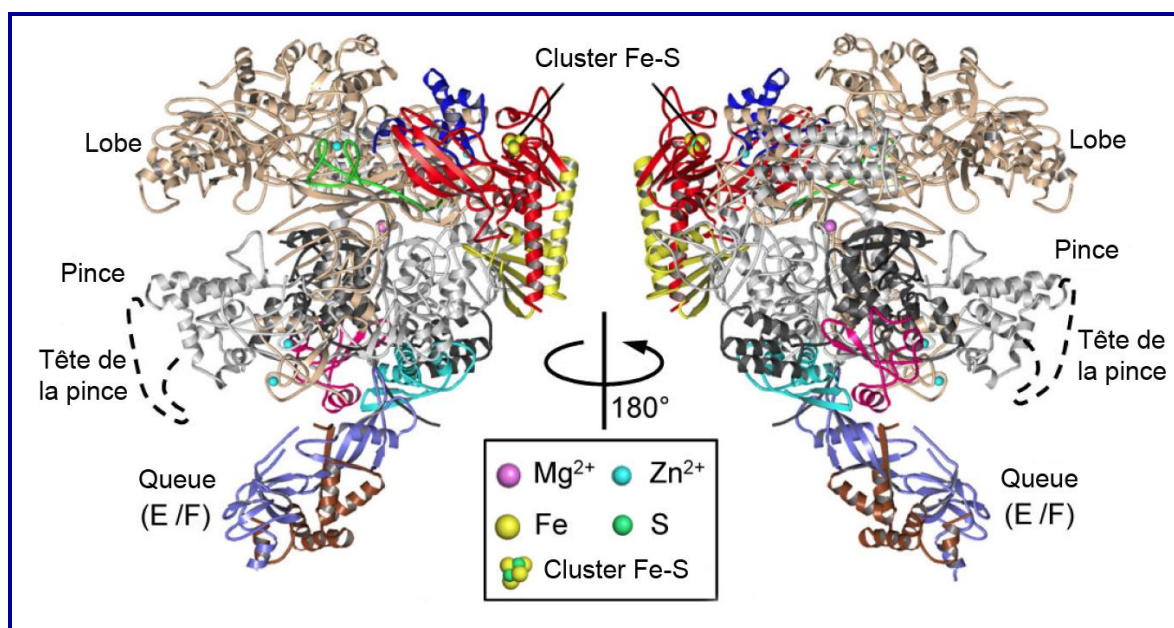


Figure 5 : Structure cristallographique de l'ARNP de *Sulfolobus solfataricus*.

La sous-unité A' est en gris, B en beige, P en vert, D en rouge, N en bleu foncé, L en jaune, K en vert-clair, E' en bleu-clair et F en marron. Le domaine non structuré de la tête de la pince est indiqué en pointillés (Hirata et al., 2008).

La particularité nouvelle de la structure de l'ARNP de *S. solfataricus* est la présence d'un cluster fer-soufre (Fe-S) dans la sous-unité D. Ce cluster n'est pas directement impliqué dans la formation de l'hétérodimère D/L mais l'absence de ce cluster entraîne la formation des agrégats de la sous-unité D qui empêche la formation d'un complexe fonctionnel (D/L). Ce complexe joue un rôle structural important pour l'assemblage de l'enzyme (Hirata et al., 2008). Dans la structure de l'ARNP de *S. solfataricus*, le cluster Fe-S est situé à 45 Å de site catalytique et par conséquent, il est peu probable qu'il soit impliqué dans la synthèse d'ARN (Hirata et al., 2008; Hirata and Murakami, 2009) (Figure 5).

En théorie, le cluster Fe-S pourrait permettre une régulation rédox de la transcription générale de *S. solfataricus*. Cependant, le cluster est étroitement lié à l'enzyme, il n'est pas sensible à l'oxygène. Le motif de fixation du cluster Fe-S n'a pas été retrouvé chez *Pyrococcus furiosus*. Mais de tels motifs étaient prédits dans les séquences des sous-unités D de 16 archéobactéries et AC40 (homologue de la sous-unité D) de l'ARNP I et III de plusieurs eucaryotes (plantes : *Arabidopsis thaliana* et protozoaires : *Tetrahymena thermophila*) (Hirata et al., 2008; Hirata and Murakami, 2009).

I.3. Les ARN polymérases des cellules eucaryotes

Chez les eucaryotes, la réaction de transcription est plus complexe à cause de l'environnement plus structuré du noyau. L'ADN présent dans le noyau d'une cellule eucaryote est condensé au sein de complexes nucléoprotéiques appelés nucléosomes. Ces complexes sont constitués d'histones et d'ADN et ils forment la chromatine. Le niveau de compaction de la chromatine constitue le premier niveau de régulation des gènes. Pour activer la transcription d'un gène dans une cellule, la chromatine située dans la région de contrôle du gène doit être dans une conformation qui la rend accessible aux différentes protéines régulant la transcription. De la même façon, le gène dans sa totalité doit être accessible à la polymérase pour être transcrit et donc localisé dans une zone décondensée de la chromatine. La transcription est assurée par les ARNP nucléaires I, II et III. Chacune de ces enzymes est un hétéromultimère dont la masse est comprise entre 500 et 700 kDa.

On leur attribue la lettre A, B ou C en fonction de leur sensibilité à l' α -amanitine (Les ARNP II et III sont sensibles à l' α -amanitine, ce qui n'est pas le cas de l'ARNP I) (Kedinger et al., 1970) ; ou les numéros I, II, III selon leur ordre d'élution lors d'une séparation par chromatographie sur DEAE-séphadex (Roeder and Rutter, 1969). Nous utilisons classiquement la deuxième nomenclature pour décrire le complexe enzymatique (ARNP I, II, et III), cependant, nous utilisons pour nommer les sous-unités qui composent chaque ARNP la première nomenclature. Ainsi nous parlerons des sous-unités RPA, RPB et RPC pour ARNP I, II et III ou RPAC quand nous parlons de sous-unités communes aux ARNP I et III. Pour définir une sous-unité spécifique des ARNP I et/ou III, on leur affecte un nombre représentant leur masse

molaire en kDa. Les sous-unités de l'ARNP II ont été numérotées de RPB1 à 12, dans l'ordre décroissant de leur mobilité électrophorétique.

L'organisme dont est issue une sous-unité particulière sera précisé devant cette nomenclature : HsRPx pour les sous-unités humaines, ScRPx et PfrPx pour les sous-unités de *S. cerevisiae* et *Plasmodium falciparum* respectivement.

I.3.1. Les trois ARN polymérases (I, II, III) nucléaires

Les études biochimiques, génétiques et structurales des trois ARNP (I,II, III) de levures *S. cerevisiae* en font les ARNP eucaryotiques mieux caractérisées. Les compositions des ARNP I, II et III de la levure *S. cerevisiae* ont été établies. Des homologues humains des sous-unités de levure ont été identifiés, indiquant une même organisation polypeptidique des ARNP humaine et de levure *S. cerevisiae*. Les caractéristiques des trois enzymes seront présentées pour les polymérases de la levure de *S. cerevisiae*.

L'ARN polymérase I est responsable de la transcription de l'ARN ribosomal 35S, précurseur des ARN ribosomiaux (ARNr) 28S, 18S et 5,8S. Elle est nucléolaire et constituée de 14 sous-unités. Les ARN ribosomiaux sont les plus abondants (environ 90% des ARN totaux). La régulation de l'ARNP I est donc nécessairement un aspect très important du contrôle global de la croissance cellulaire (Stefanovsky et al., 2001). Il existe un lien étroit entre la croissance cellulaire et la quantité d'ARNr dans une cellule (Moss, 2004).

L'ARN polymérase II synthétise les précurseurs des ARN messagers (ARNm) codant les protéines et la très grande majorité des petits ARN non codant impliqués dans la régulation de l'expression des gènes (miARN, *micro-ARN*) et des ARN nucléaires (ARNsn) et nucléolaires (ARNsno). Elle est nucléoplasmique et constituée de 12 sous-unités. Depuis peu, on sait aussi qu'elle assure la synthèse d'ARN cryptiques, codés par les régions intergéniques (Wyers et al., 2005). La transcription d'un gène en ARNm est sans doute le mécanisme transcriptionnel le plus complexe et finement régulé. En effet, alors que les ARNr et les ARNt sont transcrits dans toutes les cellules sans distinction, chaque ARNm n'est transcrit que dans un certain nombre de cellules, à un moment donné et pendant un temps donné. C'est ce qui fait la spécificité de chacune de nos cellules.

L'ARN polymérase III est un complexe multiprotéiques de 17 sous-unités, responsable de la transcription des ARN de transfert (ARNt), de l'ARN ribosomal 5S (ARNr 5S), de petits ARN nucléaires (U6 snARN, U6 *Small nuclear ARN*) ainsi que de certains miRNA (Haeusler and Engelke, 2006). L'ensemble des gènes transcrits par l'ARNP III a été identifié par la technique de « ChIP on chip » (Harismendy et al., 2003; White, 2011). Sa localisation nucléoplasmique ou nucléolaire, est encore controversée : elle pourrait être en partie nucléolaire à cause de l'ARN ribosomal 5S qui est à proximité de l'ADNr transcrit par l'ARNP I (Thompson et al., 2003). Cependant ce point n'a pas été clairement établi. Il est possible que la transcription des autres transcrits de l'ARNP III, comme les ARNt, soit également nucléolaire (Thompson et al., 2003).

I.3.2. L'ARN polymérase mitochondriale

Les cellules eucaryotiques contiennent une activité polymérasique distincte de celle des ARNP nucléaires. Cette activité est responsable de la transcription de l'ADN mitochondrial (ADNmt). Peu de données structurales sont connues sur les ARNP mitochondriales (mtARNP), mais des études biochimiques et phylogénétiques ont pu mettre en évidence des homologies avec les ARNP monomériques.

La levure *S. cerevisiae* possède un ADNmt circulaire de 80 kb. Son mtARNP, codée par le gène nucléaire Rpo41, est un polypeptide de 145 kDa. La délétion ou la mutation du gène RPO41 entraîne un phénotype dit de levures « petites » (Greenleaf et al., 1986). Un facteur de spécificité est nécessaire à la reconnaissance du promoteur : Mtf1 (*Mitochondrial Transcriptional Factor 1*). Ce facteur présente des homologies avec les facteurs σ de la bactérie (Mangus et al., 1994). Mtf1 est une protéine de 43 kDa qui interagit directement avec le polypeptide de 145 kDa, pour former une « holoenzyme » fonctionnelle dont l'activité est promoteur-spécifique (Jang and Jaehning, 1991).

Le domaine C-terminal (CTD) (~ 80 kDa) du cœur catalytique de l'enzyme mitochondriale est bien conservé chez les levures, les plantes et les animaux et présente une forte homologie avec la polymérase de type phagique T3 et T7 (Cermakian et al., 1996). Cependant, le domaine N-terminal est moins conservé et il n'est pas présent dans les ARNP des bactériophages T3 et T7. Le domaine N-terminal de Rpo41 de levure interagit avec des protéines (Nam1p et Sls1p) qui sont

impliquées dans la maturation de l'ARNm et la traduction (Rodeheffer and Shadel, 2003). Ce domaine interagit directement avec Mtf1 et cette interaction est modifiée lors du passage de l'initiation à élongation (Paratkar et al., 2011).

A ce jour, aucune donnée structurale n'est disponible sur Rpo41, mais la structure cristallographique du facteur de transcription mitochondrial (Mtf1) de *S. cerevisiae* a été résolue à 2,6 Å (Schubot et al., 2000).

I.3.3. L'ARN polymérase chloroplastique

Dans le chloroplaste des plantes supérieures, la transcription est effectuée par deux types d'ARNP ADN-dépendantes :

(1) la PEP (*Plastid Encoded RNA Polymerase*) est une ARNP de type procaryotique, présente chez la totalité des végétaux. Il s'agit d'une enzyme multimérique, dont le noyau « core » est composé de sous-unités α , β , β' , β'' codées respectivement par les gènes plastidiaux *rpoA*, *rpoB*, *rpoC1* et *rpoC2* (Hu and Bogorad, 1990; Hu et al., 1991). Tout comme chez la bactérie, le noyau de l'enzyme $\alpha_2\beta\beta'\beta''$ doit être associé au préalable avec un facteur de spécificité, capable de reconnaître les promoteurs consensus -10 et -35 : un facteur de type σ ou SLF (*Sigma Like Factors*) (Schweer et al., 2010; Suzuki et al., 2004).

(2) La NEP (*Nuclear Encoded RNA Polymerase*) est une ARNP codée par le noyau, importée dans le plaste. Elle ne répond pas à des critères biochimiques caractéristiques des ARNP bactériennes ou eucaryotique.

Lerbs-Mache a réussi à séparer deux activités transcriptionnelles dans les chloroplastes de très jeunes feuilles d'épinard (Lerbs-Mache, 1993): la première présentant le profil polypeptidique et propriété de la PEP, déjà connue ; et la seconde (NEP) est associée à un polypeptide unique de 110 kDa codé par le gène nucléaire *RpoT*, qui est capable d'initier la transcription à partir d'un promoteur généralement reconnu par l'ARNP du bactériophage T7 (Liere et al., 2004; Liere et al., 2011).

I.3.4. Les ARN polymérases IV et V

Récemment, de nouvelles ARNP nucléaires ont été identifiées chez les plantes (*Arabidopsis thaliana*) et nommées ARNP IV et V. L'ARNP IV synthétise les précurseurs de petits ARN de 24 nucléotides (siRNA). Ces petits ARN vont induire la méthylation de l'ADN et des modifications des histones au niveau de séquences cibles. Ces modifications entraîneront la transformation de la chromatine en hétérochromatine, il s'agit d'une répression au niveau transcriptionnel (Matzke et al., 2006; Ream et al., 2009).

L'ARNP V synthétise les précurseurs des siRNA qui induisent la dégradation spécifique des ARNm des gènes de rétrotransposon comme AtSN1 (Wierzbicki et al., 2008; Wierzbicki et al., 2009).

Elles sont composées de 12 sous-unités. Une étude récente a montré que certaines sous-unités sont partagées entre les trois ARNP II, IV, et V (Tableau 1). Néanmoins, 4 sous-unités de l'ARNP IV (RPD1, 2, 4 et 7) et 6 sous-unités de l'ARNP V (RPE1, 2, 3b, 4, 5 et 7) sont distinctes de leurs paralogues dans l'ARNP II. Les ARNP IV et V sont différentes en 4 sous-unités (1, 3, 5, et 7) (Ream et al., 2008).

Sous-unités	ARNP II	ARNP IV	ARNP V
1	RPB1	RPD1	RPE1
2	RPB2	RPD2	RPE2
3	RPB3	RPD3	RPE3a RPE3b
4	RPB4	RPD4	RPE4
5	RPB5	RPD5	RPE5
6	RPB6	RPD6	RPE6
7	RPB7	RPD7	RPE7
8	RPB8	RPD8	RPE8
9	RPB9	RPD9	RPE9
10	RPB10	RPD10	RPE10
11	RPB11	RPD11	RPE11
12	RPB12	RPD12	RPE12

Tableau 1 : Comparaison des sous-unités de l'ARNP II, IV et V.

Les sous-unités communes de l'ARNP II, IV et V sont colorées en bleu. Les sous-unités spécifiques de l'ARNP IV sont colorées en vert. Les sous-unités spécifiques de l'ARNP V sont colorées en rouge. Les sous-unités communes de l'ARNP IV et V sont colorées en jaune (Ream et al., 2008).

II. Comparaison des ARN polymérase

La structure des ARNP cellulaires est hautement conservée à travers l'évolution, comme le montre le niveau des similitudes entre les séquences primaires des différentes sous-unités des ARNP de la bactérie à l'homme.

II.1. Conservation et spécificité des sous-unités des ARN polymérase

Les 12 polypeptides (RBP1 à RBP12) de l'ARNP II sont paralogues (ou identiques dans le cas de RPB5, RPB6, RPB8, RPB10 et RPB12) à 12 des sous unités des ARNP I et III et ont tous un équivalent chez les archéobactéries (Tableau 2). Le tableau 2 présente leur conservation avec celles des bactéries et les archées (Langer et al., 1995; Werner and Weinzierl, 2002). La comparaison des séquences des protéines eubactériennes, archéobactériennes, et eucaryotiques a permis de définir une série de régions conservées au cours de l'évolution. Cinq des sous-unités sont présentes dans tous les organismes cellulaires, y compris chez les bactéries qui forment l'enzyme « core » (Murakami et al., 2003).

Sous-unités	Procaryotes		Eucaryotes (<i>S. cer</i>)		
	Eubactéries (5)	Archées (12)	ARNP II (12)	ARNP I (14)	ARNP III (17)
Cœur eubactérien	β'	A' / A' '	RPB1	RPA190	RPC160
	β	B' / B' '	RPB2	RPA35	RPC128
	α	D	RPB3	RPC40	RPC40
	α	L	RPB11	RPC19	RPC19
	ω	K	RPB6	RPB6	RPB6
Archées et Eucaryotes		F	RPB4	RPA14	RPC17
		H	RPB5	RPB5	RPB5
		E	RPB7	RPA43	RPC25
		M/TFIIS	RPB9	RPA12	RPC11
		N	RPB10	RPB10	RPB10
spécifiques		P	RPB12	RPB12	RPB12
		G	RPB8	RPB8	RPB8
				RPA34	RPC31
				RPA49	RPC34
					RPC37
				RPC53	
				RPC82	

Tableau 2 : Classification des sous-unités des ARNP.

L'ARNP I et III possèdent en outre un jeu de sous-unités spécifiques. La grande sous-unité de l'ARNP II possède en plus à son extrémité carboxy-terminale une extension particulière, constituée de répétitions de la séquence "Tyr-Ser-Pro-Thr-Ser-Pro-Ser". Le nombre de répétitions varie d'un organisme à l'autre. Ce domaine est essentiel et contient des sites de modifications post-traductionnelles importantes pour la régulation de l'activité de l'enzyme qui sera détaillé plus tard (Corden and Patturajan, 1997; Oelgeschlager, 2002; Proudfoot et al., 2002; Steinmetz, 1997).

L'ARNP I se distingue par ses sous-unités spécifiques RPA49 et RPA34 (Gadal et al., 1997; Huet et al., 1975; Liljelund et al., 1992). L'absence de RPA49 (ou plus précisément son extrémité C-terminale) conduit en effet à un défaut de recrutement du complexe ARNP I/Rn3 (Rn3 : facteur d'initiation de la transcription) (d'où une croissance ralentie à 30 et 25°C), et aussi conduit à une forte sensibilité au mycophénolate (un inhibiteur de l'élongation) (Desmoucelles, 2002; Gadal et al., 2002; Gadal et al., 1997; Liljelund et al., 1992). Le rôle de RPA34 est moins critique, mais les données suggèrent fortement qu'elle stabilise RPA49 sur l'ARNP I et qu'elle assure donc une conformation optimale de RPA49 (Beckouet et al., 2008).

L'ARNP III possède cinq sous-unités spécifiques réparties en deux groupes. Le premier groupe est celui des sous-unités RPC82, RPC34 et RPC31 qui forment un trimère dans l'enzyme. Ces sous-unités sont importantes pour l'initiation de la transcription (Lefèvre et al., 2011; Mosrin et al., 1990; Thuillier et al., 1995). Le deuxième groupe est l'hétérodimère formé par les sous-unités RPC53 et RPC37 (Landrieux et al., 2006). Ces sous-unités interagissent avec RPC11 et seraient impliquées dans le processus de recyclage de l'ARNP III afin de faciliter la réinitiation de la transcription (Landrieux et al., 2006). Ces sous-unités spécifiques sont également importantes dans la reconnaissance correcte du signal de terminaison (Landrieux et al., 2006).

Récemment, deux isoforme de l'ARNP III (α et β) humaine ont été identifiées. Les deux isoformes se distinguent par la présence, soit de la sous-unité HsRPC32 α , soit de HsRPC32 β . (équivalent de RPC31 chez *S. cerevisiae*). L'analyse de plusieurs tissus humains et des lignées cellulaires a montrée que l'ARNP III β est largement exprimée et est indispensable pour la survie de la cellule, alors que l'ARNP III α est seulement détectée dans plusieurs lignées cellulaires de lymphomes

et leucémies et n'est pas indispensable pour la survie de la cellule. Elle est impliquée dans la différenciation cellulaire (Haurie et al., 2010; Teichmann et al., 2010).

II.2. Complémentation entre les sous-unités de l'ARNP II humaine et *S. cerevisiae*

Les structures des ARNP bactériennes et de la levure étant très proches, on peut imaginer que cette conservation s'étende aux ARNP de mammifères et humaines. Les études de remplacement des sous-unités ont montré que seulement 4 des 12 sous-unités ne sont pas fonctionnelles dans la levure : HsRBP2, 3, 5 et 11 (Tableau 3) (Khazak et al., 1998; Khazak et al., 1995; McKune et al., 1995; Schaller et al., 1999; Shpakovski et al., 1995).

Il existe une grande conservation fonctionnelle entre l'homme et la levure concernant l'ARNP II. On peut remarquer que la capacité à compléter ne suit pas le pourcentage d'homologie entre les sous-unités (Tableau 3). Cette grande conservation nous permet de considérer la structure décrite pour l'enzyme de levure *S. cerevisiae*, comme une référence fiable pour nos travaux chez l'homme.

Sous-unités	<i>H. sapiens</i>	
	Complémentation	(%) Homologie
RPB1	+	55
RPB2	-	61
RPB3	-	45
RPB4	+	31
RPB5	-	45
RPB6	+	60
RPB7	+	43
RPB8	+	33
RPB9	+	39
RPB10	+	73
RPB11	-	47
RPB12	+	38

Tableau 3 : Comparaison des résultats de complémentation Homme-levure.

Les sous-unités qui complètent sont indiquées par un (+) et celles ne complétant pas par un (-). Les pourcentages d'homologie des protéines sont indiqués entre l'homme et la levure.

III. La transcription par l'ARN polymérase II

La synthèse des ARN par l'ARNP II est l'une des principales étapes de la régulation de l'expression génétique. La transcription des gènes de classe II, les gènes codant les protéines et les petits ARN non codants est catalysée par l'ARNP II et nécessite toute une série de facteurs protéiques spécialisés appelés facteurs généraux de la transcription (GTF, *General Transcription Factors*). La transcription peut être séparée en trois grandes étapes : l'initiation, l'élongation, et la terminaison. C'est un mécanisme extrêmement complexe et finement régulé qui nécessite, à chacune de ses étapes, la reconnaissance par de multiples facteurs protéiques de nombreuses séquences régulatrices situées dans l'ADN et en particulier au niveau du promoteur des gènes.

III.1. La transcription des gènes de classe II

Les promoteurs des gènes transcrits par l'ARNP II comportent un promoteur minimal et des éléments régulateurs, proximaux ou distaux, spécifique de chaque gène.

III.1.1. Le promoteur minimal

Le promoteur minimal est défini comme la séquence nécessaire et suffisante pour une l'initiation spécifique *in vitro* par l'ARNP II en présence des facteurs généraux de la transcription.

L'élément principal du promoteur est la boîte TATA (Gannon et al., 1979; Tora and Timmers, 2010). La boîte TATA, dont la séquence consensus est TATA(A/T)A(A/T), est localisée à une trentaine de bases environ en amont du site d'initiation de la transcription (TSS, *Transcription Start Site*) (Gross and Oelgeschlager, 2006). Cette séquence est reconnue par la machinerie de base de la transcription, via la protéine TBP (*TATA Binding Protein*) du facteur de transcription TFIID (Buratowski et al., 1988; Nakajima et al., 1988).

Il a longtemps été cru qu'une majorité de gènes eucaryotes contenaient cet élément et qu'il pouvait à lui seul être responsable de la spécificité de l'initiation de la transcription. Or, depuis quelques années il apparaît très clairement que la grande

majorité de ces gènes ne contient pas de « boîte TATA » facilement identifiable (Bajic et al., 2006; Cooper et al., 2006; Gershenzon and Ioshikhes, 2005).

L'élément initiateur (Inr) est situé autour du site d'initiation TSS. Il est constitué d'une séquence riche en pyrimidines moins conservées que la boîte TATA (figure 6) (Tableau 4). Il permet aussi de spécifier l'initiation de la transcription des gènes qui ne contiennent pas de boîte TATA. Un autre type d'élément, de séquence faiblement conservée, le DPE (*Downstream Promoter Element*), est localisé environ 30 pb en aval du site de démarrage de la transcription chez certains promoteurs ne possédant pas de boîte TATA. Le DPE fonctionne de manière synergique avec l'initiateur et est reconnu par le facteur général TFIID (Burke and Kadonaga, 1997).

La séquence BRE (*TFIIB Responsive Element*) riche en guanine et située en amont de la boîte TATA, est reconnue spécifiquement par le facteur général de transcription TFIIB et elle est présente dans environ 12% des promoteurs et permet de faciliter le recrutement et de stabiliser la machinerie transcriptionnelle au niveau du promoteur (Lagrange et al., 1998). Plusieurs équipes ont démontré la présence d'un second BRE localisé cette fois-ci en aval de la « boîte TATA » (figure 6) (Chen and Hahn, 2003; Fairley et al., 2002; Tsai and Sigler, 2000).

Il existe un élément DCE (*Downstream Core Element*), également localisé en aval de site de démarrage de la transcription (Lee et al., 2005a; Lewis et al., 2000). Ce DCE, contrairement aux autres motifs présents dans le coeur des promoteurs, ne se compose pas d'un seul et même élément, mais de trois sous-séquences: SI, SII et SIII avec les motifs CTTC, CTGT et AGC respectivement (Tableau 4). Ces motifs peuvent lier une sous-unité du facteur TFIID, TAF1 (Butler, 2002; Lee et al., 2005a; Lewis et al., 2000; Thomas and Chiang, 2006).

Les éléments MTE (*Motif Ten Element*) sont un autre élément localisé de 18 à 29 pb en aval du site d'initiation de la transcription. Ils contiennent le motif consensus C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C). Il fonctionne en coopération avec l'Inr pour activer la transcription (figure 6) (Lim, 2004; Ohler et al., 2002). La présence combinée de ces différentes séquences au niveau du promoteur détermine la spécificité de l'initiation de la transcription (Gershenzon, 2004).

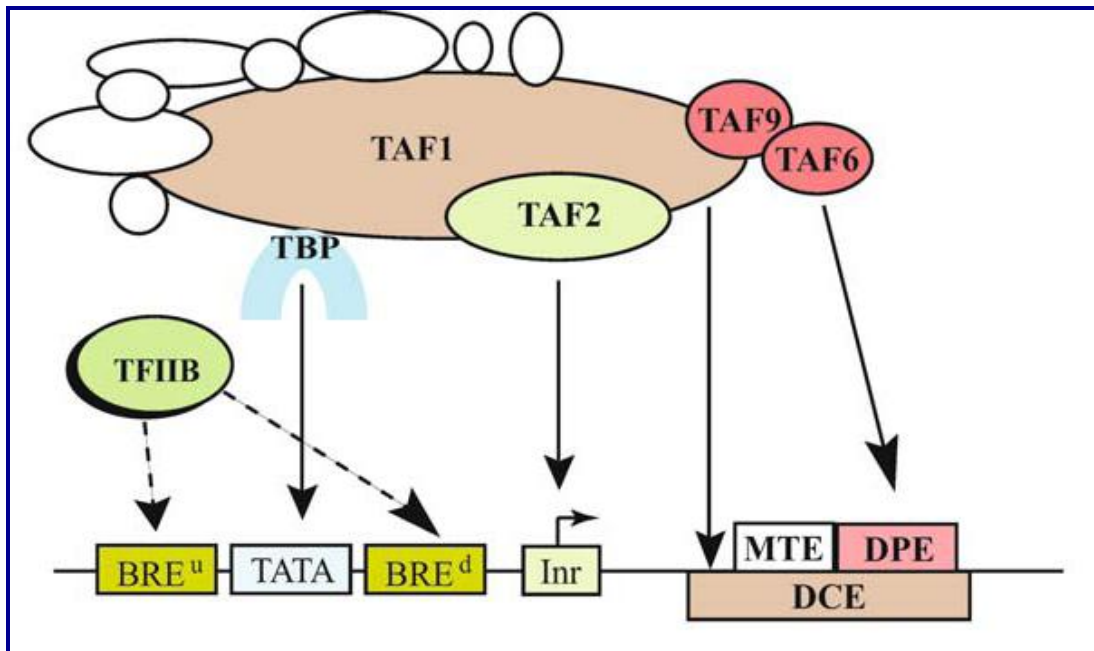


Figure 6 : Reconnaissance des éléments de promoteur par les facteurs TFIID et TFIIB.
(Thomas and Chiang, 2006)

Élément de promoteur	Position	Séquence consensus (5' à 3')	Protéine fixée
BRE ^u	-38 à -32	(G/C)(G/C)(G/A)CGCC	TFIIB
TATA	-31 à -24	TATA(A/T)A(A/T)(A/G)	TBP
BRE ^d	-23 à -17	(G/A)T(T/G)A(T/G)(G/T)(T/G)(T/G)	TFIIB
Inr	-2 à +5	PyPyAN(T/A)PyPy	TAF1/TAF2
MTE	+18 à +39	C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C)	(nd)
DPE	+28 à +34	(A/G)G(A/T)CGTG	TAF6/TAF9
DCE	+6 à +11 +16 à +21 +30 à +34	CTTC CTGT AGC	TAF1

Tableau 4 : Liste des séquences consensus et positions des éléments de promoteur.
(Thomas and Chiang, 2006)

III.1.2. Les éléments de régulation

Le niveau d'expression de chaque gène est différent. Les gènes possèdent ainsi dans leur promoteur des séquences régulatrices qui leur sont propres et qui autorisent le contrôle de leur expression par des protéines régulatrices. Ces séquences permettent d'adapter le niveau d'activité d'un gène aux besoins

physiologiques et à l'état de différenciation des cellules. Les éléments de régulation ont été classés en deux types, selon leur distance par rapport au TSS.

III.1.2.1. Les séquences régulatrices proximales

Les séquences régulatrices proximales sont généralement situées entre 40 et 110 paires de bases en amont du TSS, et sont parfois présentes en plusieurs exemplaires. Les séquences proximales peuvent avoir un effet activateur (UAS, *Upstream Activating Sequence*) ou répresseur (URS, *Upstream Repressing Sequence*) selon les protéines qui les reconnaissent (Figure 7). Les protéines se liant à ces éléments régulateurs proximaux régulent la transcription par contacts directs avec la machinerie de transcription de base (l'ARNP II et GTF).

Parmi les séquences les mieux caractérisées, on trouve les boîtes CAAT et les motifs riches en GC. Les activateurs transcriptionnels de la famille de CTF (*CAAT box Transcription Factor*) reconnaissent la séquence CAAT (Belikov et al., 2004; Gronostajski, 2000), et les motifs riches en GC sont reconnus par les facteurs transcriptionnels de la famille Sp1 via leurs motifs en doigt à zinc C- terminaux (Gill et al., 1994; Lee et al., 2005b; Marco et al., 2003).

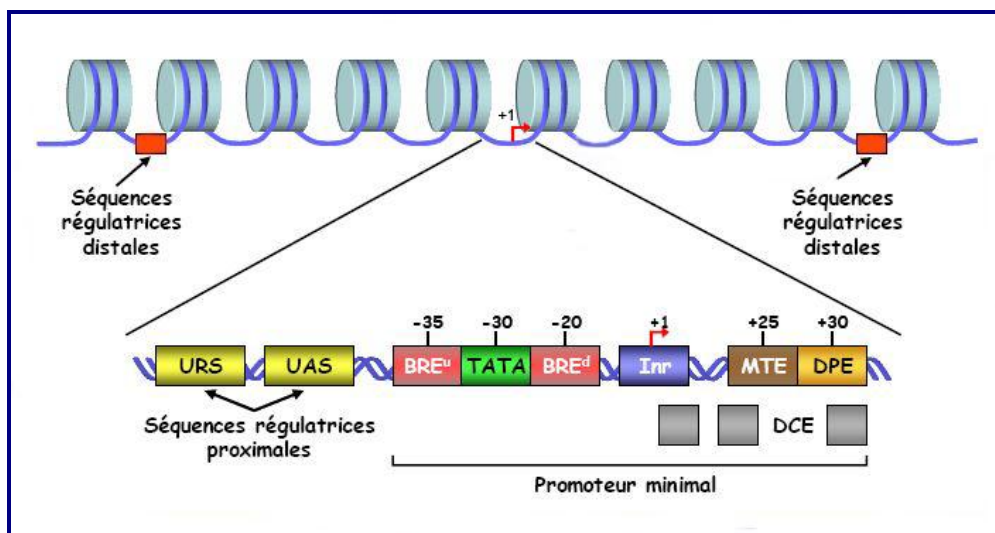


Figure 7 : Représentation schématique des éléments régulateurs de l'expression des gènes. Le promoteur du gène a été détaillé. Il contient le promoteur minimal et des séquences régulatrices proximales répressives URS (*Upstream Repressing Sequence*) et/ou activatrices UAS (*Upstream Activating Sequence*). Tous les éléments représentés au niveau du promoteur minimal n'apparaissent pas dans tous les promoteurs. Les séquences régulatrices distales sont également représentées au niveau de la chromatine.

III.1.2.2. Les séquences régulatrices distales

Les séquences régulatrices distales peuvent être situées jusqu'à plusieurs milliers de paires de bases en amont ou en aval du TSS (Ptashne, 1988). Elles sont appelées « enhancer » lorsqu'elles sont des sites de liaison d'activateurs transcriptionnels ou « silencer » lorsqu'elles sont des sites de liaison de répresseurs (Levine and Tjian, 2003). La proximité de multiples « enhancers » permet le recrutement d'un complexe transcriptionnellement actif, nommé « enhanceosome », jouant un rôle important dans la régulation spatiale et temporelle de l'expression des gènes (Gill et al., 1994; Hertel et al., 1997; Levine and Davidson, 2005; Levine and Tjian, 2003).

Les études des neurones corticaux chez la souris ont montré que l'ARNP II se lie à des milliers d'enhancers et transcrit une nouvelle classe d'ARN régulateur non codant (eARN, *enhancer ARN*). Le niveau d'expression de eARN est positivement corrélé avec le niveau de l'ARNm synthétisé à partir du gène à proximité. Ce qui suggère que les eARN synthétisés à partir d'enhancer sont activement engagés pour promouvoir de synthèse de l'ARNm (figure 8) (Kim et al., 2010; Ren, 2010).

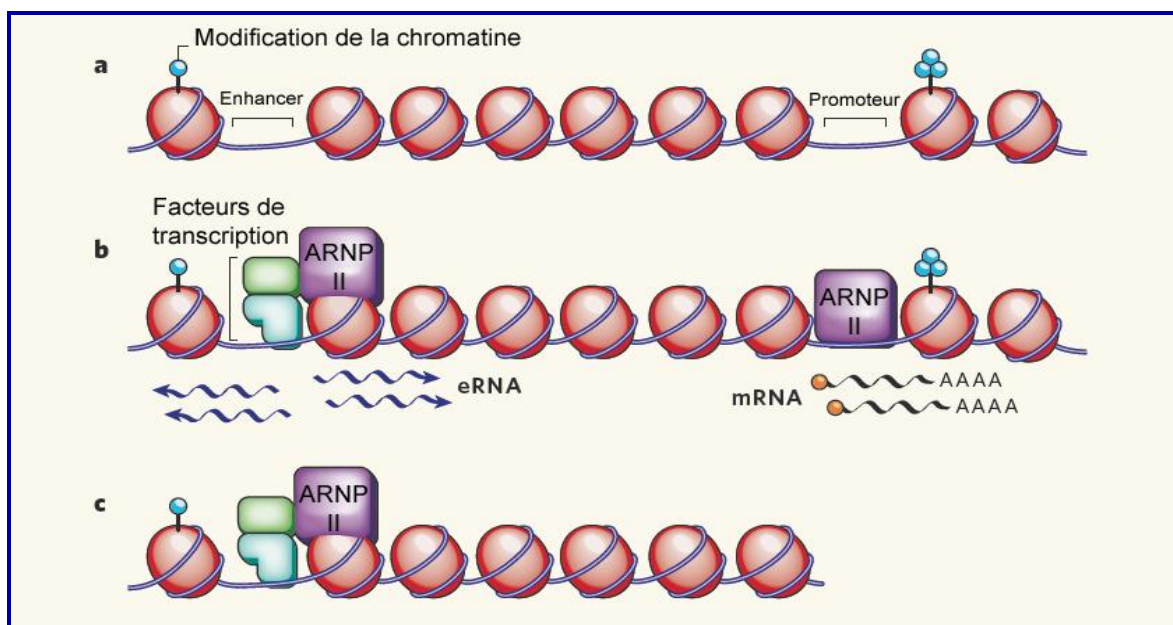


Figure 8 : Production de eARN à partir d'enhancer.

a, avant l'activation du gène, l'enhancers des neurones et les promoteurs sont dans une configuration ouverte, mais ne produisent pas de transcrits ou seulement de faibles niveaux. **b**, après une activation par la dépolarisation de la membrane suite à un traitement de KCl, les facteurs de transcription et l'ARNP II se lient à des enhancers et produisent eARN. Simultanément, ARNP II et machinerie de transcription se lient aussi aux promoteurs pour initier la transcription de l'ARNm. **c**, si le promoteur

est tronqué, l'enhancer reste lié aux facteurs de transcription et de l'ARNP II, mais sans produire de eARN (Ren, 2010).

Les enhancers se distinguent par des modifications post-traductionnelles des histones spécifiques. La définition d'une signature sur la chromatine au niveau d'enhancers telle que : la mono- et di-méthylation de la lysine 4 des histones 3, l'acétylation de la lysine 27 des histones 3 et la présence de la protéine p300 (qui a une activité acétyltransférase) a permis de prédire des enhancers putatifs dans le génome (Ong and Corces, 2011).

III.1.3. Les facteurs généraux de la transcription

L'ARNP II n'est pas capable de reconnaître, seule, le promoteur pour démarrer la synthèse d'ARN. Plusieurs protéines ou complexes protéiques sont indispensables à la fixation spécifique de l'ARNP II sur le promoteur et sont désignés comme les facteurs généraux de la transcription (GTF), tels TFIIA, TFIIB, TFIID, TFIIE, TFIIIF et TFIIF (TFII, *Transcription Factor of RNA polymerase II*) (Orphanides et al., 1996).

Ils ont été purifiés à partir de plusieurs organismes (Homme, drosophile et levure) (Conaway and Conaway, 1993; Zawel and Reinberg, 1995). Le tableau 5 présente la composition et les propriétés de chacun de ces facteurs.

Facteur général de transcription	Sous-unités	Propriétés
TFIIA	α 35 kDa	- Dissocie les dimères de TBP - Requis pour l'activation de la transcription - Stabilise le complexe TBP-boîte TATA
	β 19 kDa	
	γ 12 kDa	
TFIIB	35 kDa	- Lie les BREu et BREd - Stabilise le complexe TBP-boîte TATA - Recrutement et positionnement du complexe pol II / TFIIIF
TFIID	TBP 38 kDa	- Lie la TATA Box
	TAF1 250 kDa	- Lie l'Inr et le DCE, activité histone acétyl transférase, histone ubiquitine ligase et kinase
	TAF2 150 kDa	- Lie l'Inr en coopération avec TAF1
	TAF3 140 kDa	- Impliqué dans la fixation du promoteur. - Reconnaît spécifiquement la lysine 4 de l'histone H3 triméthylée.
	TAF4 130/135 kDa	- Interaction avec le domaine d'activation riche en glutamine de facteur de transcription Sp1 - Similitude de séquence avec l'histone H2A - Interaction avec TAF12
	TAF4b 105 kDa	- Similitude de séquence avec l'histone H2A

TFIID			- Interaction avec TAF12 - Interaction avec le facteur de transcription AP-1
	TAF5	100 kDa	- Similitude de séquence avec l'histone H4 - Joue un rôle essentiel dans l'assemblage du complexe TFIID
	TAF6	80 kDa	- Lie le DPE avec TAF9, similitude de séquence avec l'histone H4
	TAF7	55 kDa	- Inhibe l'activité histone acétyl transférase de TAF1 - Interaction avec TFIIH et P-TEFb (<i>Positive Transcription Elongation Factor b</i>)
	TAF8	43 kDa	- Interaction avec TBP et TAF10 - Similitude de séquence avec l'histone H3
	TAF9	32/31 kDa	- Lie le DPE avec TAF6, similitude de séquence avec l'histone H3
	TAF10	30 kDa	- Interaction avec TAF8, possède une structure de type histone
	TAF11	28 kDa	- Possède une structure de type histone - Interaction avec TAF13 et TFIIA
	TAF12	20 kDa	- Similitude de séquence avec l'histone H2B
	TAF13	18 kDa	- Possède une structure de type histone H4 - Interaction avec TAF11 et TBP
	TAF14	27 kDa	- Interaction avec TFIIIF, le médiateur, histone acétyl transférase NuA3 et le complexe de remodelage de la chromatine SWI/SNF chez <i>S. cerevisiae</i>
	TAF15	68 kDa	- Impliqué dans l'épissage de l'ARNm et son transport vers le cytoplasme, domaine de liaison à l'ADN, Interaction avec l'ARNP II
TFIIE	α	56 kDa	- Recrutement de TFIIH avec TFIIE
	β	34 kDa	- Lie l'ADN
TFIIF	RAP74	58 kDa	- Recrutent l'ARNP II
	RAP30	26 kDa	- Stimulent l'élongation de la transcription <i>in vitro</i> - Impliqué dans le recrutement de TFIIIB
TFIIH	XPB	89 kDa	- Activité ATPase pour l'initiation de la transcription et le dégagement du promoteur - Ouverture de la double hélice (activité hélicase) - Réparation de l'ADN - Phosphorylation de la CTD de la polymérase II par cdk7 - Activité ubiquitine ligase E3 - Activité endonucléase
	XPD	80 kDa	
	XPG	133 kDa	
	p62	62 kDa	
	p52	54 kDa	
	p44	44 kDa	
	p34	34 kDa	
	cdk7	40 kDa	
	Cycline H	38 kDa	
	MAT1	32 kDa	
TTDA	8kDa		

Tableau 5 : Composition et propriétés des facteurs généraux de transcription.

(Bhattacharya et al., 2007; Egly and Coin, 2011; Gazit et al., 2009; Gegonne et al., 2008; Kalogeropoulou et al., 2010; Law et al., 2006; Robinson et al., 2005; Schatz et al., 2010; Schulze et al., 2010; Thomas and Chiang, 2006; Ujvari et al., 2011; van Ingen et al., 2008; Woychik and Hampsey, 2002).

III.1.4. Le complexe médiateur

L'ARNP II et les GTF sont suffisants pour la transcription de base dans un système reconstitué *in vitro*, cependant ce système est incapable de répondre aux signaux de régulation. Ceci a servi de base à la recherche de facteurs qui stimulent la transcription en présence d'activateurs à partir d'extraits protéiques de *S. cerevisiae* (Kornberg, 2005).

Un groupe de gènes identifiés comme des suppresseurs extragéniques d'une troncation du CTD de l'ARNP II appelé SRB (*Suppressor of RNA polymerase B*), a pu être isolé : MED20 (SRB2), MED17 (SRB4), MED18 (SRB5), MED22 (SRB6), MED21 (SRB7), MED12 (SRB8), MED13 (SRB9), CDK8 (SRB10) et CYCC (SRB11). Les protéines codées par ces gènes ont ensuite été purifiées au sein d'un complexe qui lie le CTD et qui favorise l'initiation de la transcription *in vitro* (Hengartner et al., 1995; Nonet and Young, 1989; Thompson et al., 1993).

Indépendamment, le complexe médiateur a été purifié et caractérisé *in vitro* par l'immunoprécipitation de « l'holoenzyme » (Kim et al., 1994). Les grandes activités du médiateur ont pu être déterminées : il permet l'activation de la transcription par l'ARNP II en réponse aux activateurs, il stimule la transcription basale dont le niveau augmente en sa présence (Blazek et al., 2005; Taatjes, 2010).

La morphologie de ces complexes varie toutefois en fonction de leurs partenaires d'interactions, activateurs ou ARNP II (Davis et al., 2002; Taatjes, 2002).

Les études biochimiques (la purification de médiateur et l'interaction protéine/protéine) et structurales par microscopie électronique ont montré que le médiateur est présent dans les cellules au moins sous 3 formes : « core » qui est composé de 26 sous-unités chez l'homme disposées en trois modules (la tête, le corps et la queue), « holoenzyme » est une conformation plus ouverte de médiateur qui encercle et se lie étroitement à l'ARNP II et la troisième forme, dans laquelle une kinase (Cdk8), lie au « core » d'une manière qui empêche le médiateur de se lier à l'ARNP II (figure 9) (Conaway and Conaway, 2011; Taatjes, 2010).

Une étude récente a montré que la sous-unité MED26 peut fonctionner comme un switch moléculaire qui interagit d'abord avec TFIID dans le complexe d'initiation de la transcription. Ensuite, elle change le TFIID contre les facteurs d'élongation pour faciliter la transition de l'ARNP II vers l'étape de l'élongation (Takahashi et al., 2011).

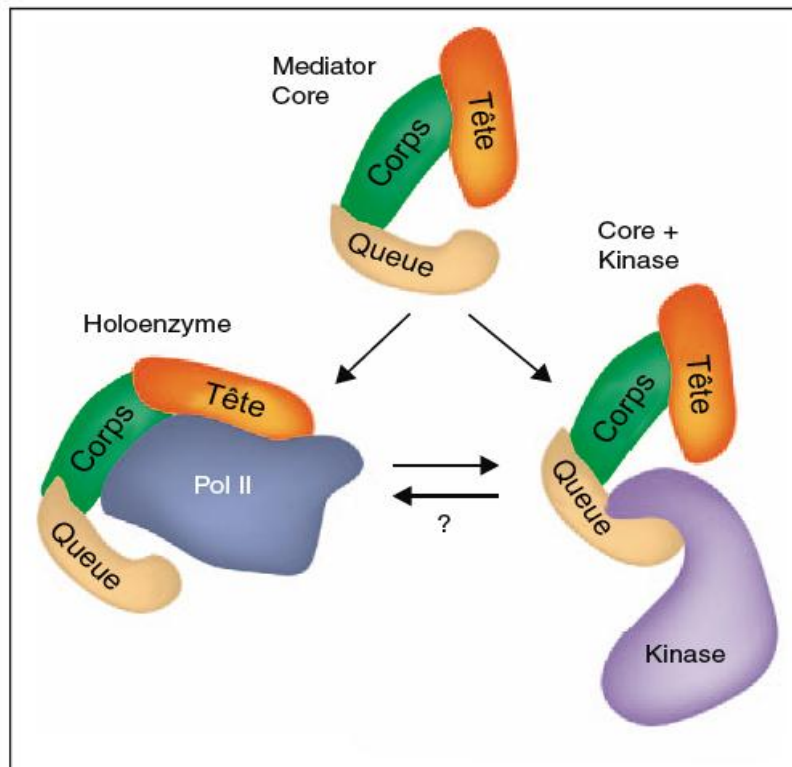


Figure 9 : Différentes formes de médiateur.

Les trois formes de médiateur : « core », « holoenzyme » et « core+kinase » sont présentées. Le médiateur avec 3 domaines : la tête, le corps et la queue. L' ARNP II est en bleu et la kinase est en violette (Conaway and Conaway, 2011).

La figure 10 indique que le site de liaison à l'ADN de l'ARNP II reste tout à fait accessible dans le complexe « holoenzyme » (Davis et al., 2002).

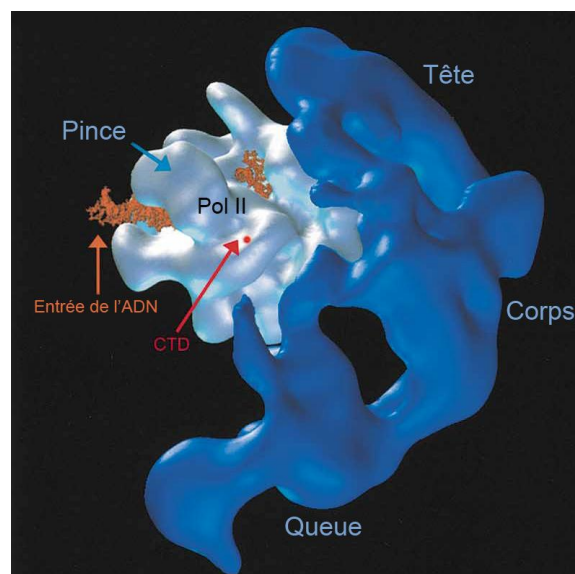


Figure 10 : Structure du complexe Pol II/médiateur.

Le médiateur en bleu avec 3 domaines : la tête, le corps et la queue. L' ARNP II est blanche avec une molécule d'ADN en amont et en aval. Le point rouge représente le début probable du CTD de la grande sous-unité de la Pol II (Davis et al., 2002).

III.2. Initiation de la transcription

La transcription d'un gène en ARNm par l'ARNP II nécessite tout d'abord la formation du complexe de pré-initiation (PIC) au niveau du promoteur. Ce complexe est constitué de l'ARNP II, de GTF et d'un complexe médiateur. Des études *in vitro* ont proposé deux modèles d'assemblage du PIC : un modèle séquentiel et un modèle de complexe pré-assemblé dit « holoenzyme » (figure 11).

Le modèle séquentiel se base sur l'observation que l'addition des GTF doit se faire dans un ordre spécifique pour initier la transcription *in vitro* à partir d'une matrice d'ADN nu. La première étape de la formation du PIC correspond à la reconnaissance du promoteur par TFIID au niveau de la « boîte TATA » (Davison et al., 1983; Fire et al., 1984). Ce facteur est un composé de la TBP (Nakajima et al., 1988) et des TAFs (*TBP Associated Factors*) (Tableau 4) (Pugh and Tjian, 1991; Tanese et al., 1991). En l'absence de boîte TATA, d'autres sous-unités de TFIID permettent la reconnaissance du promoteur, notamment TAF2 qui fixe l'élément Inr ou encore TAF6 et TAF9 qui fixent l'élément DPE (figure 6) et (Tableau 4) (Chalkley and Verrijzer, 1999).

La fixation de TFIID sur le promoteur engendre une courbure importante de la molécule d'ADN (Nikolov et al., 1995). Cette courbure permettra la fixation ultérieure des autres GTF dont TFIIA et TFIIB (Fan et al., 2004; Kim et al., 1993). TFIIA et TFIIB viennent stabiliser TFIID et recruter un complexe formé par TFIIIF et l'ARNP II. Enfin, TFIIE puis TFIIH viennent s'associer à la structure préexistante, et l'activité hélicase de TFIIH catalyse l'ouverture de la bulle de transcription et conduisant à la synthèse du premier dinucléotide. Par la suite, l'allongement de la bulle de transcription permet la stabilisation du PIC et l'allongement de l'ARN naissant. Enfin, le déplacement de la bulle de transcription plus en aval entraîne l'échappée du promoteur de la Pol II ce qui correspond au passage de l'étape d'initiation à celle d'élongation (Holstege et al., 1997).

Le modèle d'assemblage de « l'holoenzyme » a été proposé suite à la purification *in vivo* d'un énorme complexe protéique, non lié au promoteur. Ce complexe contenant l'ARNP II, TFIIB, TFIIIF, TFIIH ainsi que certaines protéines SRB et protéines impliquées dans le remodelage de la chromatine (*SWI/SNF*, *mating type SWItching/Sucrose Non Fermenting*) (Kim et al., 1994; Koleske and Young, 1994; Wu and Chiang, 1998; Wu et al., 1999). Malgré la grande hétérogénéité des

préparations d'holoenzymes décrites dans plusieurs études, une propriété constante était l'absence de TFIID. Ceci suggère le recrutement de TFIID au niveau du promoteur, puis celui de l'holoenzyme (figure 11).

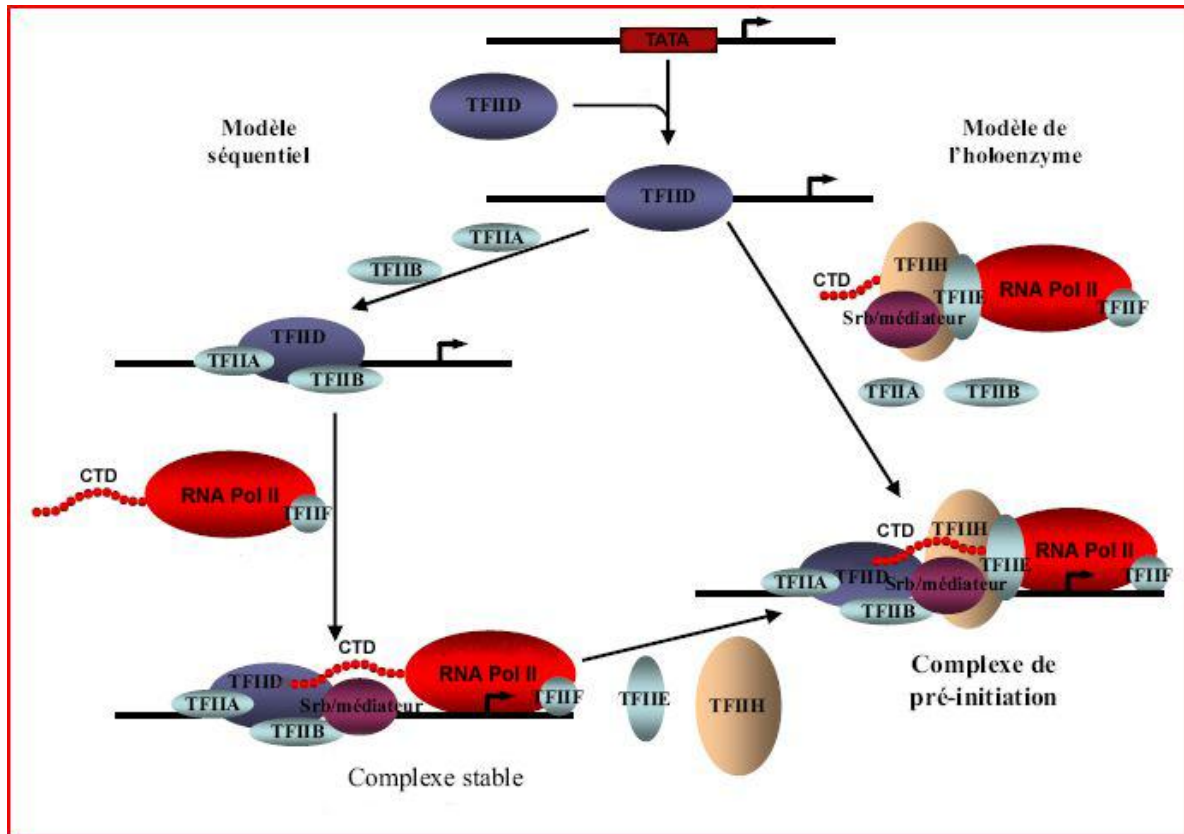


Figure 11 : Les modèles d'assemblage du PIC.

La formation du PIC peut se faire par recrutement par étape des GTF (modèle séquentiel) ou par recrutement d'un complexe contenant l'ARNP II pré assemblé (modèle de l'holoenzyme). Le facteur TFIID est au préalable lié à la boîte TATA dans les deux modèles.

III.3. Elongation de la transcription

Lors de la transition entre l'étape d'initiation et d'élongation de la transcription, les deux brins de la molécule d'ADN sont séparés par la sous-unité XPB/RAD25 (chez l'homme/chez la levure) de TFIIH de façon ATP dépendante (Guzder et al., 1994; Holstege et al., 1996; Yan and Gralla, 1997). La phosphorylation du domaine C-terminal (CTD) de la grande sous-unité de l'ARNP II semble être l'évènement clé de la transition entre l'initiation et l'élongation de la transcription. Cette phosphorylation pourrait déstabiliser les interactions entre l'ARNP II et les facteurs d'initiation. Ensuite, l'ARNP II se dissocie des facteurs généraux au niveau du promoteur.

L'ARNP II peut être en situation de pause à proximité du promoteur (Uptain et al., 1997). Cette pause est un phénomène très important, car il régule la transcription de nombreux gènes chez les eucaryotes (Levine, 2011). Il a été particulièrement bien étudié chez *Drosophila melanogaster* dans le cas des gènes dont l'expression est induite par un choc thermique (Gilmour, 2009; Saunders et al., 2006). L'état de pause de l'ARNP II peut être causé par des répresseurs de l'élongation, DSIF (*DRB Sensitivity Inducing Factor*) et NELF (*Negative Elongation Factor*) (Saunders et al., 2006; Sims, 2004) alors que des régulateurs positifs comme P-TEFb peuvent stimuler la reprise de l'élongation en induisant la dissociation de DSIF et NELF de l'ARNP II (figure 12) (Barboric et al., 2001; Core and Lis, 2008; Rahl et al., 2010).

La vitesse de progression de l'ARNP II peut être augmentée en réduisant le nombre des arrêts ou en limitant leur durée. Les facteurs TFIIF, ELL (*Elevenineteen Lysine-rich in Leukemia*) et les elongines stimulent l'élongation de la transcription de l'ARNP II en réduisant ces pauses (Lei et al., 1999; Lin et al., 2010; Shilatifard et al., 1996).

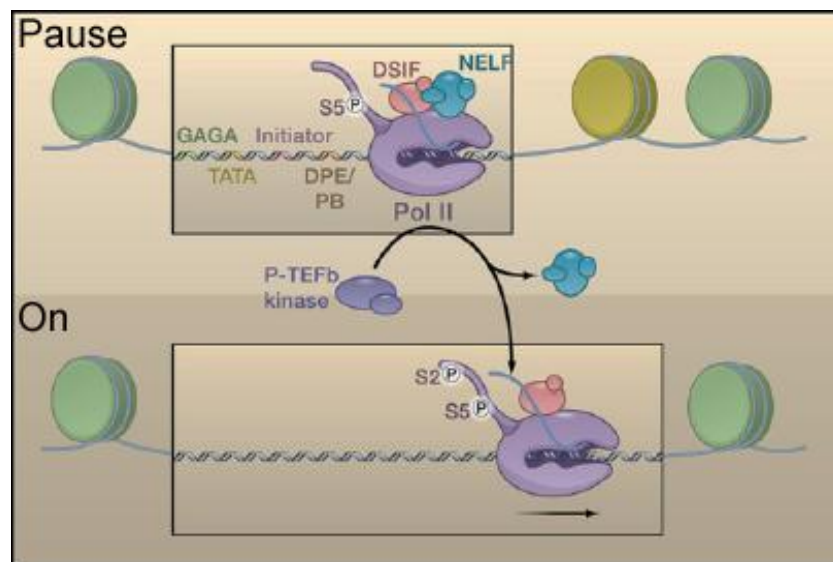


Figure 12 : Pause de l'ARNP II et sa libération de promoteur proximal.

Le panneau supérieur montre un promoteur avec l'ARNP II en situation de pause. La région promotrice contenant des éléments, y compris GAGA, TATA, Initiateur, et les motifs DPE/PB est présentée. L'ARNP II est généralement en pause juste en aval de la région DPE. DSIF et NELF sont liés à l'ARNP II et aussi au transcrit naissant. Le recrutement de P-TEFb provoque la libération de NELF, la phosphorylation de domaine carboxy-terminal de la grande sous-unité et la processon de l'ARNP II (Levine, 2011).

III.4. Terminaison de la transcription

La terminaison de la transcription par l'ARNP II est une étape essentielle permettant d'éviter une interférence dans la transcription des gènes situés en aval du transcrit. La terminaison dépend du signal de polyadénylation et au cours de cette étape, l'ARNP II pourrait subir des changements de conformation qui réduirait sa processivité (Birse, 1998; Osheim et al., 1999).

Deux modèles ont été proposés en ce qui concerne le déroulement des évènements. Le premier modèle « anti-terminateur » propose que l'émergence de la séquence de polyadénylation sur l'ARN induise un échange des facteurs associés à l'ARNP II. Ainsi, la fixation de facteurs reconnaissant la séquence de polyadénylation sur l'ARN entraînerait le départ de facteurs positifs de l'élongation et le recrutement de facteurs négatifs de l'élongation. Par conséquent l'ARNP II devient moins processive, et finirait par s'arrêter (Buratowski, 2005).

Un second modèle appelé « torpedo » propose que le clivage de l'ARN par la machinerie de poly-adénylation génère une nouvelle extrémité 5' qui serait le substrat d'une nucléase dont l'activité entraîne la dissociation de l'ARNP II (Buratowski, 2005).

Il est intéressant de noter que ce dernier modèle ressemble beaucoup au mécanisme de terminaison bactérien dépendant du facteur Rho (Richardson, 2003). Il semble qu'il existe plusieurs mécanismes de terminaison de la transcription en fonction du type cellulaire et même du gène concerné. Dans tous les cas, la terminaison de la transcription reste étroitement liée à la polyadénylation du pré-ARNm (Buratowski, 2005). Par ailleurs, un mode de terminaison alternatif a récemment été mis en évidence dans le cas de la transcription des sn/sno ARN. Celui-ci fait intervenir le complexe Ndr1p, qui reconnaît des sites spécifiques à l'extrémité 3' de l'ARN (Lykke-Andersen and Jensen, 2007).

IV. ARN polymérase II

IV.1. Organisation de l'ARN polymérase II

L'ARNP II est une protéine multimérique de 0,6 MDa composée de 12 sous-unités, nommées RPB1 à RPB12. L'ARNP II de *S. cerevisiae* peut être dissociée en un coeur catalytique de dix sous-unités et un hétérodimère composé des sous-unités Rpb4 et Rpb7 (Edwards et al., 1991; Orlicky, 2001). Cinq sous-unités (RPB5, RPB6, RPB8, RPB10, RPB12) sont communes aux trois ARNP, et sept sous-unités (RPB1, RPB2, RPB3, RPB4, RPB7, RPB9, RPB11) ont des paralogues chez les ARNP I et III.

Les gènes codant chacune des douze sous-unités de l'ARNP II sont conservés au cours de l'évolution. Dix des douze gènes codant les sous-unités de l'ARNP II sont essentiels. Chez la levure seuls *RPB4* et *RPB9* peuvent être délétés, bien que ces délétions provoquent des phénotypes de croissance cellulaire ralentie (Hampsey, 1998b). De nombreuses études génétiques et biochimiques ont permis de mieux définir les fonctions des différentes sous-unités de l'ARNP.

IV.2. Les sous-unités de l'ARN polymérase II

IV.2.1. Les sous-unités RPB1 (220 kDa) et RPB2 (140 kDa)

Les deux grandes sous-unités présentent des homologies avec les sous-unités β' et β des eubactéries respectivement. Elles sont responsables de l'activité catalytique de l'ARNP et essentielles pour la formation de la liaison phosphodiester (Hampsey, 1998b; Lee and Young, 2000).

De plus, RPB1 et β' sont impliquées dans la fixation de l'ADN alors que RPB2 et β fixent les nucléotides (Hampsey, 1998a). En effet, durant la transcription par l'ARNP II, les nucléotides entrent dans le site catalytique en association avec un ion Mg^{2+} . La structure cristallographique de la ARNP II de la levure a révélé la position de deux ions Mg^{2+} dans le site catalytique de l'enzyme (Cramer et al., 2001; Gnatt et al., 2001; Westover et al., 2004).

Le premier Mg^{2+} (A) est coordonné par trois résidus aspartate de RPB1 (D481, D483 et D485) contenus dans le motif strictement conservé NADFDGD (Figure13). Le deuxième Mg^{2+} (B) a récemment été localisé près de résidus D481 et D483 de

RPB1 ainsi que le résidu hautement conservé D837 de RPB2 (Westover et al., 2004).

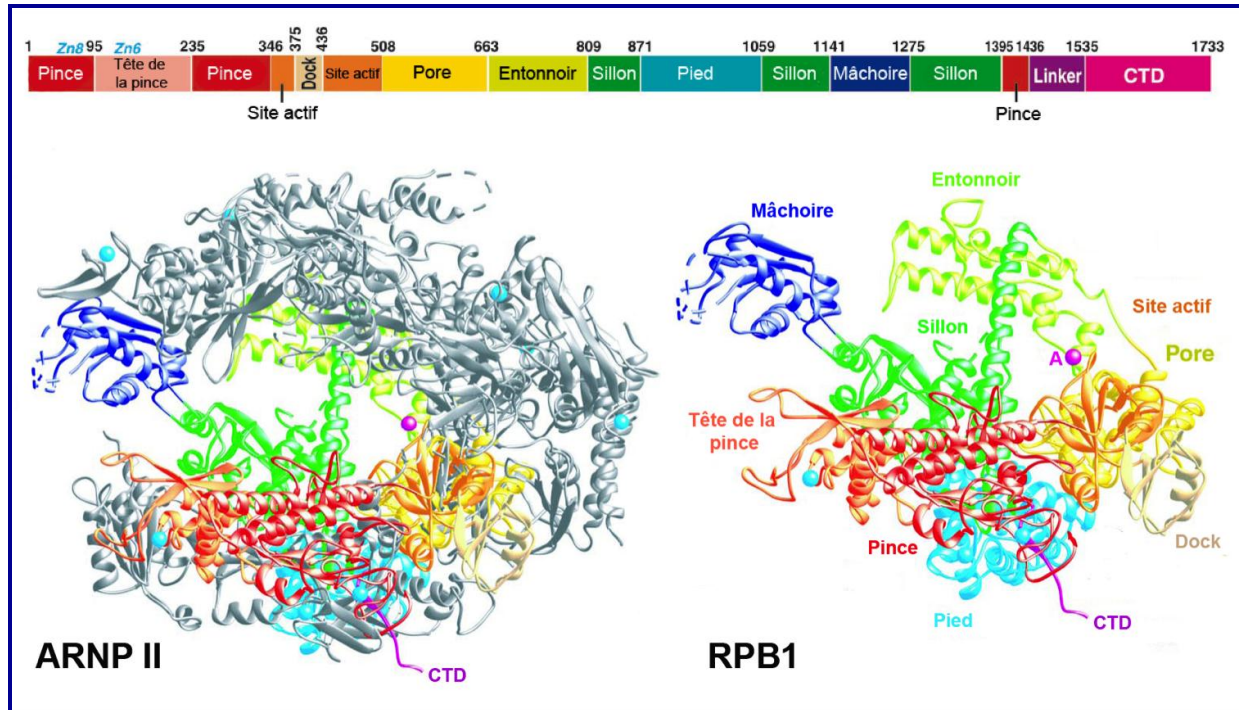


Figure 13 : Structure cristallographique de la sous-unité RPB1.

La position de la grande sous-unité RPB1 est présentée au sein du complexe de l'ARNP II. La sous-unité RPB1 isolée est colorée en fonction de ses différents domaines. Les ions Mg^{2+} (A) et Zn^{2+} sont représentés dans la structure par une sphère violet et bleu respectivement (Cramer et al., 2001).

Par ailleurs, des études antérieures ont suggéré que le résidu E836, très conservé, est également impliqué dans la coordination de l'ion Mg^{2+} (B) dans le site actif. L'activité de transcription du mutant E791 homologue de E836 dans la protéine humaine de RPB2 est altérée, *in vitro* et *in vivo* à faible concentration de nucléotides. De plus, la polymérisation des nucléotides et les activités de clivage du transcrit sont diminuées aux faibles concentrations en magnésium. En définitive, la proximité du complexe nucléotide/ Mg^{2+} avec le résidu E791 de RPB2 permettrait un contact direct de ce résidu avec les sites d'entrée et d'ajout des nucléotides lors de la transcription (Langelier, 2005).

Des expériences de délétions des résidus ($\Delta 458-459$) de la boucle 1 de la sous-unité RPB2 montrent que cette boucle est essentielle pour l'initiation et l'élongation de la transcription, mais n'est pas impliquée dans la formation du complexe de préinitiation (figure 14). Des mutations, dans une autre boucle de RPB2 localisée près de l'hybride ADN-ARN, démontrent que celle-ci est requise pour une formation efficace du PIC (Jeronimo et al., 2004). Il a été découvert une interaction entre RPB2 et TFIIB, qui est importante pour le recrutement de l'ARNP II au promoteur (Chen and Hampsey, 2004).

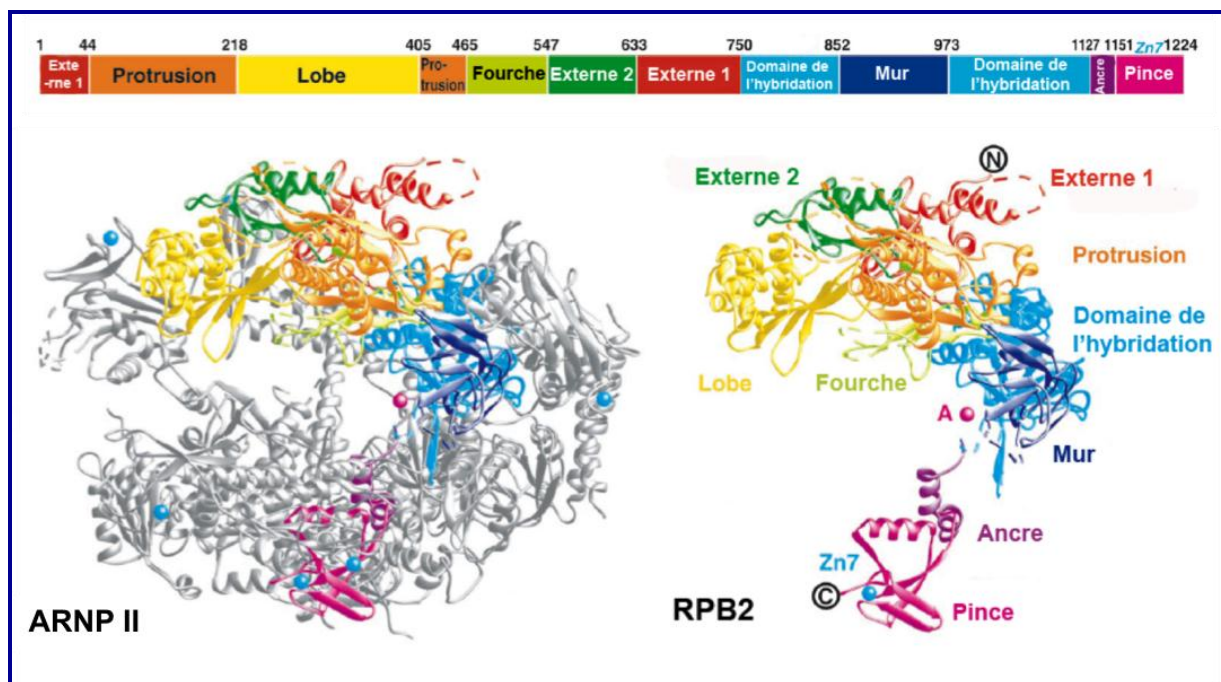


Figure 14 : Structure cristallographique de la sous-unité RPB2.

La position de la grande sous-unité RPB2 est présentée au sein du complexe de l'ARNP II. La sous-unité RPB2 isolée est colorée en fonction de ses différents domaines. Les ions Mg^{2+} (A) et Zn^{2+} sont représentés dans la structure par une sphère violet et bleu respectivement (Cramer et al., 2001).

L'analyse bio-informatique des interactions entre RPB1 et d'autres protéines dans la levure, a suggéré que RPB1 interagit avec un grand nombre de protéines impliquées dans différentes étapes de la transcription, modification et remodelage de la chromatine, réparation de l'ADN, et d'autres processus biologiques (Shukla et al., 2009).

Le domaine Carboxy-terminal de RPB1 (CTD)

Une des particularités de l'ARNP II par rapport aux autres ARNP, est la présence du domaine carboxy terminal de RPB1 (CTD) qui se caractérise par la répétition de l'heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser ($Y_1S_2P_3T_4S_5P_6S_7$). Ce motif est très conservé au cours de l'évolution. Cependant le nombre de répétitions varie en fonction des organismes de 26 chez *S. cerevisiae*, 42 chez la drosophile et 52 chez l'homme (Hampsey, 1998b; Lee and Young, 2000).

Le domaine CTD propre à l'ARNP II n'est pas toujours indispensable pour la transcription *in vitro* (Akoulitchev et al., 1995) mais il est essentiel pour l'activité de l'ARNP II *in vivo* (West and Corden, 1995). Sa délétion est létale chez la souris (Bartolomei et al., 1988), la drosophile (Zehring et al., 1988) et la levure (Allison et al., 1988). Chez la levure, le CTD représente une extension mobile et non structurée du «coeur» catalytique de l'ARNP II. Il est lié de manière flexible à la région adjacente au canal de sortie de l'ARN via un linker de 80 acides aminés et constitue une plate-forme pour l'interaction avec plusieurs protéines impliquées dans la formation de la coiffe en 5' de l'ARNm, l'épissage de l'ARNm, et la terminaison de la transcription (Ghosh et al., 2011; McCracken et al., 1997; Morris, 2000).

Les modifications post-traductionnelles du CTD dépendent de l'état de la polymérase et de l'étape de la transcription (figure 15) (Egloff and Murphy, 2008). Dans cette séquence peptidique, cinq résidus sur sept sont phosphorylables, mais la phosphorylation des sérines aux positions 2 et 5 joue un rôle particulièrement important pour la transcription (Hampsey, 1998b; Hirose and Ohkuma, 2007).

Une étude récente dans la levure *Schizosaccharomyces pombe* a montré que les acides aminés Tyr1, Pro3, Ser5 et Pro6 du CTD sont essentiels pour la survie, tandis que Ser2, Thr4, et Ser7 ne le sont pas. La sérine 2 est cependant essentielle pour la transcription pendant la différenciation sexuelle et la Ser5 est essentielle pour la croissance végétative et sa phosphorylation est nécessaire pour le recrutement des enzymes qui ajoutent la coiffe en 5' des ARNm (Schwer and Shuman, 2011).

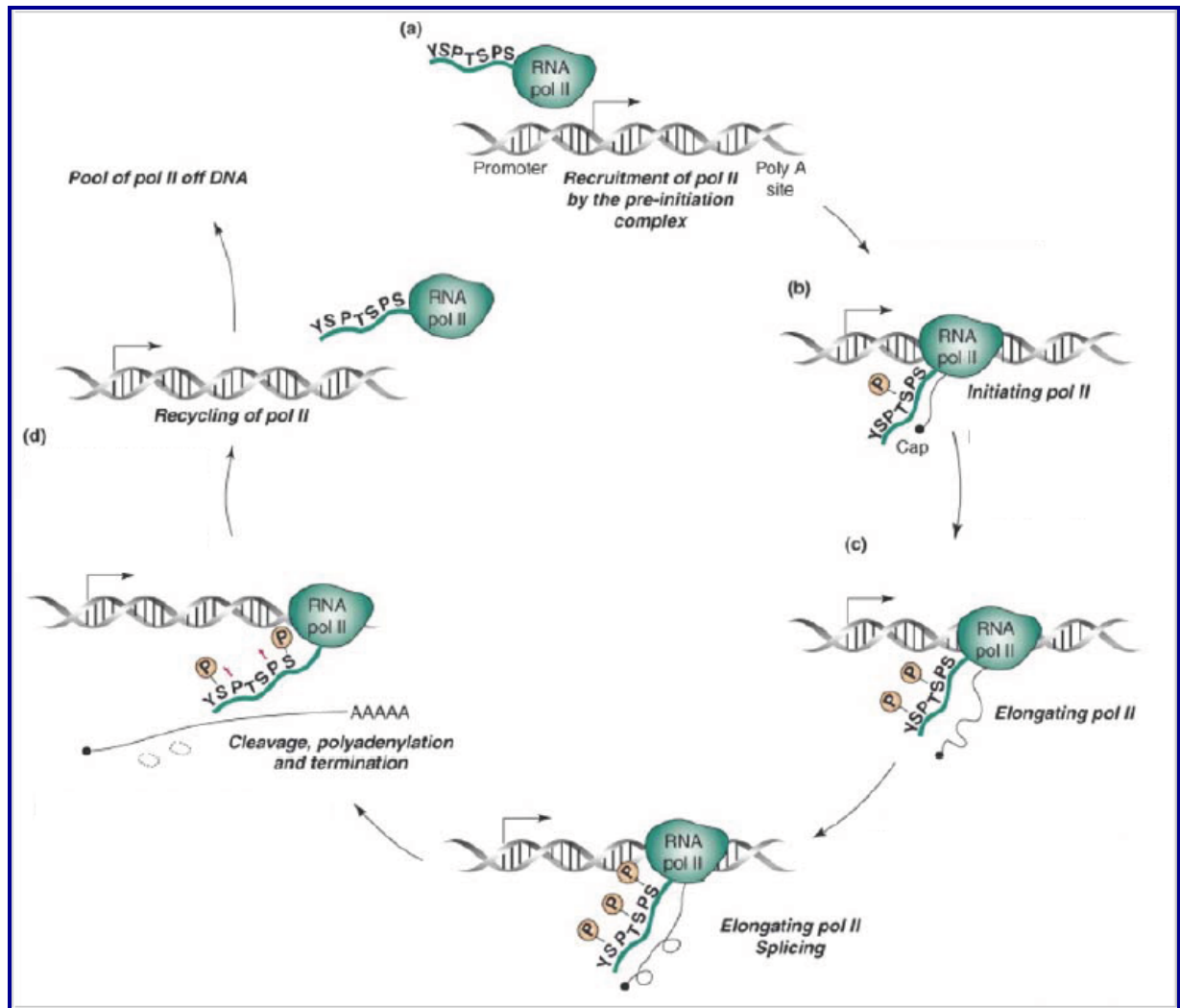


Figure 15 : Schématisation des modifications post-traductionnelles du CTD au cours de la transcription.

(a) La Pol II ayant le CTD non-phosphorylé est recrutée par le complexe de pré-initiation au niveau du promoteur. (b) Au cours de l'initiation, la phosphorylation de la Ser5 du CTD par CDK7 facilite le recrutement des enzymes qui ajoutent la coiffe (représentée par un rond noir) à l'extrémité 5' du transcrit émergent. (c) La phosphorylation progressive de la Ser2 par la sous-unité CDK9 du facteur P-TEFb stimule l'élongation et la maturation. La phosphorylation de la Ser7 croît au cours de l'élongation tandis que la Ser5 est progressivement déphosphorylée. (d) Après le clivage et la polyadénylation de l'extrémité 3' du pré-ARNm, le CTD est déphosphorylé. La Pol II peut alors se dégager pour un autre cycle de transcription (Egloff and Murphy, 2008).

L'ARNP II libre, avant d'être recrutée sur les promoteurs, possède un CTD qui est hypophosphorylé mais qui peut être glycosylé (Kelly et al., 1993; Phatnani and Greenleaf, 2006) dans le complexe de préinitiation. La phosphorylation des Ser5, qui atteint son maximum pendant l'initiation, alors que la Ser2 est phosphorylée au cours l'élongation (Phatnani and Greenleaf, 2006).

Plusieurs protéines kinases sont impliquées dans la phosphorylation de ces résidus. Les Ser5 ainsi que la Ser7 sont phosphorylées par la sous-unité CDK7 (*Cycline-Dependent Kinase 7*) du facteur TFIIH (Akhtar et al., 2009; Boeing et al.,

2010; Glover-Cutter et al., 2009; Kim et al., 2009; Komarnitsky, 2000). La phosphorylation de la Ser7 est indispensable pour l'association de l'ARNP II avec la machinerie de maturation des snRNA (Akhtar et al., 2009; Chapman et al., 2007; Egloff et al., 2010). La kinase CDK9, associée au facteur d'élongation de la transcription P-TEFb, phosphoryle la Ser2 (An et al., 2011; Peterlin and Price, 2006; Prelich, 2002), alors que la sous-unité CDK8 du complexe Médiateur mène à la formation d'une polymérase hyperphosphorylée qui serait inactive en transcription (Palancade and Bensaude, 2003). L'hélicase RECQ5 se lie spécifiquement aux Ser2 et 5 phosphorylées et joue un rôle dans la stabilité du génome lors de la transcription (Kanagaraj et al., 2010).

La polymérase est recyclée par déphosphorylation, différentes phosphatases ciblant des résidus spécifiques. La phosphatase Fcp1 (*TFIIF-associated CTD Phosphatase*) permettrait de déphosphoryler les Ser2 (Meinhart, 2005; Tombácz et al., 2009). La Ser5 est déphosphorylée par les phosphatases Ssu72 de CPF (*Cleavage/Polyadenylation Factor*) (Krishnamurthy et al., 2004), Scp1 (Yeo, 2003) et Rtr1 (Mosley et al., 2009). Aucune phosphatase n'est encore connue pour la Ser-7 phosphorylée.

Une étude récente a montré que la forme non phosphorylée du CTD lie l'ADN et que le complexe CTD/ADN est reconnu et phosphorylé par CDK7, mais pas par CDK9. Lorsqu'il est phosphorylé par CDK7, le CTD se dissocie de l'ADN et permet ainsi l'initiation de la transcription. Le CTD alors devient accessible pour la phosphorylation par CDK9 qui entraîne le passage vers l'élongation (Lolli, 2009).

Le CTD de l'ARNP II de *S. cerevisiae* est impliqué dans le recrutement de deux facteurs HATs (*Histone Acetyltransferase*) et HDAC (*Histone DeAcetylase*). Ceci entraîne le relâchement de la chromatine, la faisant passer à l'état euchromatinien, et permet ainsi une accessibilité de l'ADN au complexe de transcription (Spain and Govind, 2011).

Une étude récente a montré que le CTD de la polymérase humaine est méthylé à une arginine unique (R1810) par CARM1 (*Coactivator-Associated aRginine Methyltransferase 1*) *in vivo* et *in vitro*. Cette méthylation de R1810 est présente sur la polymérase hyperphosphorylée, en revanche la phosphorylation de Ser2 ou Ser5 inhibe l'activité de CARM1 (*in vitro*) vers ce site. Cette observation suggère que la méthylation se produit avant l'initiation de la transcription. Cette modification joue un rôle dans la régulation de l'expression des snRNA et snoRNA (Sims et al., 2011).

Le CTD peut être glycosylé par fixation d'un groupement monosaccharidique N-acétylglucosamine sur le groupe hydroxyle d'une sérine ou d'une thréonine (Kelly et al., 1993). Seule la forme hypophosphorylée de l'ARNP II peut être glycosylée. Cette glycosylation serait mutuellement exclusive avec la phosphorylation du CTD (Comer and Hart, 2001; Iyer et al., 2003). La glycosylation des résidus Thréonine/Sérine bloquerait stériquement l'accessibilité des kinases au CTD et régulerait ainsi l'activité de l'ARNP au cours du premier stade de l'initiation de la transcription. A l'opposé, une glycosylation étendue altérerait la conformation du CTD et modulerait ainsi la dynamique d'interaction des protéines avec le CTD (Thomas and Chiang, 2006).

Contrairement à la glycosylation qui est inhibée par la phosphorylation du CTD, la polyubiquitination de l'ARNP II est corrélée à une augmentation de la phosphorylation du CTD (Bregman et al., 1996). L'exposition des cellules à des agents qui induisent des dommages dans l'ADN entraînant une hyperphosphorylation du CTD. Dans ces conditions, l'ARNP II bloquée, est polyubiquitinée et entraîne le recrutement des protéines impliquées dans la voie de réparation (Harreman et al., 2009; Jeong et al., 2005; Ratner et al., 1998; Wu et al., 2001).

IV.2.2. Les sous-unités RPB3 (33 kDa) et RPB11 (13,3 kDa)

Les sous-unités RPB3 et RPB11 présentent des homologies avec la sous-unité α de l'ARNP d'*E. coli*. Deux homologues à α sont retrouvés dans les trois polymérases eucaryotiques, RPB3 et RPB11 pour l'ARNP II et AC40/AC19 pour les polymérases I et III. Les hétérodimères formés de RPB3/RPB11 ou AC40/AC19 seraient les homologues de l'homodimère α_2 présent dans l'ARNP eubactérienne. Comme la sous-unité α , la sous-unité RPB3 de *S. cerevisiae* est impliquée dans l'assemblage de l'enzyme.

En effet, une analyse génétique de l'assemblage a permis d'obtenir des mutations affectant les trois grandes sous-unités RPB1, 2 et 3. Il a été montré que la sous-unité RPB3 permet l'assemblage des deux grandes sous-unités via un complexe intermédiaire formé de RPB3 et RPB2 (Kolodziej and Young, 1991). Ce sous complexe stable pourrait ensuite recruter la sous-unité RPB1 et ainsi constituer,

comme pour l'enzyme eubactérienne, le premier intermédiaire d'assemblage d'un complexe stable (Boulon et al., 2010).

La protéine RPB3 est codée par un gène unique chez *S. cerevisiae*, la souris et l'homme. La sous-unité RPB11 est également codée par un gène unique chez la levure et la souris, mais au contraire par une famille multigénique chez l'homme (Grandemange et al., 2001).

Il est intéressant de noter que l'hétérodimère RPB3/11 joue un rôle important dans l'interaction de l'ARNP II avec le complexe protéique médiateur (Davis et al., 2002). Chez *S. cerevisiae*, des expériences de mutagenèse de RPB3/RPB11 suggèrent que le signal de terminaison serait transduit par l'hétérodimère RPB3/RPB11 (Steinmetz et al., 2006). La sous-unité RPB3 joue aussi un rôle dans l'activation de la transcription inductible toujours chez la levure (Tan et al., 2000). Il a été montré par des expériences en double hybride que HsRPB3 interagit avec les facteurs de transcription ATF4 (De Angelis et al., 2003) et myogénine, spécifique de cellules musculaires (Corbi et al., 2002).

HsRPB3 augmente la transactivation d'ATF4 et son niveau d'expression augmente durant la différenciation musculaire dans des lignées cellulaires C2C7 (De Angelis et al., 2003). L'expression de RPB3 est régulée lors de la différenciation musculaire (Corbi et al., 2002). Cette sous-unité a une localisation subcellulaire particulière qui est régulée au cours de la différenciation musculaire. En plus de sa présence dans le complexe ARNP II, RPB3 s'accumule comme une sous-unité libre dans le cytoplasme des cellules C2C7, ainsi que dans le cytoplasme de plusieurs autres lignées cellulaires, y compris les cellules HeLa (lignée cellulaire provenant d'un cancer du col de l'utérus) (Corbi et al., 2005).

La protéine (IGFBP)-3 (*Insulin-like Growth Factor-Binding Protein 3*) est impliquée dans la prolifération et/ou la différenciation des myoblastes. Elle interagit avec RPB3 ce qui faciliterait le recrutement de l'ARNP II et de facteurs de transcription spécifiques et serait nécessaire à la transactivation des gènes. (Oufattole et al., 2006).

Une corrélation a été mise en évidence, chez l'homme, entre le taux d'expression de la sous-unité HsRPB11 et la sensibilité à une drogue anticancéreuse, la doxorubicine (Bruno et al., 1998; Fanciulli et al., 1996). En effet, la surexpression de la sous-unité HsRPB11 dans des cellules résistantes à la doxorubicine et qui présentent un faible taux de l'ARNm codant cette sous-unité

entraîne une augmentation de sensibilité à la drogue sans changer la sensibilité à d'autres drogues telles que l'étoposide ou le cisplatine. De plus, cette étude montre que la surexpression entraîne une diminution de la prolifération cellulaire.

Cette sous-unité pourrait avoir un rôle dans la régulation de la transcription de gènes spécifiques codant, par exemple, des marqueurs de différenciation. La même équipe, par des techniques de double hybride, a caractérisé une interaction entre HsRPB11 et HsRPB3, mais aussi avec la kératine 19 et la protéine Che-1 (Bruno et al., 1999; Fanciulli et al., 2000; Fanciulli et al., 1998). Takramah et ses collaborateurs ont montré une interaction entre HsRPB11a et P35 une protéine anti apoptotique de baculovirus. Dans des cellules humaines de carcinome du colon, P35 est capable, après transfection, de doubler l'activité des promoteurs E-cadhérine et β -actine. En présence de HsRPB11a, cette activité est augmentée d'un facteur 3 à 4 (Takramah et al., 2003).

IV.2.3. Les sous-unités RPB4 (16 kDa) et RPB7 (19 kDa)

Les sous-unités RPB4 et RPB7 de *S. cerevisiae* forment un sous-complexe dissociable de l'enzyme. La structure tridimensionnelle du complexe RPB4/RPB7 de levure a été déterminée à 2,3 Å (Armache et al., 2005). La structure cristallographique du dimère RPB4/7 humain a été résolue à 2,7 Å (Meka et al., 2005). Elle est similaire à l'hétérodimère de levure et l'homologue E/F de l'archéobactérie *Methanocaldococcus jannaschii* (Figure 16). La présence de domaines putatifs de liaison à l'ARN sur la sous-unité RPB7 et la position de l'hétérodimère à la sortie de l'ARN dans le complexe polymérase II de levure suggèrent fortement un rôle dans la liaison et la stabilisation de l'ARN naissant (Meka et al., 2005; Újvári and Luse, 2006).

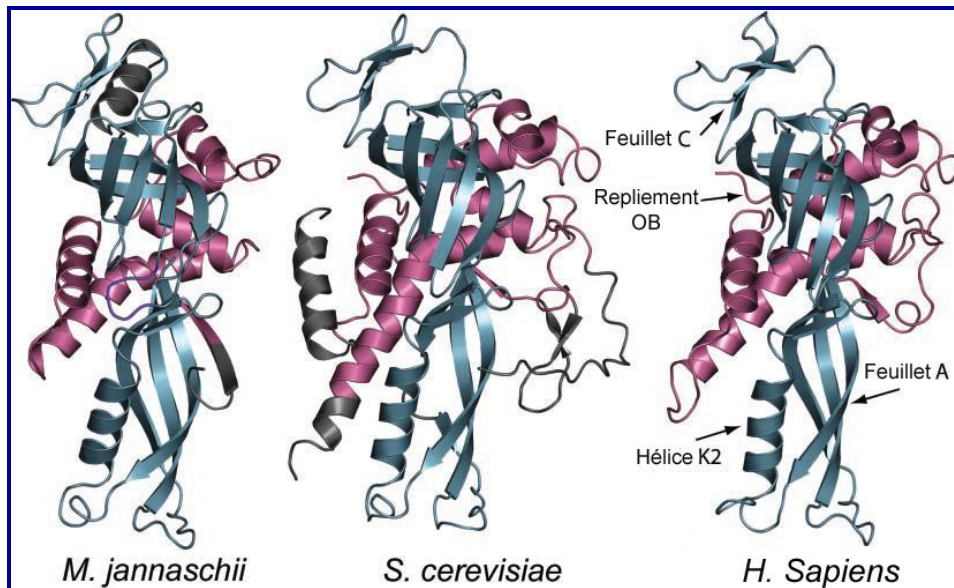


Figure 16 : Structures de complexes RPB4/7.

Structure du complexe E/F de *Methanocaldococcus jannaschii*, celle du complexe RPB4/7 de levure et de l'Homme. Les domaines structurellement conservés de RPB4 et RPB7 sont colorés en rose et violet respectivement. Les régions des structures de levure et d'archéobactérie qui diffèrent de la protéine humaine sont en gris (Meka et al., 2005).

Une ARN polymérase Δ RPB4/7 présente une activité indistinguable de l'ARNP sauvage dans un test de transcription non spécifique *in vitro* (Ruet et al., 1980). Par contre, cette même polymérase est inactive dans un test de transcription spécifique (Edwards et al., 1991). Les études récentes ont montré que les profils d'occupation de l'ensemble du génome par RPB7 et RPB3 sont essentiellement identiques (Jasiak et al., 2008). Cependant, on observe pour le mutant Δ RPB4 de *S. cerevisiae*, une réduction du niveau d'ARNP II près de l'extrémité 3' de messagers de plusieurs gènes, de même qu'un faible recrutement des facteurs protéiques de terminaison de la transcription. De plus, la perte de RPB4 induit une altération de l'usage du site de polyadénylation du gène *RAN14* (Runner et al., 2008) et un défaut de transcription des séquences de grande taille riches en paires de bases GC (Verma-Gaur et al., 2008).

Ces résultats indiquent le rôle de RPB4 dans le processus de maturation des ARNm (l'extrémité 3') et aussi dans l'élongation de la transcription.

RPB4 et RPB7 de *S. cerevisiae* sont également impliquées dans la dégradation de l'ARNm dans le cytoplasme (Lotan et al., 2007; Selitrennik et al., 2006).

Le complexe E/F de l'archéobactérie (homologue RPB4/7 des eucaryotes) stimule la processivité de l'ARNP en diminuant la fréquence de pause. Il a aussi un

rôle dans la terminaison de la transcription (Grohmann and Werner, 2011; Hirtreiter et al., 2010).

IV.2.4. La sous-unité RPB5 (25 kDa)

La sous-unité RPB5 est partagée par les trois ARNP eucaryotiques (Woychik et al., 1990). La structure de la sous-unité de *S. cerevisiae* a été résolue à 1,9 Å (figure 17) (Todone et al., 2000). Elle a été comparée aux structures homologues de sous-unités H, de l'ARNP d'archéobactéries (Thiru et al., 1999; Yee et al., 2000). La similitude de séquences de RPB5 à travers l'évolution des eucaryotes et même chez les archéobactéries, indique que sa fonction a bien été conservée.



Figure 17 : Structure cristallographique de la sous-unité RPB5 de *S.cerevisiae*.

Le domaine d'interaction avec TFIIB est représenté en bleu. Le domaine cible de HBx est représenté en rouge (Todone et al., 2000).

Le domaine C-terminal de cette protéine est fixé sur RPB1 (Zaros et al., 2007). Ce domaine interagit avec l'ADN (Cramer, 2000). La sous-unité humaine (HsRPB5) est spécifiquement associée *in vitro* et *in vivo* par son domaine N-terminal avec la protéine X du virus humain de l'hépatite B (HBx). Cette interaction semble stabilisée grâce à des contacts supplémentaires avec le facteur de transcription TFIIB (Cheong et al., 1995; Todone et al., 2000). HsRPB5 interagit également avec la sous-unité RAP30 du complexe TFIIF et cette interaction contribue à l'association entre TFIIF et

l'ARNP II (Le, 2005; Wei, 2001). Il a été prouvé que la protéine RMP (*RPB5 Mediation Protein*) qui joue un rôle dans la prolifération des cellules de carcinome hépatocellulaire humain est associée à la sous-unité RBB5 (Dorjsuren et al., 1998; Yang et al., 2011).

IV.2.5. La sous-unité RPB6 (14,4 kDa)

RPB6 est une sous-unité essentielle et commune entre les trois polymérases nucléaires. Cette sous-unité est très conservée à travers les espèces. Ceci est confirmé par la complémentation sans phénotype d'une souche Δ RPB6, aussi bien par la sous-unité RPB6 de *S. pombe* (Shpakovski, 1994) que par celle de l'homme (Shpakovski et al., 1995) et du hamster (McKune and Woychik, 1994). Cette sous-unité est phosphorylée dans les trois polymérases (Bell et al., 1976, 1977). La caséine kinase II (CKII) pourrait être responsable de cette phosphorylation *in vivo* (Kayukawa et al., 1999).

La structure de HsRPB6 a été établie par résonance magnétique nucléaire (figure 18). Elle comprend un domaine central très conservé formant deux hélices α antiparallèles. Ce domaine est flanqué par des régions N et C-terminales qui forment une structure de 5 feuillets β en sandwich (del Rio-Portilla et al., 1999).

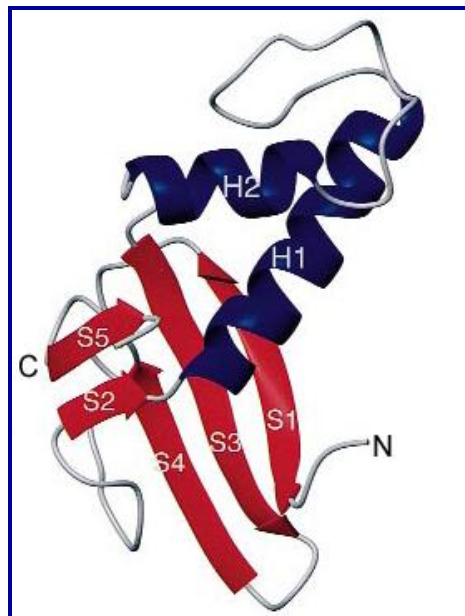


Figure 18 : Structure de la sous-unité HsRPB6 établie par résonance magnétique nucléaire.

Sur ce modèle en rubans, les hélices α (H1 et H2) sont colorées en bleu et les feuillets β en rouge. Les deux hélices sont connectées par une boucle (boucle ω), qui n'a pas de structure secondaire définie (del Rio-Portilla et al., 1999).

En accord avec les données génétiques, une homologie de structure tridimensionnelle entre la sous-unité RPB6 et la sous-unité ω de l'ARNP de l'*E. coli* a été constatée (Minakhin et al., 2001). Il est proposé que cette sous-unité permettrait et/ou stimulerait l'assemblage des trois polymérases eucaryotiques par l'intermédiaire de son interaction avec les trois grandes sous-unités (A190, RPB1 et C160). En dehors de son rôle dans l'assemblage des polymérases, cette sous-unité joue un rôle dans l'interaction entre l'ARNP II et le facteur d'élongation de la transcription TFIIIS (Ishiguro et al., 2000). L'examen de la structure tridimensionnelle de l'ARNP II (Cramer, 2000) a montré que RPB6 interagit directement avec RPB4 et constitue l'un des points de contact entre le sous-complexe RPB4/RPB7 et le cœur de l'ARNP II (Tan et al., 2003).

IV.2.6. La sous-unité RPB8 (17,1 kDa)

La sous-unité RPB8 est essentielle et modérément conservée au cours de l'évolution. Cette sous-unité est présente dans les trois types d'ARNP. Elle a la particularité de n'avoir pas d'homologue évident dans les polymérases des archéobactéries. Cependant deux études récentes montrent que RPB8 est un orthologue d'une protéine appelé G, spécifiquement présente chez les Crénarchées, une sous-groupe des archéobactéries (Koonin et al., 2007; Kwapisz et al., 2008). La structure tridimensionnelle de la sous-unité RPB8 de *S. cerevisiae* et RPB8 humaine (figure 19) a été résolue par résonance magnétique nucléaire (Kang et al., 2006; Krapp et al., 1998). Les auteurs proposent l'existence d'une surface potentielle d'interaction avec des oligonucléotides ou le domaine OB (*Oligonucleotide/Oligosaccha Binding*). La structure montre une interaction non spécifique entre l'ADN simple brin et HsRPB8 qui peut par ailleurs adopter des changements conformationnels significatifs (Kang et al., 2006).

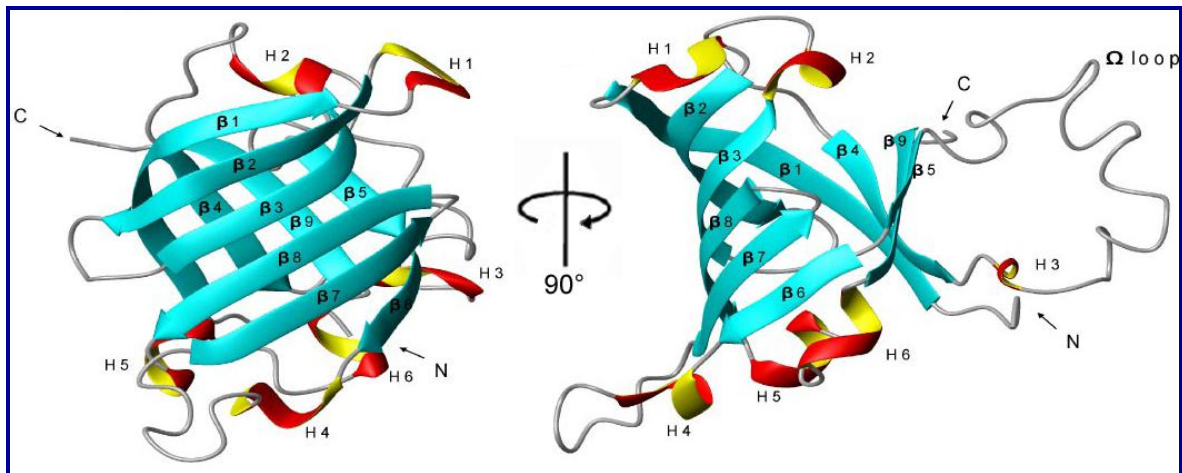


Figure 19 : Structure tridimensionnelle de HsRPB8.

Elle se compose de 8 feuillets β antiparallèles colorés en bleu. Les 6 courtes hélices sont colorées en rouge et jaune. La grande boucle Ω est non structurée (Kang et al., 2006)

L'expérience de suppression génétique de RPB8 de *S. pombe* et *S. cerevisiae* a permis de démontrer que cette sous-unité est impliquée dans l'assemblage de l'ARNP III (Voutsina et al., 1999). Il a été montré que le phénotype thermosensible d'une souche complétée avec la sous-unité RPB8 humaine peut-être diminué par la surexpression de RPB6 (Briand et al., 2001). Cette suppression rend compte d'une interaction fonctionnelle entre ces deux protéines. La sous-unité RPB6 est comme la sous unité RPB8 impliquée dans l'assemblage des complexes enzymatiques. Ainsi, le défaut d'assemblage de la polymérase III lié à la sous-unité RPB8 peut être compensé par la grande sous-unité C160 elle-même, mais aussi par une protéine RPB6 qui viendrait stabiliser C160 au sein du complexe (Voutsina et al., 1999).

Dans des cellules HeLa, en réponse à un dommage de l'ADN, le complexe BRCA1-BARD1 ((BRCA1, *Breast Cancer 1*) (BARD1, *BRCA1-Associated RING Domain protein 1*)) interagit et induit la polyubiquitination de RPB8 *in vivo* et *in vitro*. BRCA1 est une protéine suppresseur de tumeur (Wu et al., 2007).

IV.2.7. La sous-unité RPB9 (14,5 kDa)

Cette sous-unité, bien que spécifique de l'ARNP II, possède une homologie avec les sous-unités A12.2 et C11 des ARNP I et III respectivement. RPB9 est composée de trois domaines distincts: deux domaines de liaison au zinc (N et C-terminal), et le bras central. La structure de la sous-unité homologue à RPB9 chez

les archéobactéries a été déterminée (figure 20) (Wang et al., 1998). Elle ne possède que la partie contenant le motif de liaison au zinc de la partie C-terminale. Les auteurs constatent une remarquable ressemblance entre cette structure et celles de TFIIS et TFIIB. La sous-unité RPB9 n'est pas essentielle chez la levure, mais sa délétion entraîne un phénotype thermo- et cryo-sensible ainsi qu'une croissance fortement ralentie (Hazoume et al., 2011; Woychik et al., 1991).

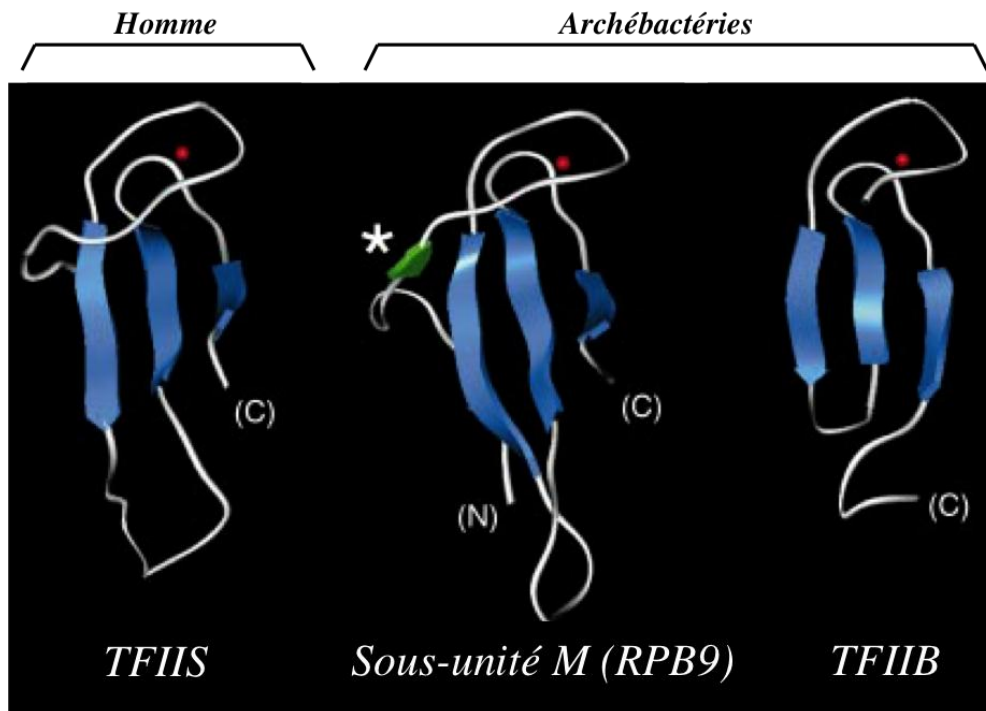


Figure 20 : Comparaison des structures des parties C-terminales de la sous-unité M (RPB9) et TFIIB d'archéobactérie et TFIIS humain.

Les feuilletés β homologues sont représentés en bleu. Un quatrième feuillet est représenté en vert dans RPB9. L'ion zinc est représenté dans chaque structure par un point rouge (Tableau 2) (Wang et al., 1998).

L'étude de la sous-unité RPB9 montre son implication dans diverses étapes de la transcription. L'implication de RPB9 dans l'initiation de la transcription et particulièrement dans la reconnaissance du site d'initiation, a été observée (Furter-Graves et al., 1994; Hull et al., 1995; Sun et al., 1996) comme par exemple celui du gène IMD2 de levure codant l'IMP déshydrogénase (Jenks et al., 2008). Le motif de liaison du zinc en position N-terminale est impliqué dans l'initiation de la transcription. Le deuxième motif de liaison du zinc en position C-terminale de la protéine, homologue à celui de TFIIS (Hemming and Edwards, 2000; Kaine et al., 1994), est impliqué dans l'étape d'élongation.

TFIIS stimule l'activité intrinsèque de clivage d'un transcrit naissant par l'ARNP

II. Cette stimulation est augmentée par RPB9. Des travaux récents démontrent que cette activité de clivage contribue au maintien de la fidélité de la transcription (Nesser et al., 2006). En effet, chez la levure, la disruption du gène *DST1* codant TFIIIS et celle de RPB9 diminuent la fidélité de la transcription. Par ailleurs, l'expression de protéines mutantes de RPB9 incapables d'augmenter l'activité de TFIIIS, ne restaure pas le phénotype alors que celle de RPB9 sauvage le peut. Ce phénotype est plus sévère dans le cas d'un double mutant. Ceci suggère que RPB9 maintient la fidélité de la transcription via deux mécanismes: l'augmentation de la stimulation de l'activité de clivage TFIIIS-dépendant et un autre mécanisme indépendant de la fonction de TFIIIS (Koyama et al., 2007; Nesser et al., 2006).

La sous-unité RPB9 et la protéine de levure RAD26 (facteur stimulant de la réparation de l'ADN) sont impliquées de manière indépendante dans le phénomène de réparation couplée à la transcription (TCR, *Transcription Coupled Repair*) (Li et al., 2006a; Li and Smerdon, 2004). Le motif de liaison au zinc en position N-terminale et le domaine central de RPB9 sont essentiels pour la TCR (Li et al., 2006b).

Walmacq et ses collaborateurs ont démontré que le mutant Δ RPB9 favorise la séquestration du NTP (*Nucleoside Tri Phosphate*) dans le centre actif de la polymérase juste avant la formation d'une liaison phosphodiester. Ils ont proposé un modèle dans lequel la sous-unité RPB9 contrôle la fidélité de la transcription en retardant la fermeture de la « trigger loop » à l'entrée de NTP via une interaction entre le domaine C-terminal de RPB9 et la « trigger loop » (Walmacq et al., 2009).

Une étude récente a montré que l'activité de clivage de l'ARNP II augmente lorsque le domaine de liaison au zinc dans la partie C-terminale de RPB9 est remplacée par son homologue de la sous-unité de l'ARNP III (C11). Les ARNP I et ARNP III contiennent un domaine catalytique de liaison au zinc dans la partie C-terminale qui complètent le site actif de l'ARNP, tandis que l'ARNP II contient un domaine non-catalytique de liaison au zinc qui est immobilisé sur la surface de l'enzyme (Ruan et al., 2011).

IV.2.8. La sous-unité RPB10 (7,6 kDa)

Cette sous-unité commune aux trois classes de polymérase est la plus conservée. Elle possède une partie C-terminale basique et un motif de liaison du zinc conservé, chez les eucaryotes et les archéobactéries (Mackereth et al., 2000).

La structure tridimensionnelle de RPB10 de l'archéobactérie *Methanobacterium thermoautotrophicum* a été résolue par résonance magnétique nucléaire (figure 21) (Mackereth et al., 2000). La sous-unité a une structure dite en «boule à zinc» qui consiste en 3 hélices α stabilisées par un ion zinc. L'ion zinc se fixe par un motif atypique (CX₂CX_nCC) et sert de pont entre la boucle N-terminale et l'hélice 3 (Mackereth et al., 2000).

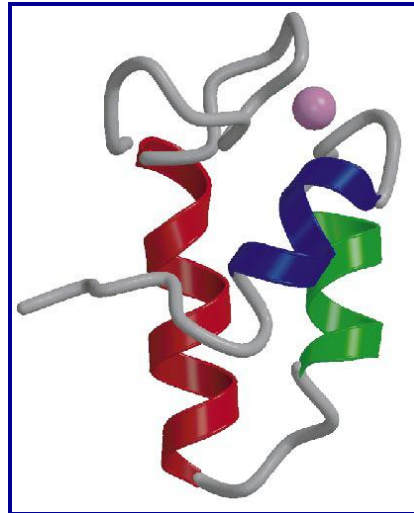


Figure 21: Structure de la sous-unité RPB10 de l'archéobactérie *Methanobacterium thermoautotrophicum*.

Les 3 hélices α sont colorées en rouge, vert et bleu. L'atome de zinc est représenté par la petite sphère rose (Mackereth et al., 2000).

La complémentation de RPB10 de *S. cerevisiae* par les sous-unités homologues de l'homme et de *S. pombe*, indique que la conservation de séquences entre ces sous-unités s'étend à une conservation de fonction (Shpakovski et al., 2000; Shpakovski et al., 1995).

IV.2.9. La sous-unité RPB12 (7 kDa)

Cette sous-unité est commune aux trois classes d'ARNP. Les sous-unités HsRPB12 et ScRPB12 présentent 34% d'identité et la sous-unité humaine comme la sous-unité de la levure *S. pombe* peuvent remplacer la sous-unité ScRPB12 (Shpakovski et al., 2000; Shpakovski et al., 1995). D'autres études dans la levure *S. cerevisiae* (Lalo et al., 1993) ou dans les archéobactéries (Eloranta et al., 1998) ont montré que cette sous-unité est souvent associée aux sous-unités RPB3/RPB11 et AC40/AC19.

Le motif de liaison du zinc de sous-unité RPB12 conservé chez toutes les espèces est essentiel à la survie de *S. cerevisiae* (Das et al., 2006; Gadad et al., 1999). Alors que la partie C-terminale basique est essentielle, seule la première cystéine est essentielle pour la croissance de *S. cerevisiae* (Rubbi et al., 1999). Une analyse de l'interaction protéine-protéine a montré que la sous-unité RPB12 interagit avec TFIIIE β chez *S. pombe* (Hayashi et al., 2005).

La sous-unité RPB12 peut interagir avec les sous-unités A135, RPB2 et C128 des trois ARNP eucaryotiques. Des mutants dans la partie C-terminale de RPB12, comprenant le motif de liaison du zinc, présentent une incapacité de liaison à la sous-unité C128 et un défaut d'assemblage de l'ARNP III (Rubbi et al., 1999).

En plus de son rôle dans l'assemblage, l'implication de la sous-unité RPB12 humaine dans l'activation transcriptionnelle a été suggérée (Schlegel et al., 2000). En effet, une interaction directe a été montrée entre le coactivateur BRCA1 et les sous-unités HsRPB2 et HsRPB12. Ces sous-unités pourraient donc avoir un rôle dans l'activation transcriptionnelle spécifiquement liée à BRCA1 (voir aussi IV.2.6.)

Une récente étude montre que la protéine NS5A du virus de l'hépatite C réprime l'expression du gène HsRPB12 et que la modulation de la transcription de la cellule hôte par NS5A serait induite par cette répression (Jung et al., 2007).

La sous-unité P de l'ARNP de l'archéobactérie sous le contrôle d'un promoteur fort de levure supprime le phénotype létal d'un mutant Δ RPB12. La sous-unité RPB12 de la levure peut remplacer fonctionnellement la sous-unité P lors de la reconstitution de l'ARNP de l'archéobactérie (*Pyrococcus furiosus*) (Reich et al., 2009).

IV.3. Structure tridimensionnelle de l'ARN Polymérase II

Les premières études cristallographiques de l'ARNP II ont été effectuées chez la levure en utilisant l'enzyme formée de dix sous-unités (c'est-à-dire toutes sauf RPB4 ET RPB7) (Cramer, 2000; Cramer et al., 2001; Fu et al., 1999). La structure cristallographique de la polymérase complète, avec ses 12 sous-unités, a été obtenue avec des complexes purifiés à partir de cellules en phase stationnaire (figure 22) (Armache et al., 2003; Armache et al., 2005; Bushnell and Kornberg, 2003; Jensen et al., 1998; Kettenberger et al., 2004).

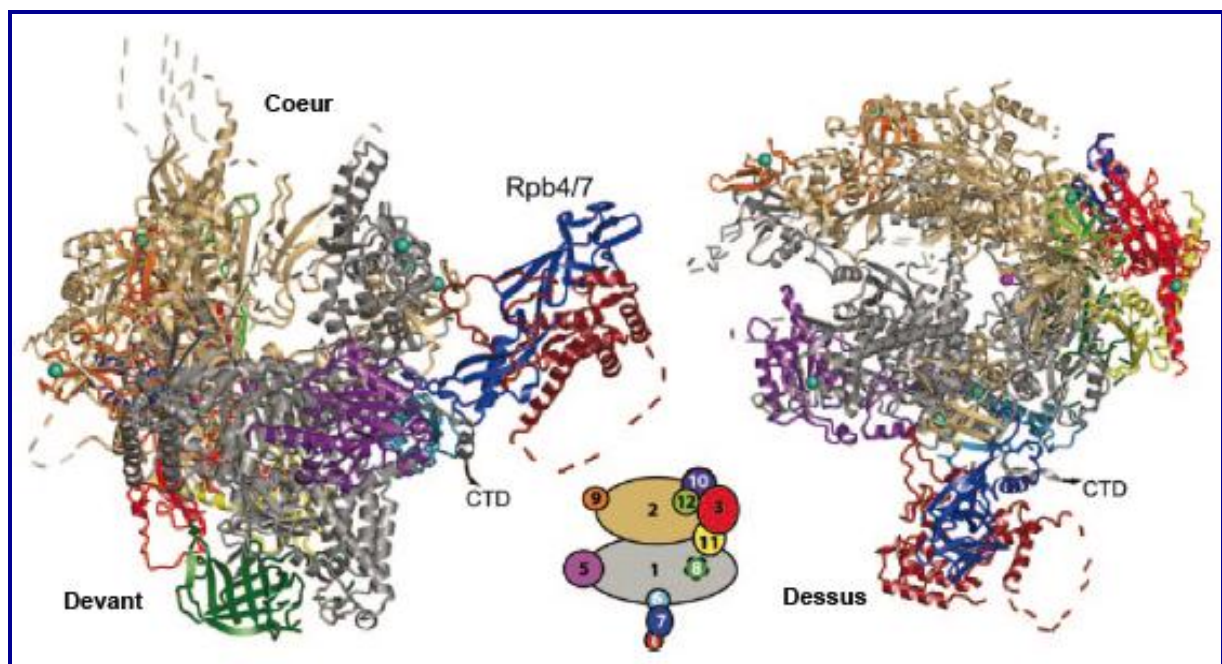


Figure 22 : Structure de l'ARNP II complète.

Le code couleur des 12 sous-unités est donné dans le schéma central. Les lignes en pointillés représentent des boucles non structurées. Les ions zinc sont montrés par des sphères bleues et l'ion magnésium du site actif par une sphère rose (Armache et al., 2005).

Deux autres approches ont été utilisées afin d'obtenir des cristaux de l'ARN Pol II à 12 sous-unités : la première a utilisé un complexe portant une étiquette TAP (*Tandem Affinity Purification*) de Rpb4 (Bushnell and Kornberg, 2003), alors que la deuxième a utilisé des complexes purifiés d'une lignée $\Delta rpb4$ qui ont été incubés par la suite avec les sous-unités RPB4 et RPB7 recombinantes (Armache et al., 2003; Armache et al., 2005).

Les deux plus grandes sous-unités RPB1 et RPB2 constituent l'essentiel de masse de l'enzyme et forment un sillon (cleft) chargé positivement par lequel s'introduit l'ADN matrice (figure 23). Les autres sous-unités, plus petites, sont arrangées à l'extérieur de l'enzyme. L'entrée du sillon est bordée par des mâchoires formées en partie de RPB1, RPB9 et RPB5. L'un des côtés du sillon forme une pince mobile (clamp) qui pourrait adopter une conformation ouverte ou fermée selon qu'elle interagit ou non avec l'ADN.

Le domaine C-terminal de la sous-unité RPB1 de l'ARNP II (CTD) n'est pas structuré, mais se situe probablement au niveau de la face amont, sous la pince mobile. Le site actif est localisé au niveau du mur (wall) définissant le fond du sillon

dans lequel se trouvent deux ions Mg^{2+} . Une hélice de RPB1 appelée l'hélice de pontage (bridging helix) traverse le sillon et rejoint RPB2. Des acides aminés de cette hélice se retrouvent en contact direct avec la base codante du brin transcrit de l'ADN formant une structure de type « fermeture-éclair ». La flexibilité de cette hélice est importante pour maintenir l'interaction ADN-protéine pendant la translocation de l'ARNP II (Cramer et al., 2001).

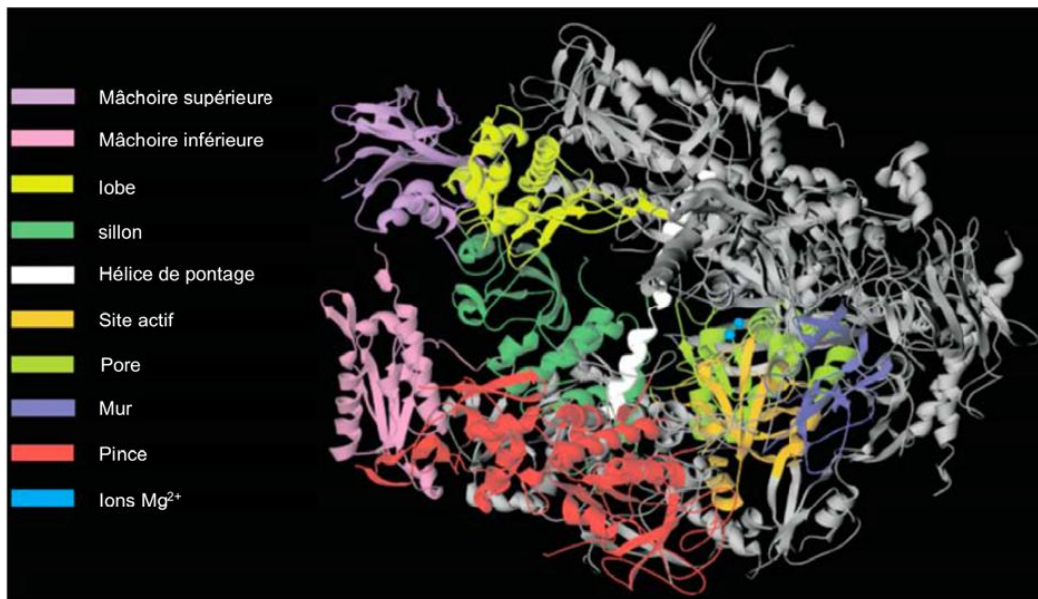


Figure 23 : Structure cristallographique de l'ARNP II de *S. cerevisiae*

L'enzyme présente un sillon (RPB1) bordé à son entrée par une paire de mâchoires, supérieure (partie de RPB1 et RPB9) et inférieure (RPB5), et au fond duquel se trouvent deux ions Mg^{2+} qui matérialisent l'emplacement du site actif. Près du site actif, l'hélice de pontage (RPB1) traverse le sillon pour rejoindre RPB2. Le pore localisé à l'arrière de l'enzyme permettrait l'entrée des ribonucléotides et la sortie de l'extrémité 5' libre de l'ARN (Cramer et al., 2001).

L'hétérodimère formé par RPB4 et RPB7 forme une structure qui ressort de la surface de la polymérase et est située entre la pince mobile et le domaine C-terminal (CTD) de la sous-unité RPB1. Cet hétérodimère s'attache au reste de l'enzyme par principalement des contacts entre RPB7 et RPB6 et RPB1 (Armache et al., 2003; Armache et al., 2005; Bushnell and Kornberg, 2003; Craighead et al., 2002).

Afin de décrire le mécanisme de sélection et d'incorporation des nucléosides triphosphates (NTP) dans le complexe, deux modèles principaux ont été proposés. Dans le premier modèle, le NTP atteint le site de pré-insertion, en conformation ouverte du centre catalytique, ce qui induirait un changement de conformation de la boucle « trigger loop » et la fermeture subséquente du centre actif (Kettenberger et

al., 2004).

Dans un deuxième modèle, le NTP se lierait tout d'abord dans un site d'entrée du pore situé au dessous du centre catalytique et dans une deuxième étape il serait orienté par le Mg^{2+} (B) directement dans le site d'insertion (Westover et al., 2004). Le premier modèle a été confirmé par des études plus récentes de l'ARNP II eucaryote (Brueckner et al., 2009; Wang et al., 2006) ou bactérienne (Vassilyev et al., 2007).

L'alpha-amanitine est l'une des mycotoxines du champignon vénéneux *Amanita phalloides*. A faible concentration, l' α -amanitine est un inhibiteur sélectif des ARNP II eucaryotiques (Kedinger et al., 1970). La structure cristallographique de l'ARNP II de la levure *S. cerevisiae* complexée à l' α -amanitine a été résolue à 2,8 Å (figure 24). Les résidus de l'ARNP II en contact avec l' α -amanitine sont situés dans l'hélice de pontage « bridging helix » entre les deux grandes sous-unités RPB1 et RPB2 (Bushnell et al., 2002).

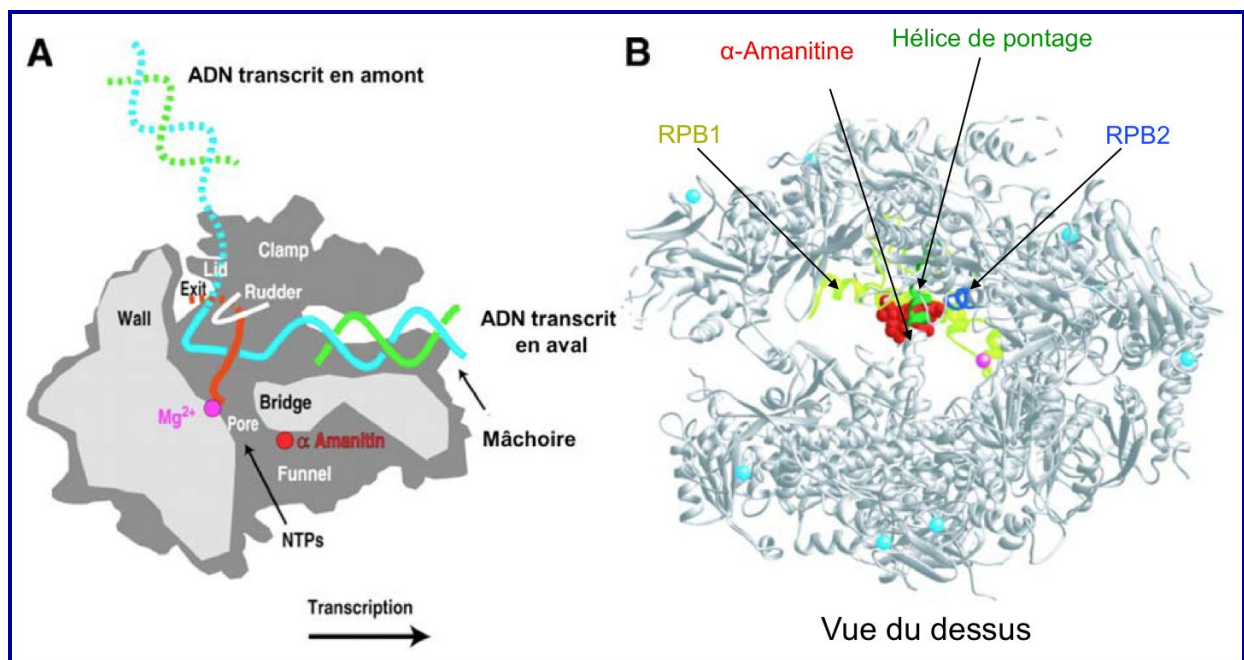


Figure 24 : Localisation de l' α -amanitine fixée sur l'ARNP II

A : Vue schématisée en coupe du complexe de l'ARNP II en cours de transcription et montrant la fixation de l' α -amanitine (sphère rouge) ainsi que les acides nucléiques et certains éléments structuraux du site actif de l'enzyme. L'ARN est figuré en rouge. B : Représentation en rubans de la structure du complexe l'ARNP II / α -amanitine. Les résidus en contact avec α -amanitine sont montrés. Les atomes de zinc sont montrés par des sphères bleu-clair et l'ion magnésium du site actif par une sphère rose.

IV.4. L'assemblage de l'ARN polymérase II

Le mécanisme par lequel chacune de ces 12 sous-unités de la polymérase est assemblée dans le complexe n'est pas encore clairement élucidé. Les études récentes ont montré que les deux sous-unités RPB1 et RPB3 s'accumulent dans le cytoplasme des cellules traitées à l' α -amanitine, ce qui provoque la dégradation de la grande sous-unité RPB1 (Boulon et al., 2010; Bushnell et al., 2002; Nguyen et al., 1996).

L'import nucléaire de RPB1 nécessite la présence de toutes les sous-unités. La grande sous-unité RPB1 libre dans le cytoplasme serait prise en charge par la protéine chaperon HSP90 et son co-facteur hSpagh. Ces protéines stabilisent RPB1 libre dans le cytoplasme. L'ARNP II est alors assemblée dans le cytoplasme et puis les enzymes entièrement assemblées sont importées dans le noyau (Boulon et al., 2010; Corden, 2011).

Chez la levure *S. cerevisiae*, l'import nucléaire de la Pol II nécessite la protéine *lwr1* (*Putative transcription factor lwr1*). *lwr1* se lie à l'ARNP II dans le centre actif entre les deux grandes sous-unités. Cette protéine possède un signal de localisation nucléaire (NLS, *Nuclear Localization Signal*) (Kosugi et al., 2009) qui est reconnu par la karyophérine α , ce qui permet de diriger la polymérase vers le noyau. La présence de l'ADN et des facteurs d'initiation de la transcription induit la dissociation du complexe *lwr1*/ARNP II et libère l'enzyme. La protéine *lwr1* ainsi dissociée est exportée vers le cytoplasme grâce à sa séquence NES (*Nuclear Export Signal*) (Czeko et al., 2011; Kosugi et al., 2009).

La protéine 4 associée à l'ARNP II (RPAP4/GPN1), joue la navette entre le noyau et le cytoplasme et régule l'importation nucléaire de deux grandes sous-unités de la polymérase humaine. Cette protéine est le membre d'une famille GTPase nouvellement découverte qui contient un motif GPN (GPN-loop GTPase 1) unique et hautement conservé. Ce motif est essentiel pour la localisation nucléaire de HsRPB1 dans un processus qui exige également l'assemblage des microtubules (Forget et al., 2010).

Les données disponibles à l'heure actuelle ne sont pas encore suffisantes pour décrire un cheminement complet, mais sont compatibles avec l'existence de deux intermédiaires assemblés dans le cytoplasme avant de produire l'enzyme complète (figure 25) (Corden, 2011).

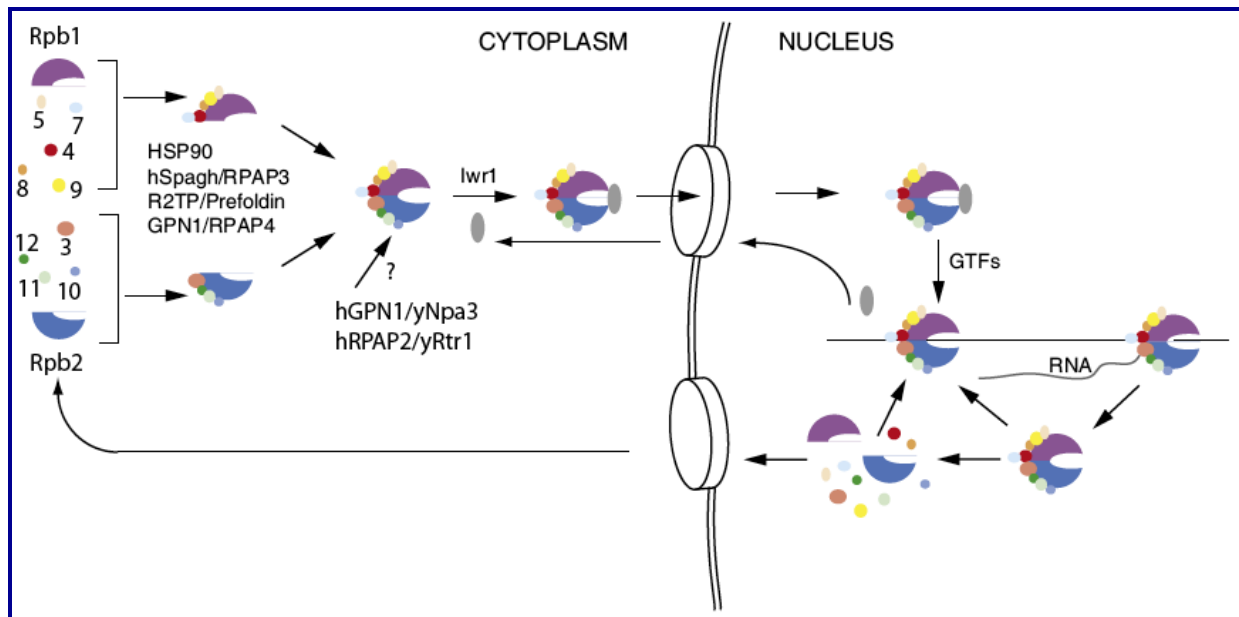


Figure 25 : Assemblage et transport de l'ARNP II.

Les sous-unités individuelles de l'ARNP II sont indiquées en formes colorées. Les intermédiaires potentiels formés au cours de l'assemblage contiennent soit RPB1 soit RPB2 et un ensemble de petites sous-unités. Ces complexes contiennent aussi les facteurs d'assemblage putatifs indiqués par leur nom. Dans le noyau, soit l'enzyme se fixe sur un promoteur, soit elle se décompose en unités libres. Ces sous-unités peuvent être, soit transportées vers le cytoplasme pour un nouvel assemblage, soit assemblées sur le promoteur comme cela a été montré pour la Pol I (Corden, 2011; Dundr et al., 2002).

V. La sous-unité HsRPB11

V.1. Organisation du gène et isoformes

La localisation chromosomique du gène codant pour cette sous-unité a été identifiée au sein de notre laboratoire en utilisant une sonde génomique comprenant la partie 5' du gène. Ceci a permis d'identifier trois locus qui sont situés sur le chromosome 7 (7p12, 7q22 et 7q11.23), indiquant la présence de plusieurs gènes indépendants (figure 26) (Grandemange et al., 2001).

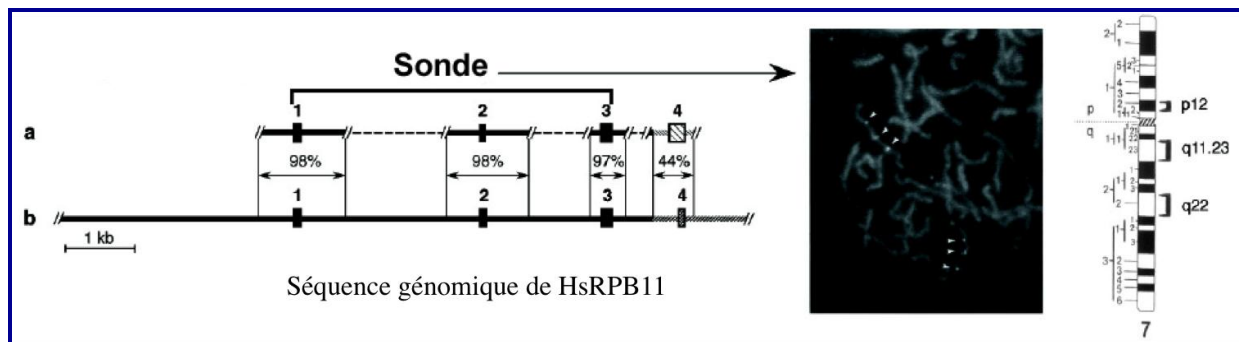


Figure 26 : Organisation génomique de HsRPB11.

Les lignes horizontales représentent les séquences génomiques de HsRPB11 humaines. Les exons sont identifiés par des boîtes noires. Les homologies entre les types a et b sont indiqués. La métaphase présentée sur la droite a été hybridée avec une sonde génomique comprenant la partie 5' du gène de HsRPB11a. Les têtes de flèches blanches indiquent la position des locus liés spécifiquement. Le chromosome 7 est représenté avec la localisation des trois locus (Grandemange et al., 2001).

Il existe dans le génome humain une famille de 4 gènes (POLR2J1 à 4) susceptible de coder plusieurs isoformes de la sous-unité HsRPB11 de la Pol II. Cette famille n'existe pas chez la souris dont le génome contient un gène RPB11 unique. Une étude récente a identifié le gène POLR2J (7q22) comme un nouvel oncogène potentiel dans le cancer du poumon. L'activation de ce gène a été confirmée chez un panel de tissus pulmonaires tumoraux comparés à des tissus pulmonaires normaux correspondants (Campbell et al., 2008b). La caractérisation de la séquence génomique codant la sous-unité HsRPB11 a permis l'identification de deux types de gènes qui sont appelés : type « a » et type « b » (figure 26).

Le type « a » ou HsRPB11a contient la séquence codante de l'ARNm précédemment caractérisée (Acker et al., 1997; Fanciulli et al., 1996).

Le type de gène « b » appelé HsRBP11b. La transcription de ce gène donne lieu à deux isoformes nommés HsRPB11b α et HsRPB11b β (Grandemange et al., 2001).

Le gène POLR2J1 code la sous-unité RPB11a dont la séquence peptidique est identique à son orthologue de souris. Cette protéine est retrouvée dans la Pol II fonctionnelle. Les gènes POLR2J2 et 3 codent les isoformes appelés HsRPB11b α et HsRPB11b β . Le gène POLR2J4 a perdu le codon ATG, et serait plutôt à considérer comme un pseudogène.

V.2. L'épissage alternatif de messenger HsRPB11 et l'expression tissu-spécifique

Les gènes POLR2J2, 3 et 4 sont transcrits et les niveaux de leurs messagers varient d'un tissu à l'autre, mais les protéines correspondantes n'ont pas encore été détectées. Ces transcrits sont soumis à des épissages alternatifs, ce qui n'est pas le cas des transcrits de POLR2J1. La différence entre ces deux types de gènes HsRPB11a et HsRPB11b se situe dans la partie 3'. Le messenger HsRPB11b β diffère de b α par l'excision de l'exon 3 suite à un mécanisme d'épissage alternatif. Ceci change la phase de lecture, l'exon 4 devenant entièrement codant. Le codon stop se retrouve alors dans le cinquième exon (figure 27 A) (Grandemange et al., 2001).

Il est intéressant de remarquer que l'analyse des séquences montre que les polypeptides HsRPB11a, HsRPB11b α et b β ont des tailles similaires : 117, 115 et 116 résidus avec un poids moléculaire d'environ 13,3, 13, et 12,7 kDa respectivement. La partie N-terminale de la sous-unité HsRPB11a diffère du polypeptide HsRPB11b par la présence d'une lysine supplémentaire à la jonction entre les exons 1 et 2.

En revanche, les parties C-terminale de ces polypeptides sont très différentes (figure 27). Des mutations ponctuelles du domaine C-terminal de la sous-unité RPB11 de levure se traduisent par un phénotype mutateur (résultats non publiés).

L'exon 4 de HsRPB11a code un peptide de 11 résidus hydrophiles, par contre dans le cas de HsRPB11b α , il génère un peptide hydrophobe de 10 résidus.

Concernant HsRPB11b β , en raison de l'absence de l'exon 3, un peptide totalement différent riche en Proline (16%), Alanine (14,5%), Acide glutamique (9%), Histidine (9%) et Cystéine (7%) est produit (figure 27 B) (Grandemange et al., 2001).

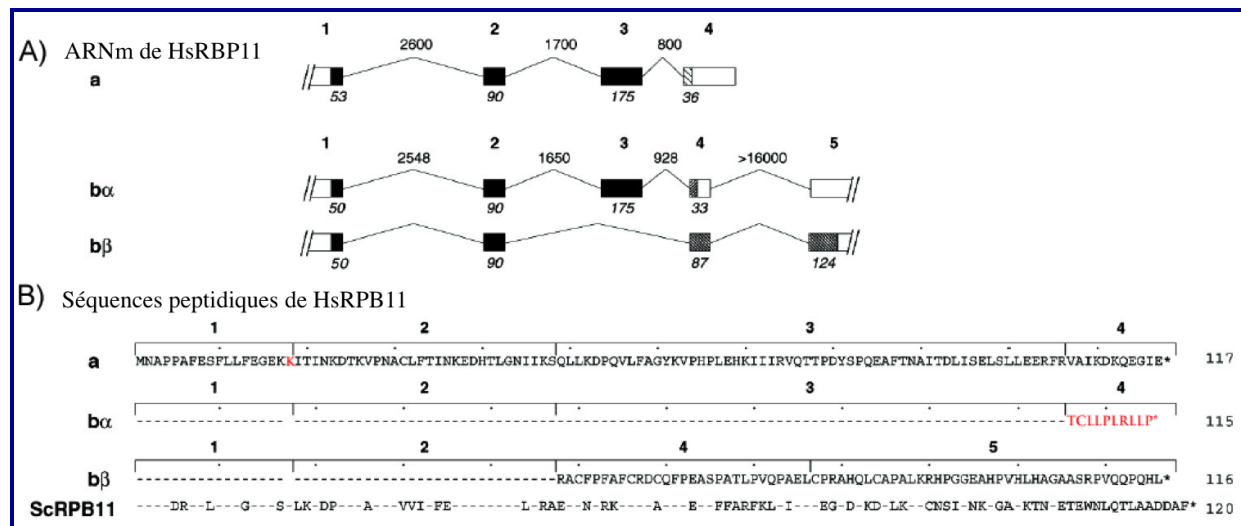


Figure 27 : Structure de l'ARN messager et la séquence peptidique de HsRPB11.

A) Structure de l'ARN messager de HsRPB11a, b α et b β . Les exons sont indiqués par des boîtes. Les tailles des introns sont indiquées en paire de base (pb). La partie 5' et 3' des régions non codantes sont indiqués par des boîtes vides. La taille (pb) de la séquence codante (CDS) dans chaque exon est indiquée. **B)** Séquences d'acides aminés des polypeptides HsRPB11a, b α et b β . Les séquences codantes des ARNm identifiées pour les deux gènes (a et b) sont alignées avec leur nom et leur taille (acides aminés), indiquée sur la gauche et à droite, respectivement. Les bords des exons pour chaque partie de la séquence et leurs numéros correspondants sont indiqués. La séquence de HsRPB11a a été utilisée comme référence et la lysine supplémentaire à la jonction entre les exons 1 et 2 est montrée en rouge. Seulement les résidus qui sont différents sont indiqués. La séquence de RPB11 de *S. cerevisiae* (ScRPB11) est montrée ci-dessous (Grandemange et al., 2001).

Les différents transcrits des gènes POLR2J ont été détectés par l'analyse de Northern-blot et RT-PCR dans la plupart des tissus humains. L'expression de ces ADNc a été testée dans 16 tissus humains. Une bande majeure a été détectée avec chaque sonde dans tous les tissus. Les niveaux relatifs d'expression de HsRPB11a ainsi que HsRPB11b varient selon les tissus. La transcription de HsRPB11a est majeure dans plupart des tissus avec des niveaux plus élevés dans le cœur et le muscle squelettique. Son expression par contre est plus réduite dans le colon, le poumon, intestin grêle et l'utérus. En revanche, l'ARNm de HsRPB11b α est le plus abondant dans le cerveau. La transcription de HsRPB11b β est faible dans tous les

tissus, mais plus importante dans le cœur, les muscles squelettiques et de l'ovaire (Grandemange et al., 2001).

Fanciulli et ses collaborateurs ont montré la surexpression des messagers de HsRPB3 et 11 dans certains tissus particulièrement dans le cœur, les muscles squelettiques après l'addition de doxorubicine. La doxorubicine est une drogue souvent utilisée dans les traitements anticancéreux. Néanmoins son utilisation clinique est limitée par l'apparition, entre autres, d'une toxicité cardiaque et de tumeurs résistantes à cette drogue. Cette drogue peut exercer un effet inhibiteur *in vivo* spécifiquement sur les composants majeurs de la machinerie transcriptionnelle (Fanciulli et al., 1998).

V.3. Propriétés des trois protéines HsRPB11

Les interactions entre l'isoforme HsRPB11b et les sous unités de l'ARNP II ont été étudiées et comparées à celles observées entre HsRPB11a et les autres sous-unités de l'enzyme. La partie N-terminale codée par les exons 1 et 2, contient d'un motif conservé appelé le « motif α » responsable de l'interaction RPB3/RPB11 (figure 28). Les différentes isoformes sont tous les trois capables d'interagir avec RPB3, ce qui n'est pas totalement inattendu puisque le motif α est présent pour toutes. L'isoforme HsRPB11b α , comme HsRPB11a, n'interagit qu'avec HsRPB3. Par contre, HsRPB11b β interagit également avec d'autres sous-unités : HsRPB4, 5, 6, 7 et 12 (Grandemange et al., 2001).

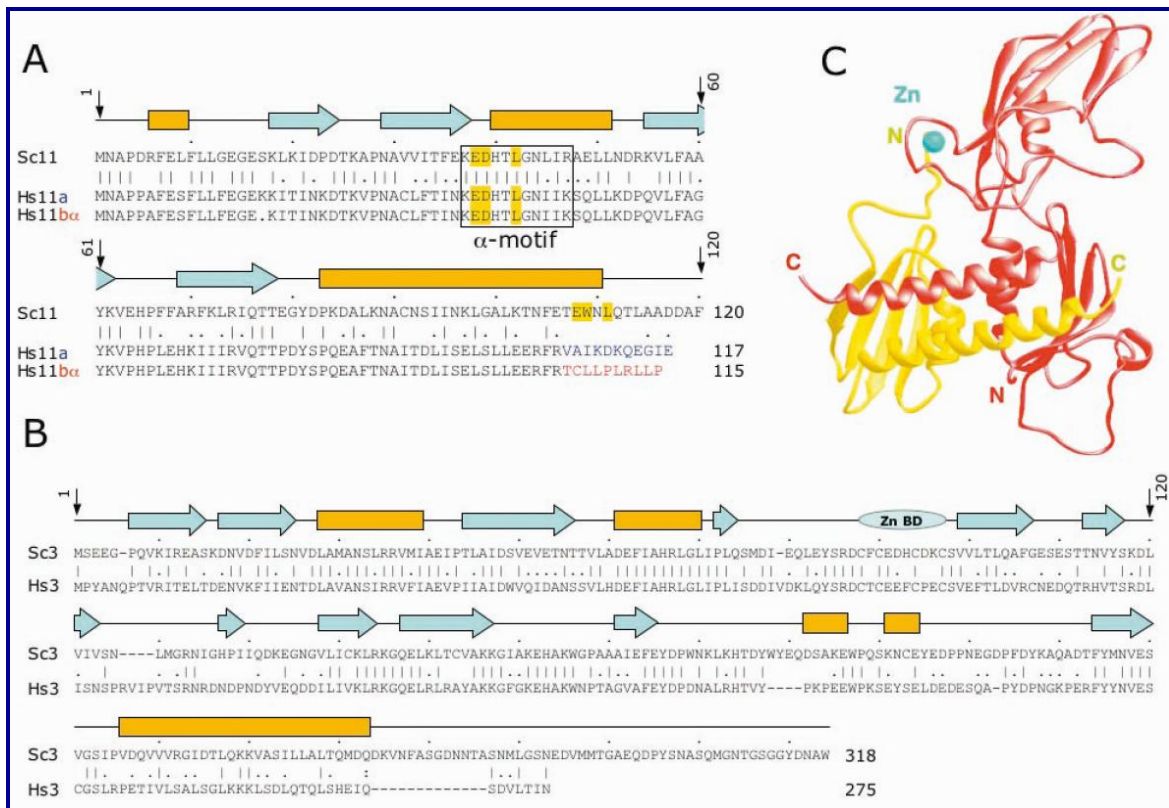


Figure 28 : Structure de la sous-unité RPB3 et RPB11 de la levure et l'homme.

(A) Alignement des séquences peptidiques de ScRPB11 (Sc11), HsRPB11a (Hs11a) et HsRPB11b α (Hs11b α). Les homologies des séquences entre la levure et l'homme sont présentées. Les barres verticales et des points indiquent les résidus identiques et similaires, respectivement. L'élément conservé (motif α) est encadré. Les résidus surlignés en jaune ont été soumis à la mutagenèse dirigée. Les hélices alpha et les feuillets β sont présentés par des boîtes jaunes et les flèches bleues, respectivement. **(B)** Alignement des séquences peptidiques de Sc3 et Hs3. La position d'un domaine de liaison au zinc (Zn BD) est montrée. **(C)** La structure tridimensionnelle de l'hétérodimère Sc3/Sc11. Les rubans rouge et jaunes représentent les polypeptides Sc3 et Sc11, respectivement. Le zinc (en bleu), coordonné par quatre résidus cystéine de domaine de liaison au zinc de Sc3 est indiqué (Benga et al., 2005).

Des expériences de complémentation dans la levure *S. cerevisiae* Δ RPB11 ont montré que le variant HsRPB11b α est capable de s'intégrer à l'ARNP pour former un complexe actif *in vivo*. Par contre, ni HsRPB11b β , ni HsRPB11a, qui est le polypeptide détecté dans l'enzyme humaine, ne sont capables de compléter dans la levure (Grandemange et al., 2001).

Les isoformes HsRPB11b α et HsRPB11b β présentent des domaines (correspondant à l'exon 4) fortement homologues à des domaines présents dans des protéines HsPMS2L (*Humain Post-Meiotic Segregation 2 Like*) (Kondo et al., 1999). Ces protéines présentent elle-même une forte homologie avec des séquences peptidiques de la protéine HsPMS2, impliquée dans les mécanismes de réparation de l'ADN par « Mismatch Repair » (Kolodner and Marsischky, 1999). Les protéines HsPMS2L appartiennent également à une famille de gènes localisés dans les mêmes régions que les gènes HsRPB11 sur le chromosome 7. Il est fortement probable que la famille de gènes HsRPB11 provienne d'évènements de recombinaison entre les gènes HsRPB11a, les gènes HsPMS2L et un autre gène (uroplakin) correspondant à l'exon 5 de HsRPB11b.

L'analyse comparative des hétérodimérisations des sous-unités RPB3 et RPB11 de l'homme et la levure a été effectuée par la méthode des double-hybrides. L'hétérodimérisation des sous-unités HsRPB3/HsRPB11a, HsRPB3/HsRPB11b α et ScRPB3/ScRPB11 était facilement détectable, tandis que dans les mêmes conditions l'hétérodimère HsRPB3/ScRPB11 ne pouvait pas être détecté. Cette absence d'interaction explique l'incapacité de HSRPB3 pour à remplacer ScRPB3 dans la levure. Par contre la sous-unité ScRPB3 peut interagir avec HsRPB11a et HsRPB11b α et les résidus qui sont importants pour ces interactions ont été identifiés. Des résidus de ScRPB3 qui sont importants dans l'hétérodimérisation de ScRPB3/HsRPB11a et ScRPB3/HsRPB11b α ont été identifiés. Ces résidus n'ont aucun effet sur l'interaction ScRPB3/ScRPB11 (Benga et al., 2005).

La capacité de HsRPB11b α , à la fois de remplacer la sous-unité ScRPB11 dans la levure et à interagir avec ScRPB3 avec une efficacité similaire à celle de ScRPB11 indique que HsRPB11b α a le potentiel pour s'intégrer dans le complexe d'ARNP II dans les cellules humaines et aussi chez la levure.

Les mutations dans le motif α ont un effet négatif sur toutes les combinaisons testées avec HsRPB11a et HsRPB11b α tandis qu'aucun effet sur les sous-unités de

la levure n'a pas été détecté. L'hétérodimère ScRPB3/ScRPB11 dépend fortement des résidus spécifiques dans la partie C-terminale contrairement aux hétérodimères HsRPB3/HsRPB11a ou $b\alpha$ qui ne semblent pas s'appuyer autant sur ce contact (figure 29) (Benga et al., 2005).

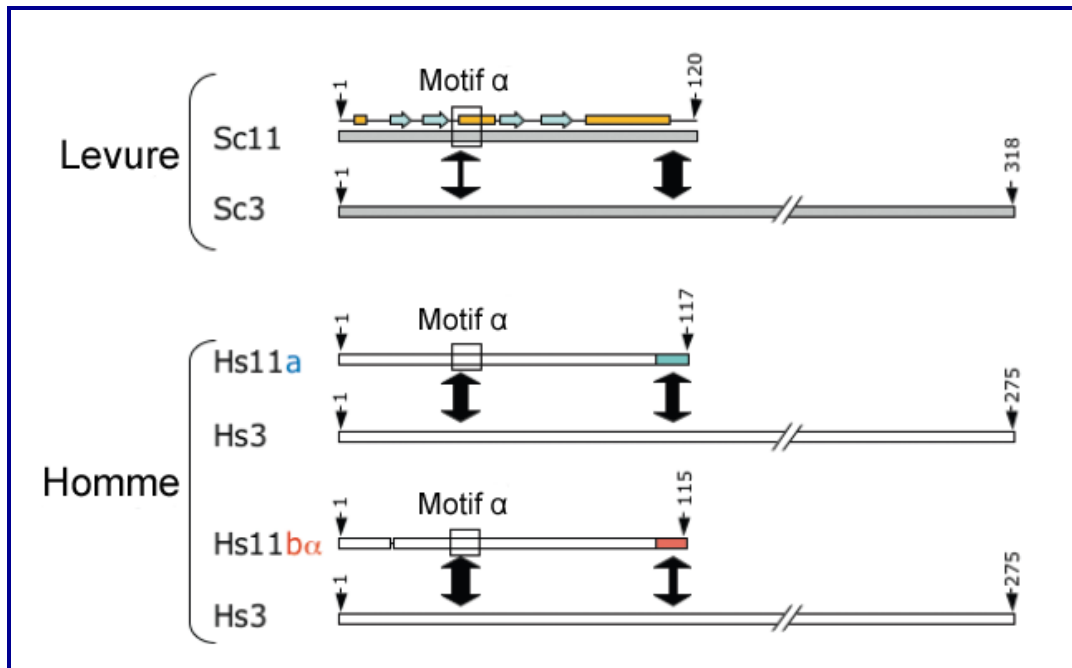


Figure 29 : Résumé des contacts identifiés entre les sous-unités RPB3 et RPB11 chez l'homme et chez la levure et leurs contributions à la stabilité des hétérodimères.

Les contributions des domaines homologues de Sc11, Hs11a, Hs11 $b\alpha$ dans hétérodimérisation avec des sous-unités Sc3 et Hs3 sont représentées par des flèches à double tête avec des épaisseurs en fonction de l'importance des contacts (Benga et al., 2005).

Ces données indiquent que les interfaces de l'hétérodimérisation ont été modifiées au cours de l'évolution d'une manière qui rend possible une diversification de la sous-unité HsRPB11 qui reste compatible, avec, l'interaction avec la sous-unité HsRPB3 (Benga et al., 2005).

VI. Objectif du travail de thèse

Cette introduction illustre l'importance de la sous-unité HsRPB11 dans le complexe de l'ARNP II. Cette sous-unité est une cible d'interaction pour certaines protéines notamment Che1, kératine 19 et P35 de baculovirus (Bruno et al., 1999; Fanciulli et al., 2000; Takramah et al., 2003). Le gène qui code cette protéine peut être surexprimé dans le cancer du poumon (Campbell et al., 2008b).

L'hétérodimérisation de HsRPB11 avec HsRPB3 joue un rôle structural fondamental dans l'assemblage de l'enzyme (Fanciulli et al., 1998; Kimura et al., 1997). La structure 3D de l'ARNP II humaine a été établie par cryomicroscopie électronique à 22 Å (Kostek et al., 2006). Cependant, très peu d'études détaillées sur la structure et l'assemblage du complexe de l'ARNP II sont disponibles à ce jour.

L'objectif de cette thèse était de participer à la détermination de la structure tridimensionnelle de l'ARNP II humaine en collaboration avec le Dr. Averell Gnat (*University of Maryland School of Medicine*). Dans le but de démarrer cette étude, nous avons décidé de résoudre la structure tridimensionnelle de la sous-unité isolée HsRPB11 par une approche cristallographique.

Au cours de la purification de cette protéine, nous avons constaté qu'elle forme d'un dimère très stable. Nous avons tenté d'analyser cette homodimérisation dans la levure *S. cerevisiae* et aussi dans les cellules humaines.

Enfin, nous avons entrepris dans une dernière partie de notre projet de thèse un essai de détection de l'isoforme HsRPB11b α en utilisant une approche de siRNA.

Résultats et Discussion

Chapitre 1 : Etude structurale de la sous-unité HsRPB11a par cristallographie

I. Clonage, expression, purification et cristallisation de HsRPB11a

Le succès de l'étape de cristallisation est déterminé par la qualité du matériel produit et purifié. La démarche qui a été utilisée est celle d'un processus itératif, dans lequel les étapes d'expression et de purification sont modifiées en fonction des résultats obtenus après les essais de cristallisation.

I.1. Clonage

À mon arrivée au laboratoire, la séquence codante de la sous-unité HsRPB11a était cloné dans le vecteur d'expression pNIR sous le contrôle d'un promoteur inductible de l'opéron nir, qui est impliqué dans l'assimilation du nitrate dans les cyanobactéries. Cet opéron est induit en présence de nitrate et efficacement réprimé par l'ammonium (Frias et al., 1997).

Ce gène a également été cloné dans un vecteur d'expression pET sous le contrôle d'un promoteur du phage T7 (Studier et al., 1990) pour produire les protéines chez *E.coli* BL21 (DE3). La souche BL21 (DE3) porte le gène codant l'ARNP de T7 en aval du promoteur *lac* de telle sorte que la polymérase du phage ne soit synthétisée qu'après induction par l'IPTG. Les protéines qui ont été produites dans les bactéries ou les cyanobactéries sont toutes munies d'une étiquette 6xHis en N-terminal qui rend possible leur purification par chromatographie d'affinité sur résine Ni-NTA.

I.2. Expression chez la cyanobactérie *Anabaena* et chez *E.coli* BL21 (DE3)

A mon arrivée, le système d'expression inductible de protéines recombinantes dans les cyanobactéries genre *Anabaena* avait été mis en place (Desplancq et al., 2001). Ce système a permis d'obtenir sous forme soluble des protéines qui formaient des agrégats insolubles chez *E.coli* (Desplancq et al., 2005).

Le principal inconvénient de ce système d'expression est le temps de division des cellules qui est relativement long (24h). Nous avons purifié la protéine HsRPB11a à partir d'un extrait de cyanobactéries. Cette protéine a été cristallisée. Les cristaux obtenus ont été testés par diffraction des rayons X. Les clichés de

diffraction obtenus n'ont pas donné de résultats exploitables, car un empilement irrégulier des molécules dans le cristal a été observé.

Pour avoir plus de matériel et dans un temps plus court, nous avons testé la production et la solubilité de cette protéine chez *E.coli*. Différentes conditions de culture (température : 20, 30, 37°C) ont été testées sur la souche bactérienne *E.coli* BL21 (DE3). Après induction, les bactéries ont été lysées par sonication. L'extrait total a été clarifié par centrifugation (5000 g). Le surnageant contenant la fraction soluble a été analysé par électrophorèse sur gel de polyacrylamide, en conditions dénaturantes (figure 30). Dans chacun des cas, la protéine HsRPB11a était présente dans la fraction soluble. La présence des protéines a été aussi vérifiée par Western-Blot en utilisant des anticorps spécifiques dirigés contre HsRPB11a (non montré).

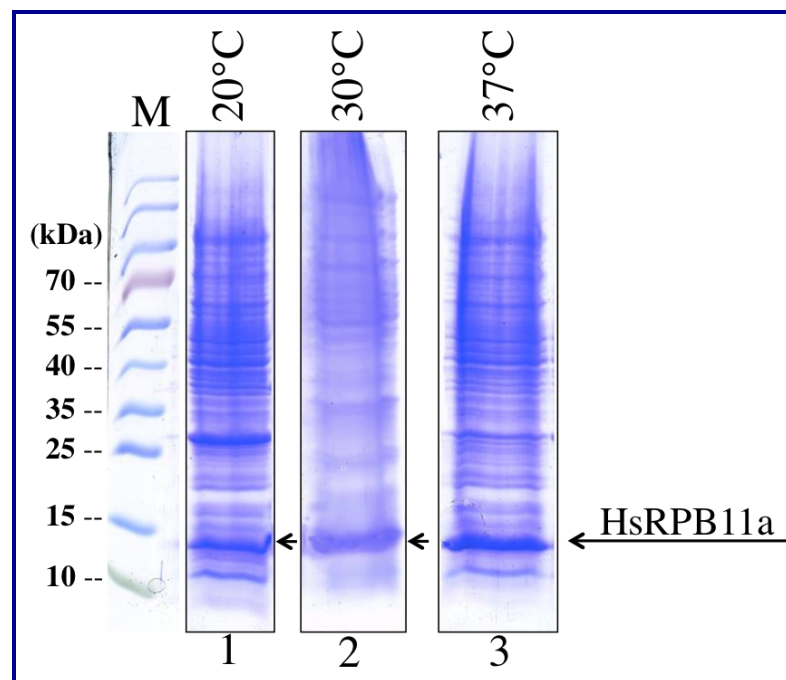


Figure 30 : Analyse SDS-PAGE de l'expression des protéines chez *E.coli* BL21(DE3).

Les échantillons correspondant aux surnageants de sonication ont été déposés sur un gel en gradient d'acrylamide (10-20%) dans des conditions dénaturantes. Le gel est coloré au bleu de coomassie. Les pistes 1, 2 et 3 correspondent aux prélèvements des échantillons incubés à 20°, 30° et 37°C pendant une nuit respectivement. M : marqueur de poids moléculaires en kDa. Les bandes (14,6 kDa) correspondant aux protéines surproduites sont indiquées par une flèche.

I.3. Purification des protéines

Les bactéries provenant d'un litre de culture incubé à 37° C et induites à l'IPTG (400 µM) ont été mises en suspension dans un tampon de lyse (20mM Tris pH 8, 500mM NaCl, 0,5mM β-mercaptoéthanol) contenant 20mM d'imidazole et lysées par sonication en présence d'inhibiteurs de protéases (Complete, Roche). Après une étape de clarification par ultracentrifugation (30 min à 31000 g), le lysat est incubé en «batch» avec la résine Ni-NTA-agarose (QIAGEN) à 4° C. Après un lavage en présence d'imidazole (20mM), les protéines fixées à la résine sont éluées par des concentrations croissantes d'imidazole ajustée à pH 8 dans le tampon Tris-HCl (100mM). Les protéines sont ensuite analysées par SDS-PAGE et les fractions d'intérêt (5 à 9) sont réunies pour une étape de gel-filtration (figure 31).

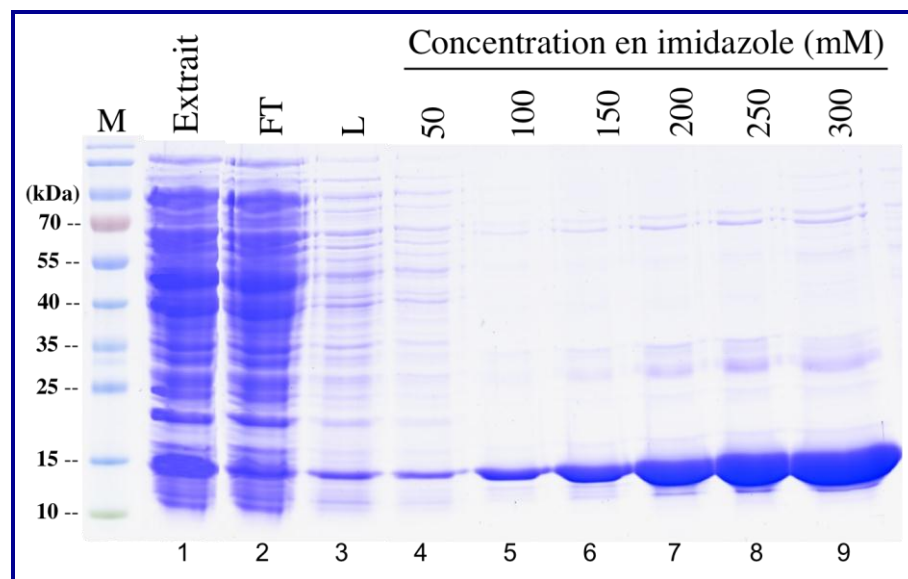


Figure 31 : Etape de chromatographie d'affinité sur Ni-NTA suivie par SDS-PAGE.

Les échantillons correspondant à des fractions éluées ont été déposés sur un gel en gradient d'acrylamide (10- 20%) en conditions dénaturantes. Le gel est coloré au bleu de coomassie. Les pistes 1, 2 et 3 correspondent à l'extrait, effluent de colonne et la fraction après lavage de résine Ni-NTA à 20 mM d'imidazole respectivement. Les pistes 4 à 9 correspondent aux concentrations croissantes en imidazole 50, 100, 150, 200, 250 et 300 mM respectivement. M : marqueur de poids moléculaires en kDa.

L'étape de gel-filtration nécessite au préalable la concentration des échantillons jusqu'à obtenir un volume final de 1-2 ml, ce qui est réalisé par ultrafiltration sur une membrane 10k (Centricon, Millipore). La concentration en protéines totales obtenue a été déterminée selon la méthode de Bradford à l'aide du réactif « Bio-Rad Protein Assay ».

La protéine concentrée (10 mg/ml) est injectée sur une colonne de gel filtration Superdex 75 16/60 (GE Healthcare), réalisée sur un système AKTA FPLC (*Fast Protein Liquid Chromatography*). Le tampon utilisé (20mM Tris-HCl pH 8, 50mM NaCl, 0,5mM β -mercaptoéthanol) dans ce système était 10 fois moins concentré en sel par rapport au tampon de lyse. L'élution du matériel a été suivie par une mesure de la DO aux longueurs d'onde de 280 nm et 215 nm en raison de faible teneur en aromatique dans de cette protéine. Un chromatogramme est présenté en figure 32. Un pic majeur est observé. D'après la courbe de calibration de la colonne, cela correspond à une protéine d'une masse moléculaire d'environ de 30 kDa.

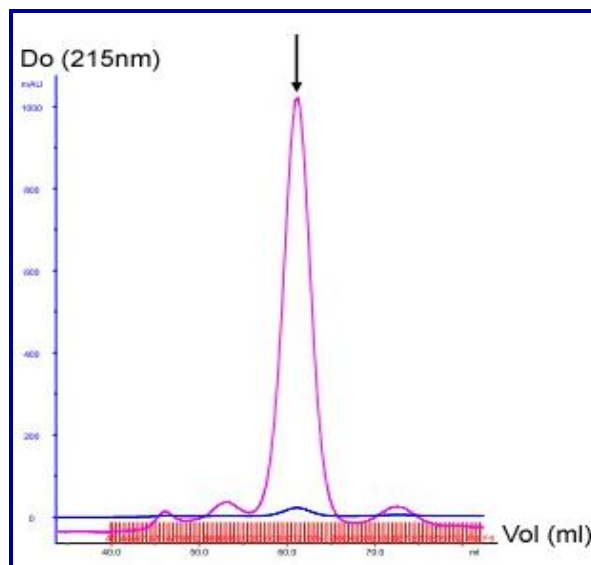


Figure 32 : Chromatogramme de gel-filtration.

Le pic majeur est montré par une flèche qui correspond à une protéine d'une masse moléculaire d'environ de 30 kDa.

L'analyse de la pureté des protéines est réalisée par électrophorèse en gel de polyacrylamide en conditions dénaturantes (SDS-PAGE 10-20%) sur toutes les fractions correspondant au pic majeur. La figure 33 montre le profil électrophorétique des 3 fractions correspondant au sommet du pic en chromatographie de gel filtration. L'électrophorèse a montré la présence d'une bande à 14,4 kDa correspondant à la protéine HsRPB11a soluble et homogène.

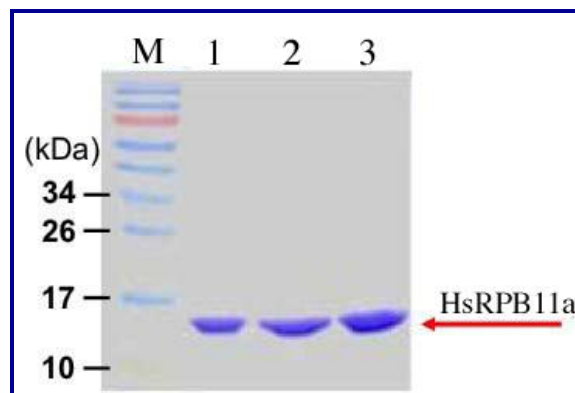


Figure 33 : L'analyse de la pureté des échantillons obtenus après gel-filtration.

La purification est suivie par électrophorèse en gels de polyacrylamide en conditions dénaturantes (SDS-PAGE, 10-20%). Le gel est coloré au bleu de Coomassie. Les pistes 1, 2 et 3 correspondent aux fractions de sommet du pic majeur (figure 28). Les bandes correspondant à la protéine HsRPB11a sont montrées par une flèche rouge. M : marqueur de poids moléculaires.

Les fractions correspondant au pic majeur sont rassemblées et la solution protéique est concentrée par ultrafiltration sur une membrane 3k (Microcon, Millipore) afin d'obtenir une concentration finale de 10 mg/ml pour les essais de cristallisation.

I.4. Criblage des conditions de cristallisation

La cristallisation est une transition de phase : la protéine passe d'un état de soluté (protéine soluble) à une phase solide (le cristal). La méthode la plus utilisée est la diffusion de vapeur en goutte suspendue ou assise, sa simplicité et son adaptation aux volumes réduits de solutions expliquent sa popularité (figure 34).

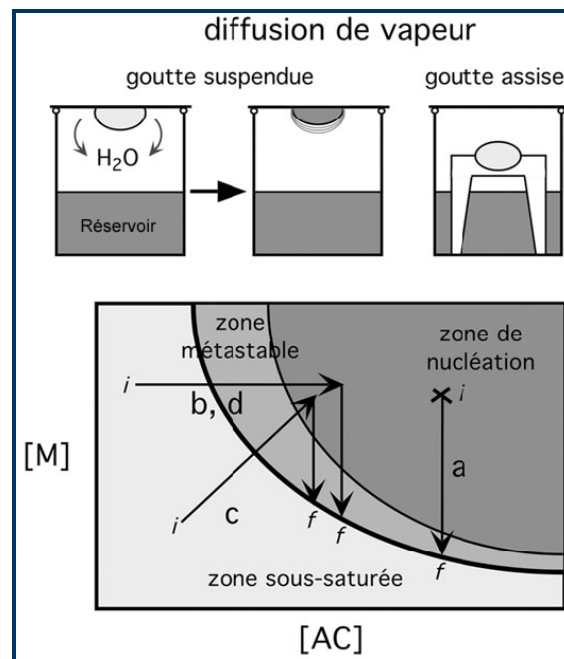


Figure 34 : Le méthode de la diffusion de vapeur en goutte suspendue et assise.

La solution de macromolécule est mise au contact (indirect) d'un réservoir contenant un agent cristallisant, dont l'augmentation de concentration va provoquer la cristallisation. Le principe d'équilibrage des solutions de cristallisation entre état initial (i) et état final (f) est indiqué dans le diagramme de phase en fonction de la méthode choisie. [M] et [AC], les concentrations de macromolécule et d'agent cristallisant.

Dans un premier temps, nous avons réalisé une recherche de conditions de cristallisation à l'aide des kits « Classics suite » et « Classics lite suite » de QIAGEN (chacune contient 96 conditions de cristallisations différentes) en utilisant un robot (Cartesian Honeybee 8+1) de la plate-forme de biologie et de génomique structurale de l'IGBMC (Illkirch). Cet appareil permet d'utiliser des microplaques de cristallisation 96 puits pour tester un grand nombre de conditions avec une faible quantité de protéine.

Les gouttes de cristallisation sont composées de 100 nl de solution de protéine à deux concentrations différentes (6 et 10 mg/ml) et 100 nl de la solution de cristallisation. Les essais ont été effectués à 25°C par la technique des gouttes

suspendues. Après 24 h d'incubation, des microcristaux ont été observés dans les conditions 46, 65 et 85a du kit « Classics suite », 83 et 85b du kit « Classics lite suite » (figure 35). Aucune différence n'a été observée entre les deux concentrations utilisées. Les compositions des solutions sont données dans le tableau 6. Les microcristaux obtenus dans toutes les conditions sont de forme hexagonale allongée.

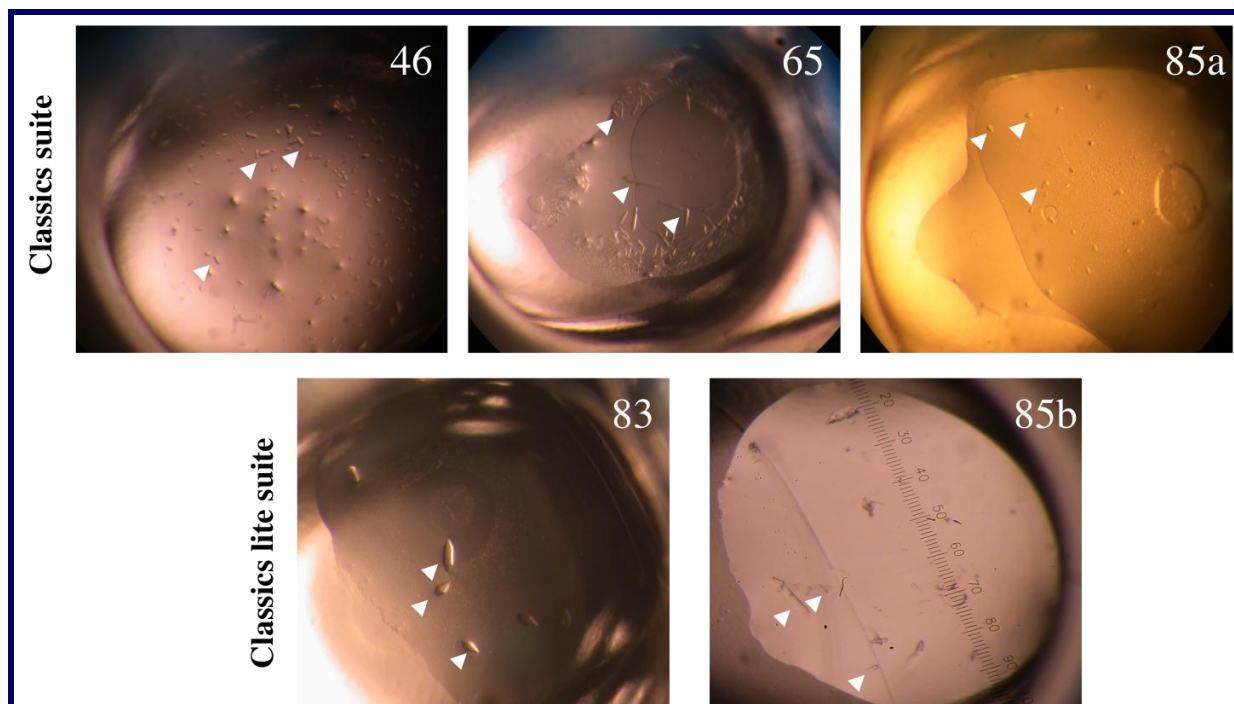


Figure 35 : Microcristaux obtenus à l'aide de kits de QIAGEN.
Les micro-cristaux sont montrés avec des flèches blanches.

Les conditions	Composition des solutions
46	100 mM HEPES pH 7,5 ; 800 mM Na ₂ HPO ₄ ; 800mM KH ₂ PO ₄ .
65	500 mM Li ₂ SO ₄ ; 15 % PEG 8000.
85a	200 mM (NH ₄) ₂ SO ₄ ; 100 mM NaAc pH 4,6; 25% PEG 4000.
83	200 mM (NH ₄) ₂ SO ₄ ; 100 mM NaAc pH 4,6; 15% PEG 2000.
85b	200 mM (NH ₄) ₂ SO ₄ ; 100 mM NaAc pH 4,6; 12,5 % PEG 4000.

Tableau 6 : Présentation des conditions obtenues lors de premier essai de cristallisation.

I.5. Optimisation des conditions de cristallisation

Les premiers cristaux obtenus ne sont pas toujours exploitables (trop petits, fragiles, agglomérats ou poly-cristallins) et il est nécessaire d'optimiser manuellement les conditions de cristallisation. Les microcristaux décrits au paragraphe précédent sont majoritairement obtenus à pH 4,6 en présence de PEG (*PolyEthylène Glycol*) qui joue le rôle d'agent cristallisant. Sur la base de ces résultats, plusieurs matrices de cristallisation ont été conçues.

Nous y avons fait varier la concentration de protéine, le volume des gouttes, la nature et la concentration en agent précipitant, la température et la concentration et la nature du sel. Nous ne décrivons pas tous les résultats obtenus, mais présentons un exemple représentatif. La figure 36 montre les cristaux obtenus en faisant varier la concentration de l'agent précipitant (PEG) pour la condition (n° 83). Les concentrations de PEG optimales sont de 13 % et 14 %.

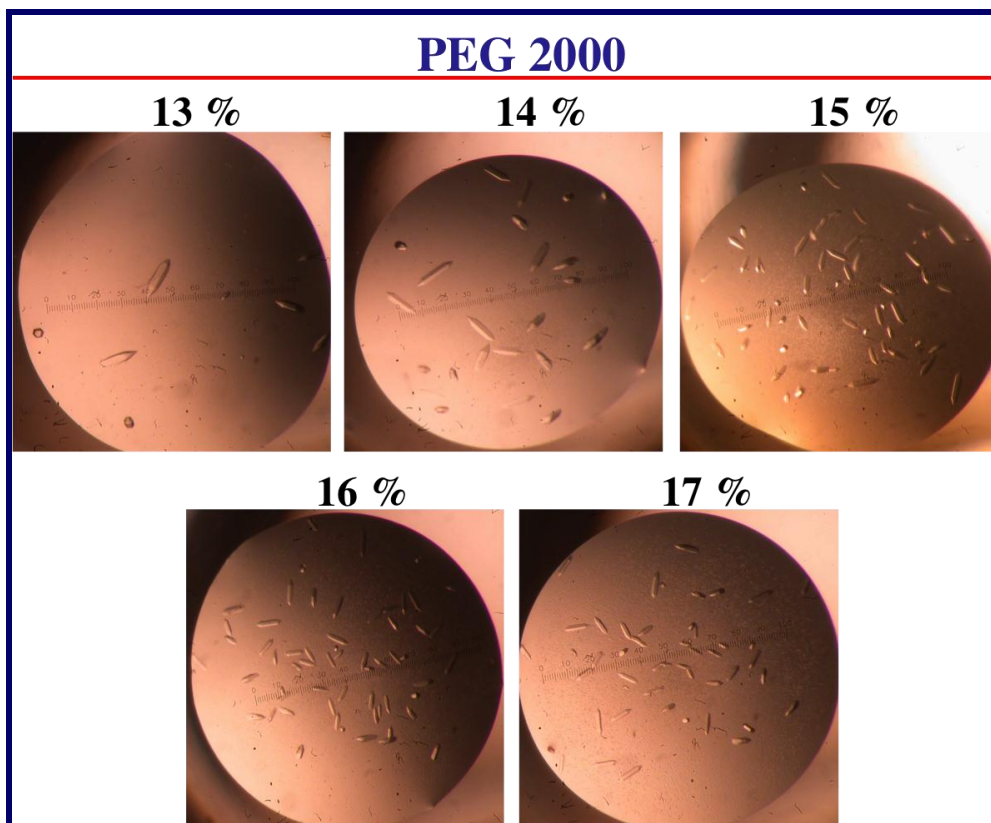


Figure 36 : Optimisation de condition de cristallisation par variation de la concentration du PEG.

La composition de la solution mère (n° 83) est décrite dans le tableau 1. Une variation de pourcentage de PEG a été effectuée sans modifier les autres paramètres. Les gouttes sont composées de 1 μ l de protéine à 6 mg/ml et 1 μ l de la solution de cristallisation. Les cristaux sont observés après 24h d'incubation à 22° C. Le volume du réservoir est 500 μ l.

Nous avons continué à optimiser cette condition en faisant varier la concentration de sulfate d'ammonium en conservant la concentration en PEG optimale de 13 %. La figure 37 montre l'effet de cette variation sur l'apparition des cristaux. Plus on augmente la concentration de sulfate d'ammonium, plus les cristaux sont petits et nombreux. La meilleure condition est difficile à identifier parce qu'à une concentration de sulfate d'ammonium 100 mM les cristaux, sont grands, mais pas complètement formés même après quelques jours, alors qu'à 200 mM ils sont plus petits et nombreux. Une concentration de 150 mM de sulfate d'ammonium a été testée et les cristaux obtenus étaient plus grands et bien formés (non montré). Nous avons considéré cette concentration comme étant l'optimum.

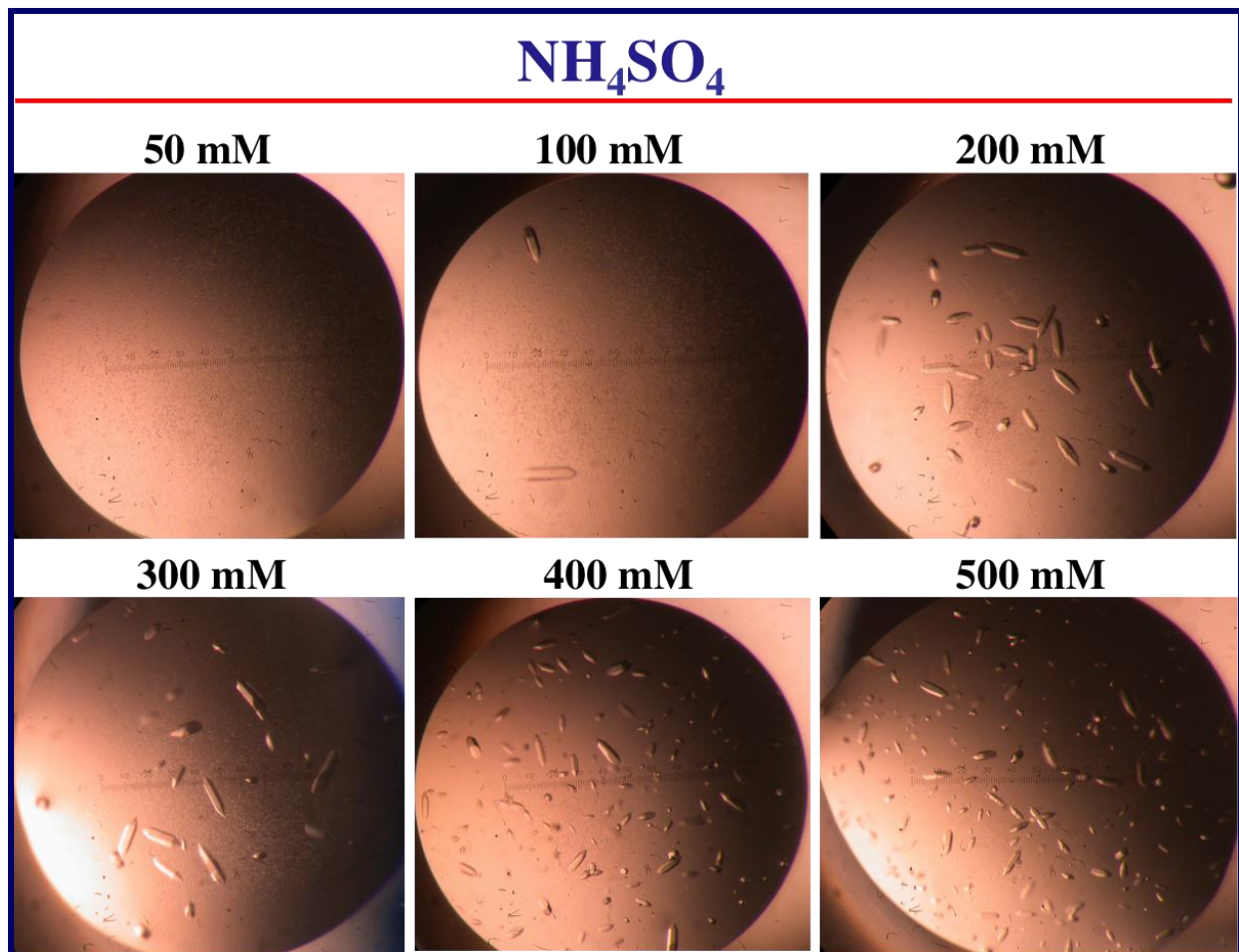


Figure 37 : Optimisation de condition de cristallisation par variation de la concentration du sel. Différentes concentrations de sulfate d'ammonium ont été utilisées. Les gouttes sont composées de 1 μ l de protéine à 6 mg/ml et 1 μ l de la solution de cristallisation. Les cristaux sont observés après 24h d'incubation à 22° C. Le volume du réservoir est 500 μ l.

L'analyse de l'effet de changement de pH dans cette condition montre que cette protéine cristallise à pH 3,6 et 4,6 (figure 38). Une gamme plus fine a été testée et le pH optimum était toujours de 4,6.

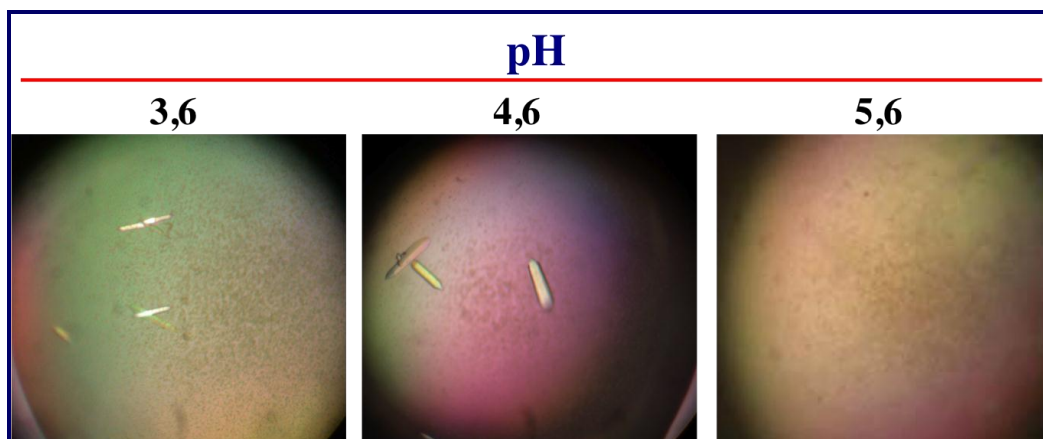


Figure 38 : Optimisation de condition de cristallisation par variation de pH.

La composition de la solution mère est la suivante : **150 mM NH_4SO_4** ; 100 mM NaAc pH [3,6, 4,6, 5,6] et 6,6 ; **13 % PEG 2000**. Les gouttes sont composées de 1 μl de protéine à 6 mg/ml et 1 μl de la solution de cristallisation. Les cristaux sont observés après 24h d'incubation à 22° C. Le volume du réservoir est 500 μl .

Les essais d'optimisation ont été effectués avec des concentrations de protéine qui varient de 5 à 16 mg/ml. Nous avons observé un rapport direct entre la taille des cristaux et la concentration en protéines (figure 39). La détermination de la concentration optimale nécessite une analyse par diffraction de rayon X.

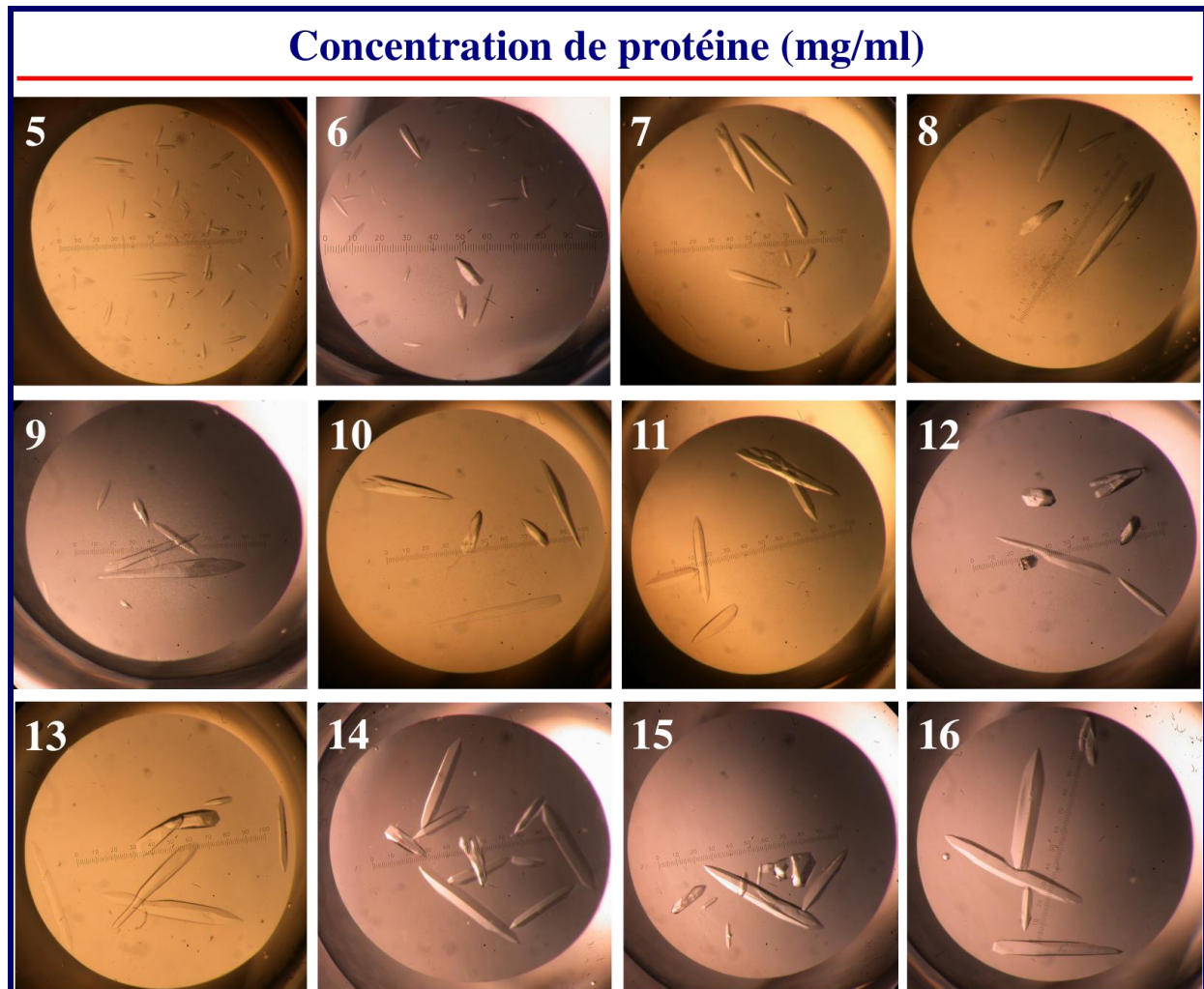


Figure 39 : Optimisation de condition de cristallisation en fonction de la concentration de protéines.

La composition de la solution mère est la suivante : **150 mM NH₄SO₄** ; 100 mM NaAc pH 4,6 ; **13 % PEG 2000**. Les gouttes sont composées de 2 μ l de protéine et 2 μ l de la solution de cristallisation. Les cristaux sont observés après 24h d'incubation à 22° C. Le volume du réservoir est 500 μ l.

Les cristaux obtenus dans différentes conditions ont été congelés directement dans l'azote liquide après une étape de cryoprotection.

I.6. Cryoprotection des cristaux et diffraction par les rayons X

Les cryoprotectants fournissent un moyen de protéger les cristaux macromoléculaires contre les effets dommageables de la formation de glace cristalline au cours du processus de congélation (Garman and Schneider, 1997). Le cristal a été pêché avec une boucle de taille adaptée et on l'a transféré dans la solution cryoprotectante. Les cryoprotectants utilisés étaient le glucose, le glycérol, l'éthylène glycol, et le MPD (*Methyl Propane Diol*) à 20 %. Les cristaux congelés dans l'azote liquide ont été testés par diffraction des rayons X à l'ESRF (*European Synchrotron Radiation Facility*, Grenoble) en collaboration avec Dr. David COBESSI de l'institut de Biologie Structurale de Grenoble.

Les données de diffraction des rayons X ont été enregistrées jusqu'à 2,9 Å de résolution. Les taches obtenues montrent une mosaïcité importante. Plus la protéine est concentrée plus la mosaïcité est grande et la divergence des faisceaux diffractés est importante. Cependant une forte mosaïcité risque d'entraîner des problèmes de superposition de réflexions, ou de causer des difficultés pour mesurer les réflexions de faibles intensités. Les données ont été traitées avec le logiciel CCP4.

La détermination des phases a été effectuée par une méthode de remplacement moléculaire (Rossmann, 1990) qui consiste à utiliser la structure de ScRPB11 comme modèle pour déterminer l'orientation et la position de la molécule recherchée. A cause de la mosaïcité observée, le phasage par la méthode de remplacement moléculaire ne nous a pas permis de valider la structure de manière définitive, mais il s'agit d'un modèle approximatif.

I.7. Construction du premier modèle

Nous disposons d'un modèle approximatif de la structure de HsRPB11a (figure 40). Les données de diffraction disponibles sont compatibles avec un dimère. Cette structure n'est pas publiable, car le modèle obtenu ne reflète pas l'ensemble des données expérimentales. Il est donc nécessaire d'utiliser une autre méthode de phasage.

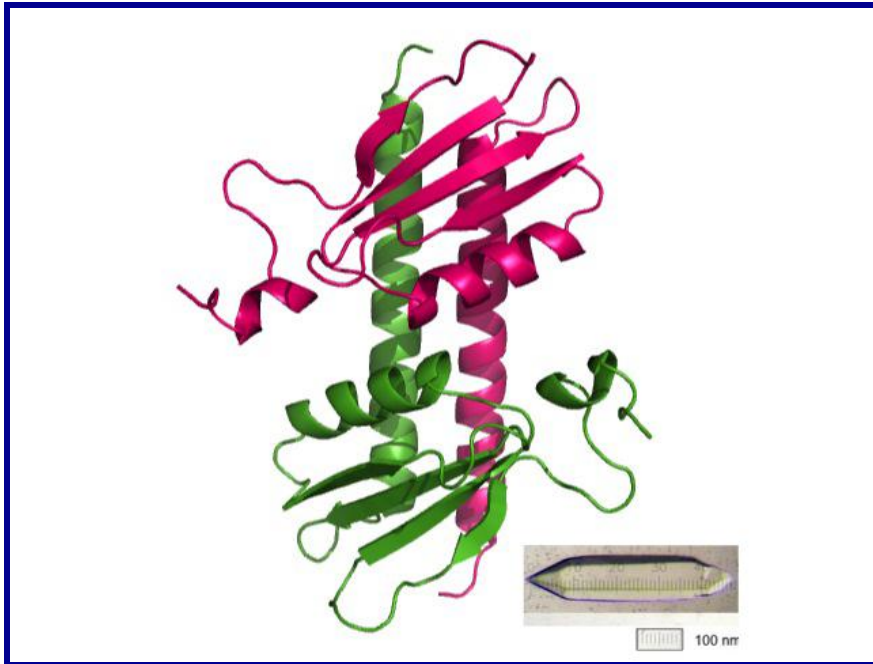


Figure 40 : Modèle approximatif de la structure de HsRPB11a.

Les données de diffraction des rayons X ont été enregistrées jusqu'à 3,4 Å de résolution à partir d'un cristal d'environ 500 microns. La composition de la solution de cristallisation est la suivante : 400 mM LiSO_4 ; 100 mM NaAc pH 4,6 ; 11 % PEG 8000. Le phasage des données a été effectué par remplacement moléculaire.

II. Résolution du problème de la phase

II.1. Remplacement isomorphe multiple

La méthode MIR (*Remplacement Isomorphe Multiple*) consiste à introduire dans le cristal de macromolécule native, par diffusion ou par co-cristallisation, des atomes lourds possédant un grand nombre d'électrons et capable de diffracter fortement les rayons X (Green et al., 1954). Si ces atomes lourds se fixent spécifiquement en des sites précis de la molécule, sans en modifier l'empilement cristallin, il est possible de déterminer la position des atomes lourds dans la maille et d'extrapoler l'information de phase. Différents atomes lourds ont été testés : le

mercure (Hg), l'euprium (Eu) et le plomb (Pb). Ces atomes lourds présentent une forte affinité pour certains résidus de la protéine, tels les cystéines et les méthionines.

Différentes concentrations comprises entre 0.05 et 1mM ont été testées par diffusion et par co-cristallisation. Nous avons réussi à obtenir des cristaux en présence des métaux lourds. Cette expérience de co-cristallisation montre l'effet de différentes concentrations des métaux lourds sur la formation et la croissance des cristaux (figure 41).

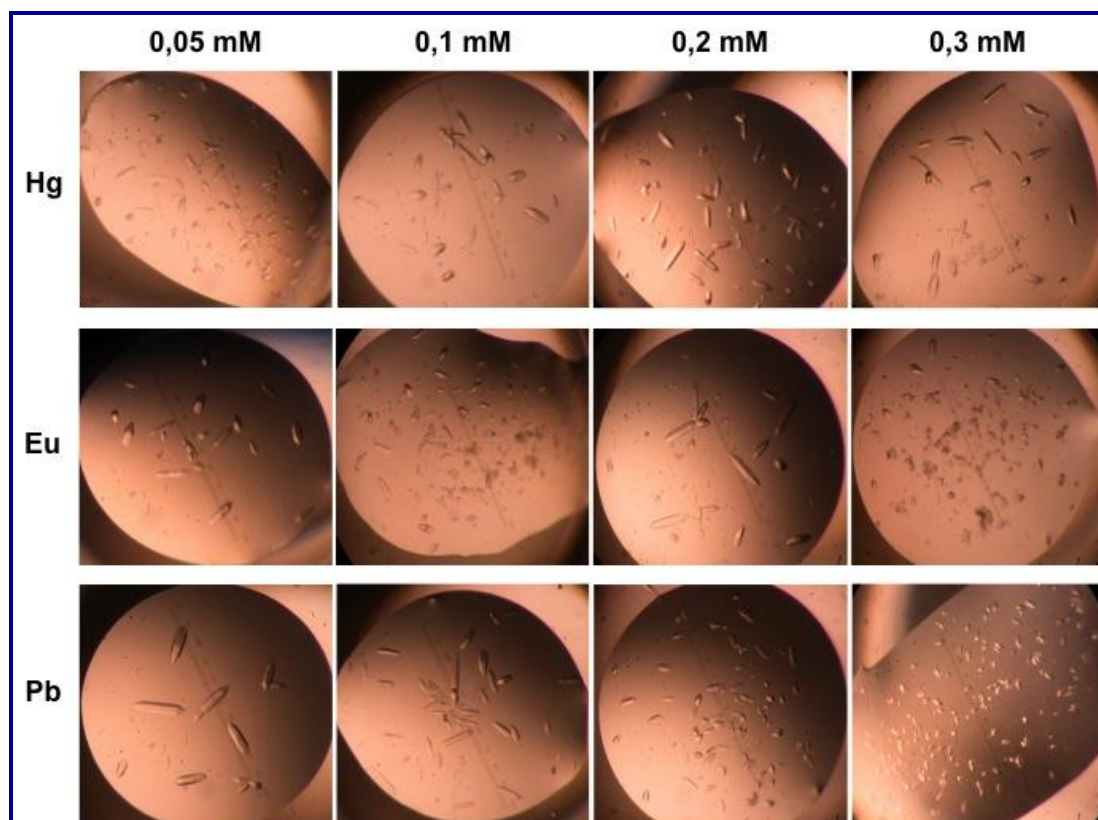


Figure 41 : l'essai de co-cristallisation des protéines avec des métaux lourds.

La composition de la solution mère est la suivante : 400 mM LiSO_4 ; 100 mM NaAc pH 4,6 ; 11 % PEG 8000. Les gouttes sont composées de 1 μl de protéine et 1 μl de la solution de cristallisation contenant 0,05 à 0,3 mM de Hg, Eu et le Pb. Les cristaux sont observés après 24h d'incubation à 22° C. Le volume du réservoir est 500 μl .

Les cristaux obtenus ont été examinés au synchrotron à l'ESRF et aucune information de phase n'a pu être extraite de la mise à l'échelle des données des cristaux natifs et des cristaux ayant subi ces traitements (il n'y avait pas de différence d'intensité entre la native et les dérivés), indiquant que la substitution n'avait pas eu lieu.

II.2. Analyse par spectrométrie de masse

La protéine purifiée a été incubée pendant 24 h avec le mercure (0,2mM) et puis soumise à une électrophorèse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE, 8%) (figure 42). Les bandes correspondant à la protéine native (A) et la protéine traitée au mercure (B) ont été découpées et analysées par spectrométrie de masse en collaboration avec la plate-forme de biologie et génomique structurale de l'IGBMC. Le résultat a confirmé que le mercure n'a pas été fixé de façon covalente sur la protéine.

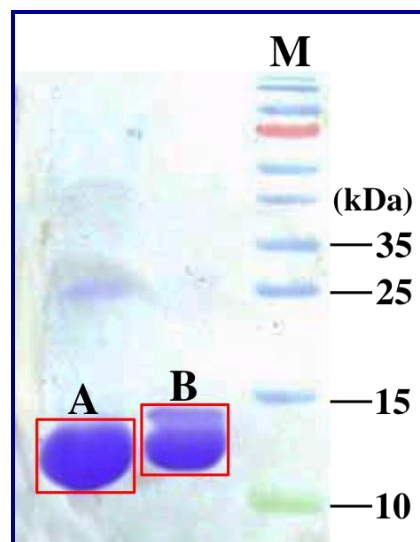


Figure 42 : Les échantillons préparés pour être analysés en spectrométrie de masse.

La bande correspondant à la protéine native (A) et la protéine traitée au mercure (B) est montrée par le rectangle rouge. Ces bandes ont été découpées et analysées par la spectrométrie de masse. M : marqueur de poids moléculaires en kDa.

III. Amélioration des conditions de cristallisation

Le but de cette nouvelle étape est d'ajuster encore plus finement les conditions afin d'améliorer encore l'aspect des cristaux, d'éliminer les problèmes de macles et surtout d'augmenter l'ordre cristallin. Nous avons essayé de réduire la vitesse de croissance des cristaux par la cristallisation à 4° C, la cristallisation en présence de 20 % de glycérol et aussi la modification du rapport volumique entre la solution de protéine et le réservoir qui permet de modifier la vitesse de concentration et

d'atteindre un niveau de sursaturation différent.

Plusieurs additifs chimiques de divers types comme les agents de solubilisation (les détergents non ioniques,...), des sels métalliques (Zn^{2+} , Mg^{2+} ,...), des composés agissant sur les potentiels d'oxydoréduction (DTT,...) ont été utilisés afin d'améliorer la qualité des cristaux.

La cristallisation a été effectuée en capillaire dans un gel d'agarose (0,2-0,3 %). Il s'agit d'une expérience de la cristallisation par contre-diffusion en microgravité. Les solutions de protéines sont mélangées avec le gel d'agarose et puis placées dans des capillaires et la cristallisation se produit grâce à une solution de cristallisation qui y diffuse progressivement. Cette méthode semble efficace pour réduire le défaut d'empilement et augmenter la qualité de diffraction des cristaux par une croissance plus régulière (Cudney et al., 1994). Les cristaux obtenus ont été examinés au synchrotron à l'ESRF. Aucune amélioration de la qualité de diffraction n'a été observée.

Beaucoup de protéines purifiées résistent à la cristallisation en raison de la nature intrinsèquement flexible de certains résidus de surface. En particulier, les résidus lysine et acide glutamique à la surface des protéines ont des chaînes latérales longues qui sont très mobiles. Ces chaînes latérales doivent être localisées dans l'espace afin de promouvoir des interactions homogènes intermoléculaires nécessaires à la formation d'un réseau cristallin. L'addition de groupements méthyle sur l'amine de la chaîne latérale de lysines a été utilisée pour favoriser la cristallisation par « immobilisation » de ces chaînes latérales (Fogg et al., 2006; Kim et al., 2008; Walter et al., 2006).

La protéine purifiée sans étiquette 6xHis a été méthylée (JBS Methylation Kit de Jena Bioscience contenant du formaldéhyde et de diméthylamine-borane complexe). La présence de groupements méthyle sur la protéine est vérifiée par spectrométrie de masse en collaboration avec Dr. VAN DORSSELAER (laboratoire de spectrométrie de masse bio-organique de Cronenbourg).

L'analyse par MALDI-TOF de la protéine non modifiée (figure 43A) fait apparaître un pic majeur d'une masse moléculaire de 13 468 Da. Dans le spectre obtenu pour la protéine méthylée (figure 43B) deux populations sont présentes avec des masses moléculaires de 13 777 et 13 804 Da. La population majoritaire (13 804 Da) présente, par rapport à la protéine non méthylée (13 468) un supplément de masse qui correspond à 24 groupements méthyle (24 fois 14 Da par méthyl). Ceci

est conforme à une réaction complète des 11 résidus lysine et de l'extrémité N-terminale. L'autre population (13 777 Da) correspond aux produits de méthylation partielle. Nous avons tenté de cristalliser les protéines méthylées, mais cela s'est soldé par un échec.

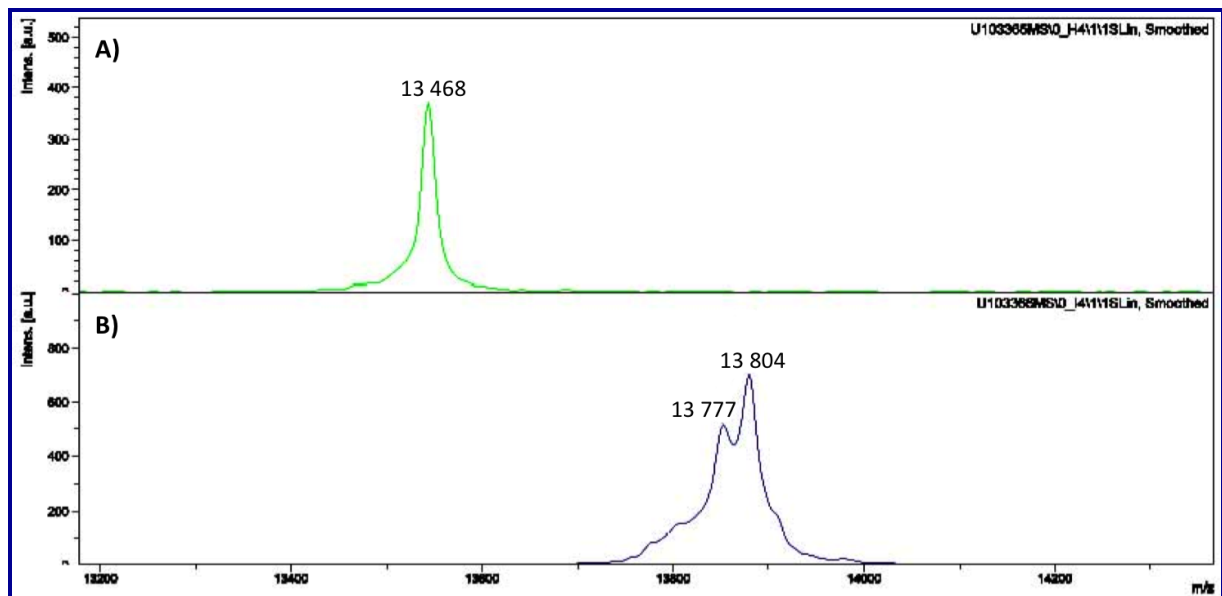


Figure 43 : Analyse de produit de méthylation de HsRPB11a par MALDI-TOF.

A) La masse moléculaire de la protéine non méthylée. B) Les masses moléculaires obtenues à partir des protéines méthylées. Les masses moléculaires sont présentées en Da.

Plusieurs autres kits de cristallisation (Additive Screen, Salt Rx, PEG Ion pH Screen de Hampton Research, The JCSG + Suite, The Nucleix suite de QIAGEN, et The Sparse Matrix 1 (SM1) de Nextal) ont été utilisés afin de trouver une autre forme cristalline. Plusieurs conditions ont été obtenues (Tableau 7). Après une étape d'optimisation, les cristaux hexagonaux de forme semblables obtenus ont été testés au synchrotron. Aucune amélioration de la qualité de diffraction n'a été observée.

Conditions	Kits	Composition des solutions
19 33 91	Salt Rx	0,7M tri-Sodium Citrate dihydrate ; 0.1 M Bis-Tris Propane pH 7.0 1.2 M DL-Malic Acid pH 7.0 ; 0.1 M Bis-Tris Propane pH 7.0 0.5 M Potassium Thiocyanate : 0.1 M Tris pH 8.5
1 2 3 6	Additive Screen	(0,2 M Ammonium sulfate ; 0,1 M Sodiome acetate pH 4,6 ; PEG 2000 13%) + 1) : 0.1 M Barium chloride dihydrate 2) : 0.1 M Cadmium chloride hydrate 3) : 0.1 M Calcium chloride dihydrate 6) : 0.1 M Magnesium chloride hexahydrate
86	SM1	86) : 2.5 M Sodium chloride ; 0.1 M Imidazole pH=8

Tableau 7 : Présentation des nouvelles conditions de cristallisation

Aucune de ces démarches n'a donné de résultat satisfaisant, car nous avons toujours observé la même forme de cristal hexagonal avec le même problème de macle et une forte mosaïcité. La persistance du problème nous a amené à tenter de modifier le polypeptide lui-même.

IV. Autres approches / nouvelles construction

IV.1. cristallisation de la protéine sans étiquette 6His

La séquence correspondant au site de clivage (ENLYFQ(G/S)) reconnu par la protéase TEV (*Tobacco Etch Virus*) (Dougherty et al., 1989) est inséré derrière de l'étiquette 6xHis. Ceci nous a permis, après le clivage par la protéase TEV (qui porte une étiquette 6xHis), d'obtenir les protéines (13,4 kDa) sans étiquette 6xHis (figure 44). Cette protéine a été purifiée par trois étapes de chromatographie. Après une étape de purification sur les résines Ni-NTA, l'échantillon a été dessalé sur une colonne (PD-10 de Amersham Biosciences) et incubé avec la protéase 6xHis+TEV à 8°C pendant 12 heures.

Une deuxième chromatographie d'affinité sur résine Ni-NTA agarose à été effectuée pour piéger la protéase TEV et éventuellement les protéines non digérées. Finalement la protéine est ensuite purifiée sur une colonne de gel filtration Superdex 75 16/60 (GE Healthcare).

Cette protéine a été testée dans toutes les conditions où nous avons observé des microcristaux avec la protéine étiquetée 6xHis. Nous avons aussi réalisé une

recherche de conditions de cristallisation à l'aide des kits « Classics suite » et « Classics lite suite » de QIAGEN. Les conditions obtenues sont pratiquement identiques à celles de la protéine étiquetée 6xHis. Les cristaux obtenus ont été testés au synchrotron et aucune différence de la qualité de diffraction n'a été observée.

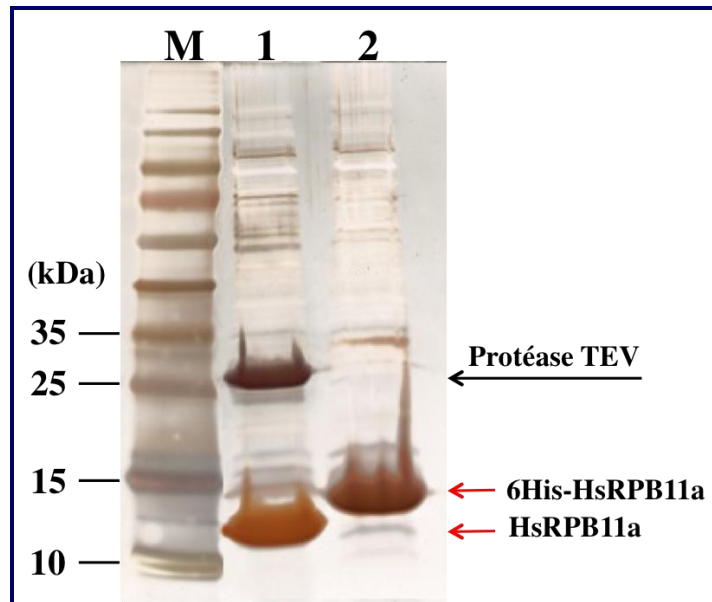


Figure 44 : Analyse de la digestion des protéines par la protéase TEV.

La protéine purifiée sur Ni-NTA est incubée une nuit avec la protéase TEV à 8°C. Les pistes (1) et (2) correspondent à des échantillons incubés avec et sans TEV protéase respectivement. Les échantillons sont déposés sur un gel d'acrylamide de gradient (10 - 20 %) en conditions dénaturantes. Le gel est coloré à l'argent. M : marqueur de poids moléculaires en kDa.

IV.2. Cristallisation de la protéine de fusion MBP-HsRPB11a

La MBP (*Maltose-Binding-Protein*) est une protéine extrêmement soluble. Cette étiquette est couramment utilisée afin d'augmenter la solubilité de la protéine à laquelle elle est fusionnée (Austin et al., 2009). Les structures de plusieurs protéines ont été résolues en les fusionnant avec la MBP (Ke and Wolberger, 2003; Kobe et al., 1999; Liu et al., 2001; Potter et al., 2008; Song et al., 2003; Ullah et al., 2008).

IV.2.1. Expression et purification de la protéine de fusion MBP-HsRPB11a

La protéine MBP a été fusionnée à la protéine HsRPB11a en position N-terminale (afin de trouver une autre forme cristalline). L'expression du gène de fusion a été induite en utilisant un milieu d'auto-induction (Studier, 2005) : un litre de culture d'*E. coli* BL21(DE3) transformées par le vecteur recombinant est incubé pendant 48h

à température ambiante (25°C). Le milieu d'auto-induction contient à la fois du glucose (280 mM) et du lactose (560 mM) comme source de carbone et d'énergie. Il permet d'obtenir une densité importante de bactéries (DO des cultures de l'ordre de 11 à 13 contre 5 à 7 dans un milieu LB induit par l'IPTG). L'induction de la production de la protéine de fusion a été vérifiée par analyse du contenu des bactéries sur gel dénaturant SDS-PAGE (figure 45 piste1) et aussi par Western-Blot (non montré).

La protéine de fusion est ensuite purifiée, à partir de l'extrait brut contenant les protéines bactériennes (20mM Tris-HCl pH 8, 500mM NaCl, 1mM DTT, inhibiteurs de protéases (complete, Roche)), par une incubation en «batch» avec les billes de résine amylose (QIAGEN) à 4°C (figure 45 pistes 2 et 3). Après un lavage avec le tampon de lyse, les protéines fixées à la résine sont éluées par une solution de maltose (10 mM). La protéine est ensuite purifiée sur une colonne de gel filtration Superdex 200 (GE Healthcare) (20mM Tris-HCl pH 8, 50mM NaCl, 1mM DTT). Les fractions correspondant au pic sont rassemblées et la solution protéique est concentrée pour les essais de cristallisation.

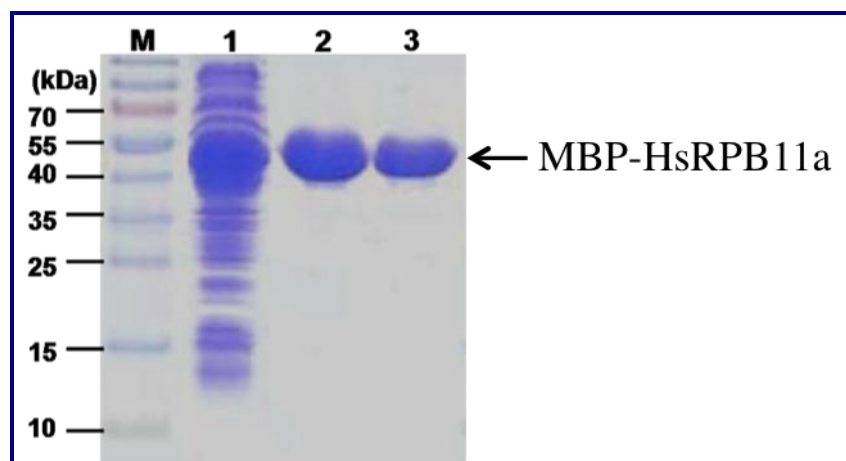


Figure 45 : Purification de protéines de fusion MBP-HsRBP11a.

Analyse de la protéine de fusion MBP-HsRBP11a exprimée dans *E.coli* BL21(DE3). Après 48 heures d'auto-induction, les cellules sont récoltées et lysées. Une partie du lysat cellulaire (piste 1) et les protéines purifiées (piste 2 et 3) ont été analysés par électrophorèse en gels de polyacrylamide en conditions dénaturantes (SDS-PAGE, 10-20%). Le gel est coloré par le bleu de coomassie. Une bande d'environ 55 kDa caractéristique de la fusion MBP-HsRBP11a est visible.

IV.2.2. Criblage des conditions de cristallisation

Les essais de cristallogénèse sur la protéine MBP-HsRPB11a ont été effectués avec le crible des 864 conditions proposées par la plateforme à haut débit de l'IGBMC. Pour cela, un premier essai a été effectué avec une protéine concentrée à 10 mg/ml. Lors de la première inspection après 48h, une séparation de phase est observée dans environ 60 % des gouttes (même après 1 mois) ce qui indique que la concentration de protéine est trop faible. Un deuxième essai a été réalisé avec une concentration de protéine de 30 mg/mL. 35 % des gouttes sont claires, 35 % des gouttes présentent une séparation de phase et 30 % des gouttes présentent une précipitation.

La recherche des conditions de cristallisation a été poursuivie avec une solution de MBP-HsRPB11a à une concentration de 70 mg/ml. Dans ces conditions, nous avons observé des gouttes claires et beaucoup de précipitation. Les gouttes sont observées pendant 2 mois et aucun cristal n'a pu se former dans toutes les conditions testées, pour cette construction.

IV.2.3. Analyse par diffusion dynamique de la lumière (DLS)

La protéine de fusion MBP-HsRPB11a a été analysée par la technique de la diffusion de la lumière (DLS : *Diffusion Light Scattering*). Cette technique permet de déterminer la taille approximative des molécules en solution et la distribution des différentes populations (agrégats, etc.).

Le résultat de DLS présenté dans la figure 46, montre un indice de polydispersité relativement élevé de l'ordre de 35 %. Cette polydispersité traduit une hétérogénéité qui explique peut-être l'absence de la cristallisation de cette protéine de fusion.

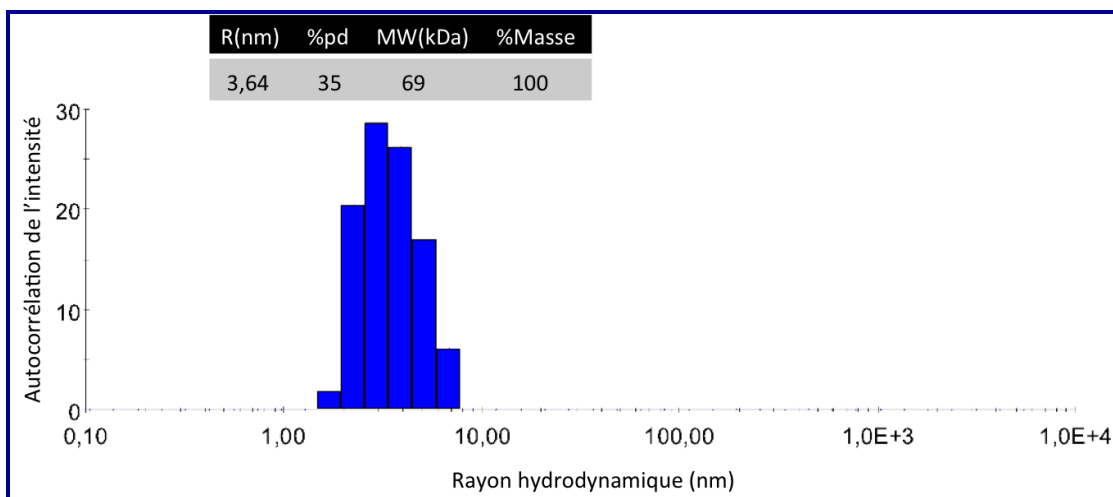


Figure 46 : Analyse de l'homogénéité structurale de la protéine par DLS.

L'analyse de fonction d'autocorrélation est reportée dans le graphique et sous forme chiffrée dans le tableau (R : rayon hydrodynamique, pd : polydispersité, MW : masse molaire approximative).

IV.3. Cristallisation de mutants de délétion de HsRPB11a

Le but de cette expérience est d'améliorer l'homogénéité du réseau cristallin ou d'obtenir de cristaux d'une nouvelle morphologie. Il s'agit de réduire la taille des protéines par ingénierie moléculaire. Le principe consiste à déléter les parties mobiles sans altérer la fonctionnalité des protéines, pour augmenter leur compaction et favoriser leur aptitude à cristalliser. Cette stratégie peut permettre d'obtenir ensuite des cristaux qui diffractent à plus haute résolution.

IV.3.1. Les différentes constructions

Le choix des résidus à déléter a été fait en analysant le modèle obtenu (figure 40) et aussi en fonction de la structure de la sous-unité RPB11 de la levure *S. cerevisiae*. Le vecteur d'expression que nous avons utilisé dans cette expérience comporte un site TEV entre l'étiquette 6xHis et la séquence codante. Les différentes constructions sont présentées dans la figure 47.

Nous avons réalisé des délétions successives de 5 acides aminés dans la partie N-terminale et 6 acides aminés dans la partie C-terminale (figure 47B). Ces acides aminés correspondent à de petits peptides qui semblent mobiles dans la structure.

Un double mutant auquel il manque 5 acides aminés du côté N-terminal et 6 acides aminés du côté C-terminal a également été construit (figure 47C).

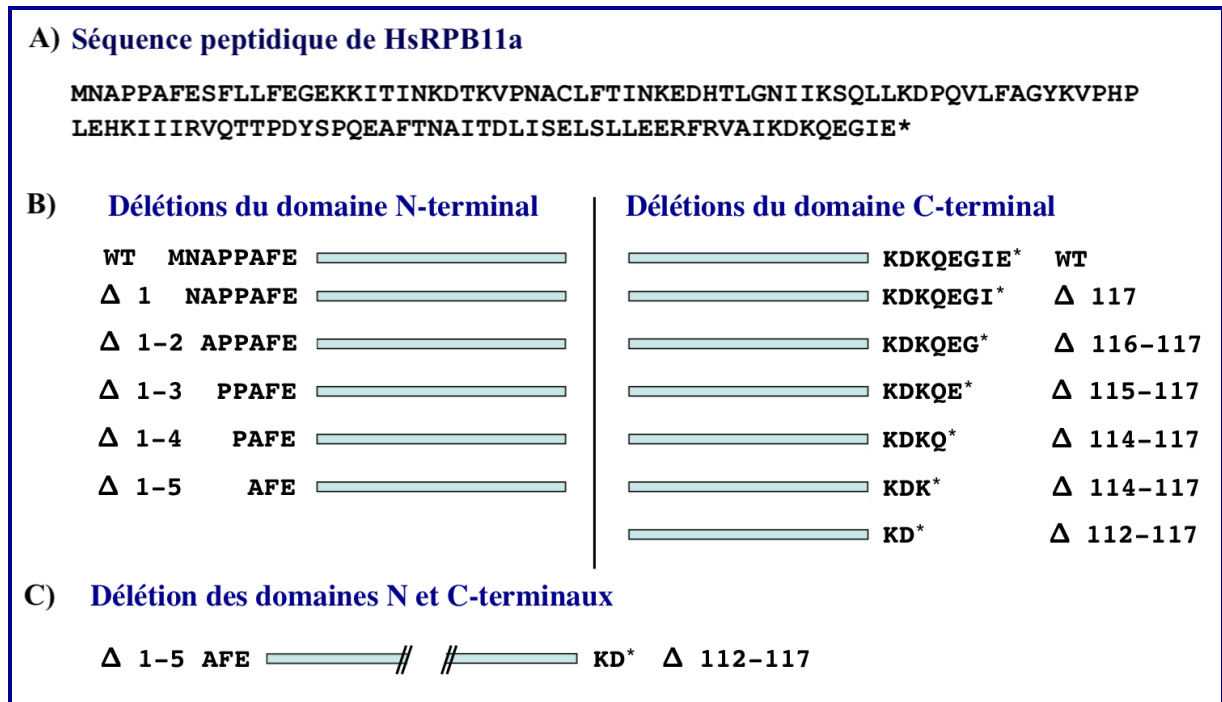


Figure 47 : Présentation des mutants de délétion de HsRPB11a

A) La séquence peptidique de la protéine HsRPB11a. B) Présentation des délétions effectuées dans les parties N et C-terminale par rapport à la protéine sauvage (WT). C) La délétion de 5 acides aminés dans la partie N-terminale et 6 acides aminés dans la partie C-terminale.

IV.3.2. Expression et purification des protéines

Les protéines mutantes ont été exprimées dans les bactéries *E.coli* BL21 (DE3). Dans un premier temps, l'induction des protéines a été réalisée dans deux litres du milieu LB à 37 °C par addition d'IPTG (0,4 mM final). Afin d'optimiser la production, différentes conditions d'expression ont été testées en jouant sur différents paramètres au cours de l'induction : concentration en IPTG, température et temps d'induction.

La solubilité des protéines est vérifiée par analyse des extraits bactériens par Western-Blot. Le surnageant du lysat cellulaire est incubé avec des billes de Ni-NTA. Après une digestion par la protéase TEV et une chromatographie d'affinité sur résine Ni-NTA agarose, la purification a été effectuée sur une colonne de gel filtration Superdex 75.

De nombreux essais de purification du double mutant de délétion ((Δ N (1-5) et Δ C (112-117)) ont échoué en raison d'une très faible production de cette protéine. Les profils chromatographiques des protéines Δ N (1-5) et Δ C (112-117) étaient différents de ceux des autres mutants (figure 48).

Outre un faible rendement pour ces mutants, on constate la présence d'une protéine de 80 kDa lors de la séparation sur une colonne Superdex 75 (figure 48 B). Le profil d'élution des autres mutants est comparable à celui obtenu pour la protéine sauvage (figure 48 A).

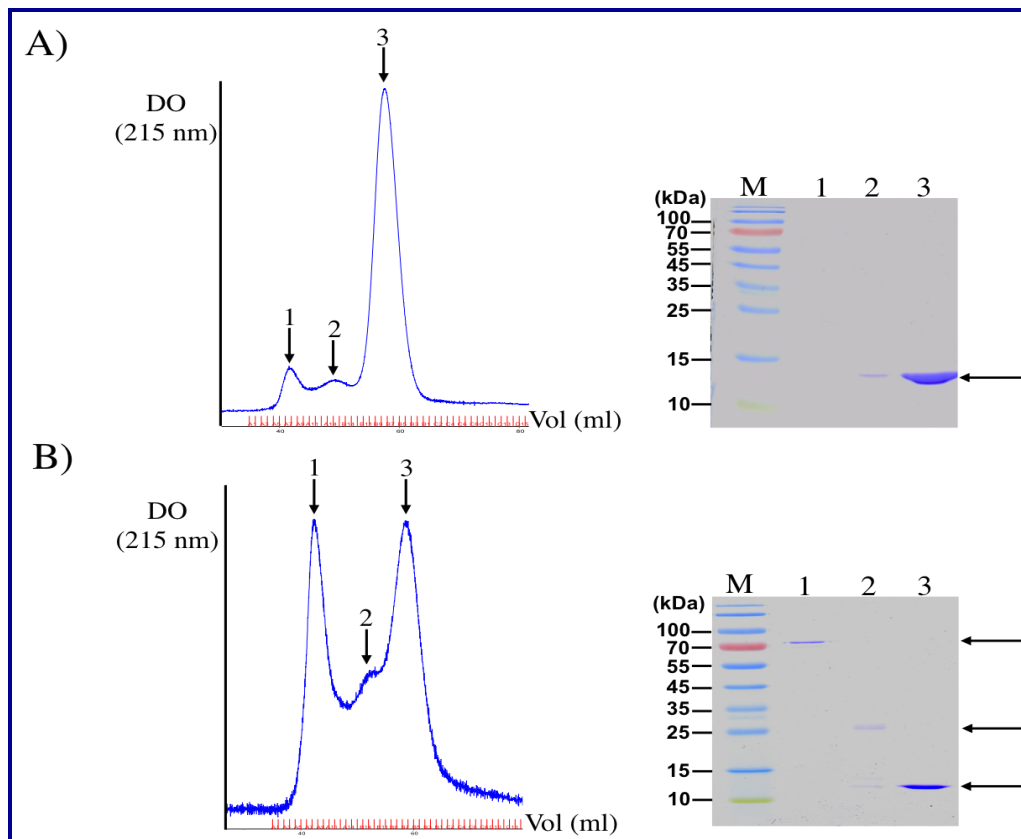


Figure 48 : Chromatogramme de gel-filtration et analyse des fractions par électrophorèse.

Courbes d'élution de la chromatographie des mutants de délétion. La séparation est effectuée par une colonne de gel filtration Superdex 75. Les fractions correspondant au sommet des pics ont été analysées sur gel en gradient de polyacrylamide (10-20 %), en présence de SDS. Le gel est coloré au bleu de Coomassie. (A) le type de chromatogramme obtenu pour tous les mutants à l'exception des mutants ΔN (1-5) et ΔC (112-117). Les pics 1, 2 et 3 correspondent à des masses moléculaires d'environ 80, 40 et 26 kDa respectivement. L'analyse de fraction des pics 1 à 3 sur gel (piste 1 à 3) montre une seule bande d'environ 13 kDa pour les pics 2 et 3 (pistes 2 et 3). Aucune bande correspond au pic 1 n'a été observée (piste 1). (B) le profil chromatographique des mutants de délétion ΔN (1-5) et ΔC (112-117) montre trois pics correspondant à des masses moléculaires d'environ 80, 40 et 26 kDa respectivement. L'analyse de fraction des pics 1 à 3 sur gel (piste 1 à 3) montre une bande d'environ 80 kDa correspond au pics 1 (piste 1) et deux bande d'environ 26 et 13 kDa correspond au pics 2 (piste 2) et une bande d'environ 13 kDa correspond au pics 3 (pistes 3). M : marqueur de poids moléculaire

Les fractions obtenues ont été réunies et la pureté de chaque échantillon a été vérifiée sur un gel de polyacrylamide en conditions dénaturantes (figure 49). La

solution protéique est concentrée à 6 mg/ml pour les essais de cristallisation.

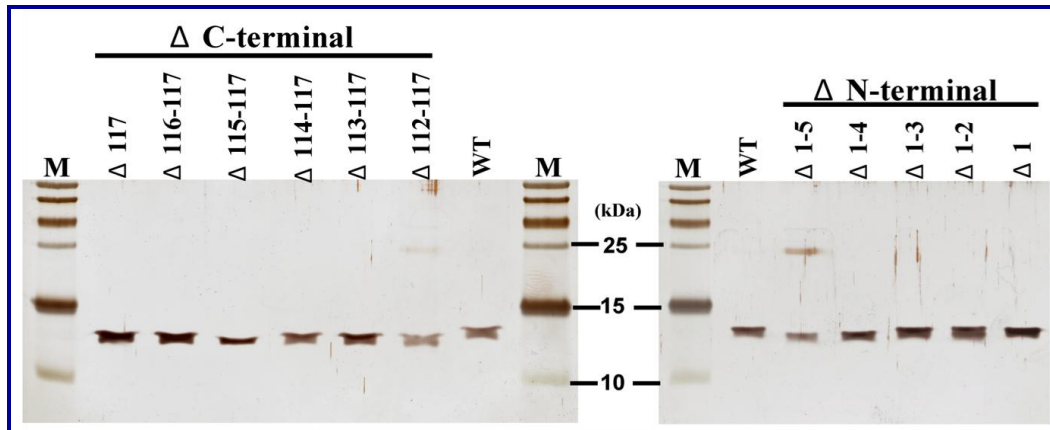


Figure 49 : Analyse des protéines purifiées en gel SDS-PAGE

La pureté des préparations a été vérifiée sur un gel de polyacrylamide en conditions dénaturantes (SDS-PAGE, 10-20 %). Le gel est coloré à l'argent. M : le marqueur de poids moléculaires en kDa.

IV.3.3. Analyse par spectrométrie de masse

Les protéines purifiées ont des tailles très proches. Elles sont difficiles à distinguer par électrophorèse en gel de polyacrylamide. L'analyse des protéines purifiées par spectrométrie de masse a été effectuée en collaboration avec le laboratoire de spectrométrie de masse bio-organique de Cronenbourg. La liste des masses théoriques est ensuite comparée à la liste des masses expérimentales relevées sur les spectres. Toutes les masses obtenues correspondent exactement à la taille attendue. Le tableau 8 représente la taille exacte obtenue de chaque protéine.

Délétions du domaine N-terminal			Délétions du domaine C-terminal		
Protéine	Masse attendue (Da)	Masse mesurée (Da)	Protéine	Masse attendue (Da)	Masse mesurée (Da)
Δ 1	13 337	13 337	Δ 117	13 339	13 339
Δ 1-2	13 223	13 226	Δ 116-117	13 226	13 226
Δ 1-3	13 152	13 151	Δ 115-117	13 169	13 168
Δ 1-4	13 054	13 054	Δ 114-117	13 040	13 039
Δ 1-5	12 957	12 957	Δ 113-117	12 911	12 911
La protéine WT (1-117)	13 468	13 468	Δ 112-117	12 783	12 782

Tableau 8 : Masses des polypeptides mesurées par MALDI-TOF.

IV.3.4. Criblage des conditions de cristallisation

Nous avons réalisé une recherche de conditions de cristallisation à l'aide du kit « Classics suite » de QIAGEN. C'est le kit qui a donné les meilleurs résultats avec la protéine sauvage. Les différents mutants ne se sont pas comportés de la même façon que la protéine sauvage. Certains nous ont permis d'obtenir rapidement des micro-cristaux (après 24 h à 22°C), d'autres ont cristallisé plus lentement (après 48 à 72 h à 22°C), enfin certains mutants n'ont pas cristallisé du tout dans ces conditions (ΔN (1-4), ΔN (1-5) et ΔC (112-117)). Le tableau 9 présente les conditions obtenues pour chaque mutant ainsi que leurs compositions.

Les mutants	Les conditions	La composition des conditions
ΔC (117)	26	26) : 100 mM Citrate de sodium pH 5,6, 1M Phosphate d'ammonium.
ΔC (116-117)	26	
ΔC (115-117)	65	27) : 100 mM Tris-HCl pH 8,5, 2M Phosphate d'ammonium.
ΔC (114-117)	27	
ΔC (113-117)	65,46,26	46) : HEPES pH 7,5, 800 mM Phosphate de sodium, 800 mM Phosphate de potassium.
ΔN (1)	26	
ΔN (1-2)	65	65) : 500 mM Sulfate de lithium, 15% PEG 8000.
ΔN (1-3)	65	

Tableau 9 : Conditions obtenues lors des premiers essais des cristallisation des mutants de délétion.

IV.3.5. Optimisation des conditions de cristallisation

Les conditions obtenues ont été optimisées de la même façon que décrit précédemment. Les cristaux sont toujours de forme hexagonale sauf dans la condition 65 avec les mutants ΔC (115-117) et ΔN (1-2), où une forme hexagonale sans pyramides a été observée (figure 50).

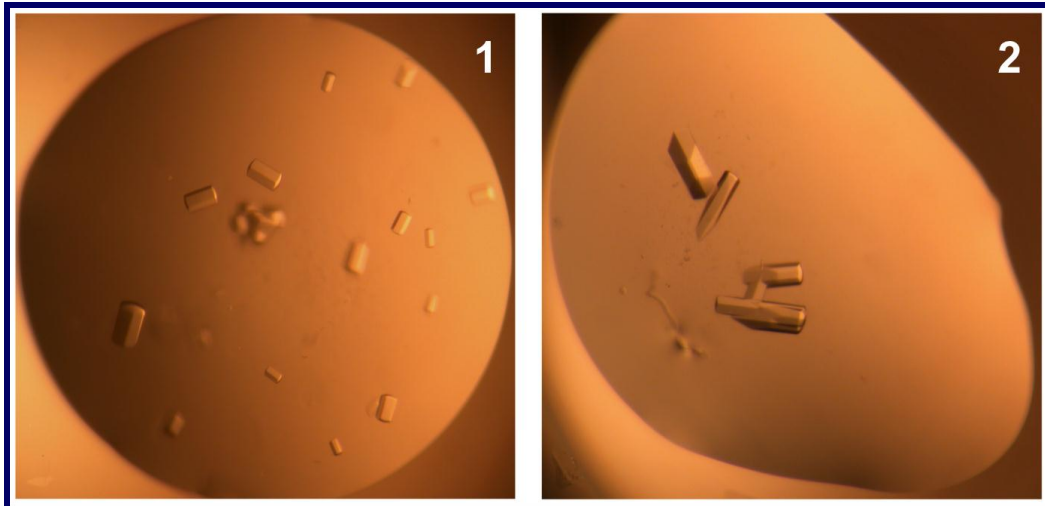


Figure 50 : Nouvelle forme cristalline obtenue à partir des mutants de délétion.

Les cristaux obtenus à partir des mutants (1) ΔC (115-117) et (2) ΔN (1-2). La condition est la suivante : (65) 500 mM Lithium sulfate , 12% PEG 8000.

IV.3.6. Recherche des conditions cryoprotectantes et diffraction par les rayons X

Chaque cristal obtenu avec les différents mutants est cryoprotégé avec la solution de cristallisation correspondante, additionnée de 20% du glycérol, glucose, éthylène glycol ou MPD avant d'être congelé dans l'azote liquide. Le test de diffraction a été effectué en collaboration avec Dr. Bruno P. KLAHOLZ de l'IGBMC. Dans ces conditions cryo-protectantes, les cristaux n'ont pas diffracté ou à très basse résolution. Nous avons observé la formation de cristaux d'eau à l'intérieur de l'échantillon. Ceci montre que les cryo-protectantes utilisées n'ont pas empêché la formation de cristaux d'eau lors de congélation. Les différents cryo-protectants ont été testé en espérant améliorer leur qualité de diffraction. Parmi ceux testés, l'huile de paratone a permis de réduire la formation des cristaux de glace. La meilleure résolution (5 à 7Å) a été obtenue avec les deux mutants (ΔN 1-2) et ΔC (116-117). Les jeux de données sur les différents cristaux montrent une mosaïcité en général élevée pour nos cristaux.

V. Discussion et perspectives

La purification de la protéine est une étape clé pour parvenir à sa cristallisation. Cette étape a fait l'objet de nombreuses optimisations. Notamment, le milieu d'auto-induction a permis d'obtenir un taux d'expression des protéines 2 à 3 fois plus élevé par rapport au milieu induit par l'IPTG. Nous avons observé au cours de la purification par la chromatographie d'exclusion, que la protéine HsRPB11a se comporte plutôt comme un dimère. En effet, le profil d'élution de HsRPB11a a révélé la présence d'un seul pic. Le calcul du poids moléculaire a montré que la protéine a une masse apparente d'environ 28 kDa, ce qui correspond au double de la masse calculée de HsRPB11a.

Nous avons purifié HsRPB11a et obtenu des cristaux. Les données de diffraction des cristaux obtenus avec la protéine étiquetée (6xHis) ont révélé une mosaïcité importante, probablement liée à un problème de maclage des cristaux, correspondant à plusieurs phases cristallines d'orientations différentes dans le cristal. Ces problèmes se sont révélés plus importants avec des cristaux obtenus avec une concentration plus élevée en protéine. En revanche, les cristaux obtenus ainsi ont mieux diffracté par rapport aux cristaux obtenus avec une faible concentration en protéine.

L'ensemble de ces jeux de données de diffraction ne nous ont cependant pas permis de résoudre la structure de HsRPB11a par la méthode de remplacement moléculaire malgré une homologie importante avec la protéine de levure. En revanche, l'analyse de ces données indique que cette protéine est sans doute sous forme de dimère dans le cristal.

Il est néanmoins important de noter que le processus de cristallisation de cette protéine était très rapide (5h à 22°C), ce qui pourrait expliquer un empilement incorrect des molécules dans le cristal, favorisant ainsi un maclage. Nous avons réussi à réduire la vitesse de nucléation et de croissance des cristaux (3 à 5 fois) par la cristallisation à 4°C, la cristallisation en présence de 20% de glycérol et la modification du rapport volumique entre la solution de protéine et le réservoir, mais aucune amélioration des données de diffraction n'a été observée. D'autres pistes telles la cristallisation dans l'huile et l'utilisation des différents additifs pourraient permettre de diminuer encore la vitesse de croissance des cristaux et seront testées.

Nous avons obtenu des cristaux en co-cristallisation avec des métaux lourds, mais les tests de diffraction ne montrent aucun signal correspondant aux métaux testés. L'analyse de spectrométrie de masse a confirmé l'absence de mercure fixé sur la protéine. Il est probable que les positions des acides aminés dans la structure de la protéine n'étaient pas favorables pour la fixation des métaux utilisés.

Une autre méthode de phasage est le MAD (*Multiwavelength Anomalous Diffraction*). Elle est actuellement l'une des méthodes les plus utilisées pour déterminer la phase des facteurs de structure des macromolécules biologiques. En effet, l'un des avantages de la méthode MAD est qu'il est possible d'obtenir plusieurs jeux de données nécessaires à la résolution de la structure, à partir d'un seul cristal. La méthode consiste à exprimer la protéine d'intérêt en incorporant dans le milieu de culture de la sélénométhionine à la place de la méthionine, l'atome de soufre de cet acide aminé étant remplacé par un atome de sélénium. Le problème rencontré dans le cas de HsRPB11a été la présence d'une seule méthionine dans la séquence peptidique (le codon d'initiation), ce qui limite fortement l'utilisation du MAD. Dans cette expérience nous avons besoin au minimum une méthionine par 7 kDa. Donc il faudrait remplacer quelques acides aminés par des résidus méthionine sans toucher aux résidus impliqués dans l'homodimérisation.

Les propriétés des cristaux de protéines peuvent souvent être améliorées par *déshydratation* des cristaux. La déshydratation peut avoir plusieurs effets observables sur les propriétés des cristaux parmi lesquels un changement de groupe d'espace, la réduction de la maille cristalline, des changements de mosaïcité, une amélioration des profils des taches de diffraction, une amélioration de la résolution ou du pouvoir de diffraction de l'échantillon.

Une autre méthode que nous avons employé pour améliorer la qualité des cristaux est la méthylation des protéines. La méthylation hétérogène de HsRPB11a a peut-être été un obstacle pour la cristallisation. Il faudra optimiser les conditions de méthylation en jouant sur le temps d'incubation avec le réactif et la concentration de protéines.

Nous avons aussi tenté de cliver le peptide polyhistidine, car de façon générale sa présence s'avère gênante dans la cristallisation, mais dans le cas de HsRPB11a aucune différence n'a été observée par rapport à la protéine non clivée.

Un essai de cristallisation de la protéine de fusion MBP-HsRPB11a a été effectué, car la MBP est une protéine très soluble qui a la capacité d'améliorer la

solubilité et le repliement correct de ses partenaires de fusion. La fusion de HsRPB11a avec la MBP a généré une protéine de fusion avec des comportements complètement différents de HsRPB11a au point de vue cristallographique. Plusieurs conditions de cristallisation (à différentes concentrations) ont été utilisées. Le test préliminaire de précipitation par le PEG donne un indice sur la concentration nécessaire de la protéine à cristalliser. Ce test a été positif avec la protéine concentrée à 10 mg/ml mais la recherche des conditions de cristallisation a montré que la protéine n'était pas assez concentrée. Malgré de la pureté de cette protéine de fusion, aucune condition de cristallisation n'a été trouvée. L'analyse de DLS a été effectuée avec une concentration de 70 mg/ml. Le résultat a montré une polydispersité assez élevée. Cette polydispersité observée peut être due à la concentration élevée des protéines ou à un repliement hétérogène.

Nous avons tenté de co-exprimer les sous-unités RPB3 et RPB11a et de les cristalliser. Nous avons rencontré d'importants problèmes de solubilité.

L'analyse des propriétés de mutants de délétion affectant les extrémités N- et C-terminales de la protéine HsRPB11a a montré que la délétion d'un seul acide aminé peut entraîner un changement important de ses capacités de cristallisation. Nous n'avons d'ailleurs pas obtenu de cristaux pour tous les mutants construits.

L'expression de certains mutants, tels que le double mutant de délétion (figure 47) n'était détectable qu'en Western-Blot. Malgré leur faible expression, la production et la purification des mutants ΔN (1-5) et ΔC (112-117) a pu être compensée en augmentant le volume de culture.

Les cristaux obtenus à partir des mutants ont montré le même problème de mosaïcité. Les clichés de diffraction obtenus ont été plus ou moins contaminés par la formation des cristaux de glace au cours de la congélation. Néanmoins, l'optimisation de la composition du cryo-protectant n'a pas permis d'empêcher la formation de la glace. Cependant, la recherche des cryo-protectants est en cours.

Les cristaux obtenus à partir des deux mutants ΔN (1-2) et ΔC (116-117) ont diffracté le mieux (5 à 7Å), si on les compare aux autres mutants (8 à 12Å). Des tests de cristallisation dans l'huile sont en cours afin d'améliorer la qualité de diffraction des cristaux.

Malgré toutes ces optimisations, malgré le changement des conditions de cristallisation, ou l'utilisation de différentes constructions, nous n'avons pas encore réussi à résoudre la structure de la protéine HsRPB11a.

Chapitre 2 : Analyse des interactions de HsRPB11a

I. Analyse de l'homodimérisation de la sous-unité HsRPB11a

L'étape de purification de la protéine HsRPB11a par chromatographie d'exclusion a permis de constater que cette protéine se comportait comme un dimère. Afin de confirmer cette observation, nous avons testé sa capacité de dimérisation en utilisant la technique du double-hybride chez *S. cerevisiae*. Dans ces expériences, nous avons comparé en parallèle l'homodimérisation des protéines HsRPB11b α et ScRPB11.

Nous avons également utilisé la technique du double-hybride pour comparer la formation de l'homodimère HsRPB11a avec celle de l'hétérodimère HsRPB3-HsRPB11a déjà caractérisé. Différents mutants ponctuels ou de délétion de HsRPB11a ont été utilisés afin d'identifier les résidus qui sont importants dans cette dimérisation.

La capacité d'homodimérisation de HsRPB11a dans les cellules humaines a également été testée par des expériences de FRET et de co-immunoprécipitation.

I.1. Analyse de l'homodimérisation par double-hybride chez *S. cerevisiae*

La capacité d'interaction des différentes sous-unités a été testée *in vivo* en utilisant le test du double-hybride dans les cellules de levure (figure 51). Pour cela, nous avons fusionné au domaine de liaison à l'ADN issu du répresseur LexA, la protéine qualifiée d'« appât ». D'autre part le partenaire d'interaction potentiel est fusionné au domaine activateur de VP16. L'interaction entre des fusions va reconstituer un facteur de transcription fonctionnel qui va induire l'expression du gène rapporteur β -galactosidase intégré dans le génome de la levure L40 (Mat α , his3 Δ 200, trp1-901, leu2-3,112, ade2, LYS2::(lexAop)₄-HIS3, URA3::(lexAop)₈-LacZ, GAL4, GAL80).

Les vecteurs pLexA-X et pVP16-Y ont été transformés en deux étapes dans des levures L40 par électroporation. Les transformants ont été sélectionnés en utilisant des milieux sélectifs appropriés. Nous les avons sélectionnés les auxotrophes pour le tryptophane (pLexA porte le gène TRP1) et la leucine (pVP16 porte le gène LEU2). L'activation du gène rapporteur β -galactosidase a été révélée qualitativement par conversion du substrat X-gal en un produit de couleur bleue.

L'expression de la protéine de fusion LexA-HsRPB11a seule induit une faible activité β -galactosidase (β -gal) (figure 51, N° 1). En présence de la fusion VP16-

HsRPB11a, l'activité β -gal est fortement augmentée (figure 51, N° 7), ainsi qu'en présence de la fusion VP16-Sc11 (figure 51, N° 19). En revanche, en présence de la fusion VP16-HsRPB11b α , seule une légère augmentation de l'activité β -gal peut être observée (figure 51, N° 13). Quand une forme mutée de la LexA-HsRPB11a dans le motif α est exprimée (E38A, D39A, L42A), pratiquement aucune activité β -gal n'est détectable (figure 51, N° 2, 8, 14, 20). La protéine HsRPB11a est capable aussi bien d'interagir avec elle-même et qu'avec la protéine de levure. Les mutations dans le motif α de HsRPB11a affectent fortement ces interactions.

L'expression de la protéine de fusion LexA-HsRPB11b α seule induit une activité β -gal légèrement plus élevée (figure 51, N° 3) que celle de la fusion Lex-HsRPB11a (figure 51, N° 1). En présence de la fusion VP16-HsRPB11b α , une forte augmentation de l'activité β -gal est visible (figure 51, N° 15), alors qu'elle est plutôt marginale en présence de la fusion VP16-HsRPB11a ou VP16-Sc11 (figure 51, N° 9 et 21). A nouveau, quand une fusion LexA-HsRPB11b α mutée dans le motif α est exprimée (E37A, D38A, L41A), aucune activité β -gal n'est détectable (figure 51, N° 4, 10, 16, 22). La protéine HsRPB11b α a donc la capacité d'interagir avec elle-même et légèrement avec HsRPB11a et ScRPB11. Les mutations dans le motif α de HsRPB11b α inhibent ces interactions.

L'expression de la fusion LexA-Sc11 n'induit aucune activité β -gal (figure 51, N° 5). En présence d'une fusion VP16-HsRPB11a, l'activité β -gal est fortement augmentée (figure 51, N° 11). En présence de la fusion VP16-HsRPB11b α , seule une augmentation très marginale de l'activité β -gal peut être détectée (figure 48 N° 17). En revanche, en présence de la fusion VP16-Sc11 (figure 51, N° 23) une très légère augmentation de l'activité β -gal est présente. Quand une forme mutée dans le motif α de la protéine de fusion LexA-Sc11 est exprimée (E38A ; D39A, L42A) au lieu de la fusion de type sauvage, l'activité β -gal est très similaire (figure 51, comparer N° 5 et 6, 11 et 12, 17 et 18). Seule la faible activité induite par le type sauvage VP16-Sc11 est affectée par cette mutation (figure 51, comparer N° 23 et 24). La protéine ScRPB11 présente une très faible capacité à former un homodimère, mais elle peut fortement interagir avec la protéine humaine HsRPB11a.

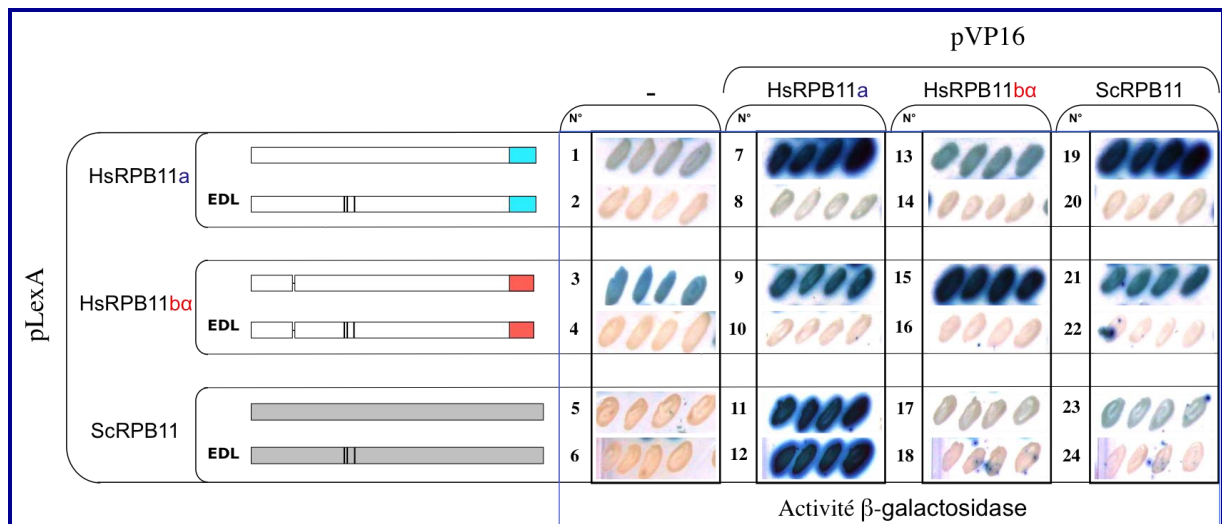


Figure 51: Étude de l'interaction protéine-protéine dans la levure par double-hybride

La souche de levure L40 a été transformée avec des plasmides exprimant le domaine de liaison à l'ADN de la protéine LexA (pLexA) fusionnée soit à une protéine sauvage ou à une protéine mutée comme indiqué. Chaque protéine est schématiquement représentée comme une barre. Dans le cas des isoformes humaines, les domaines présentés en couleur bleu et rouge correspondent à l'exon 4 qui code des peptides distincts (Benga et al. 2005). Les résidus mutés dans chaque séquence sont représentés par des barres verticales. Les levures sont également transformées avec un plasmide (pVP16) exprimant le domaine d'activation de VP16 fusionné ou non avec les mêmes protéines sauvages, comme indiqué en haut de la figure. Dans chaque cas, quatre colonies indépendantes ont été testées. Les levures ont été étalées côte à côte sur un milieu minimum YNB (*Yeast Nitrogen Base*) complété en adénine et histidine. La présence de l'activité bêta-galactosidase dans chaque souche de levure est ensuite révélée par un test de conversion de X-gal. Chaque combinaison de plasmides est référencée par un numéro (N°) qui est indiqué à gauche de des colonies de levures.

Les protéines humaines (HsRPB11) et la protéine de levure (ScRPB11) présentent des propriétés intrinsèques différentes. Les deux isoformes humaines HsRPB11a et HsRPB11bα ont montré une capacité d'homodimérisation similaire, mais elles interagissent très peu l'une avec l'autre. En revanche, la protéine ScRPB11 présente une très faible capacité d'homodimérisation.

Les trois résidus conservés dans le motif α s'avèrent indispensables pour l'homodimérisation dans le cas des protéines humaines (HsRPB11a). En revanche, la mutation des mêmes résidus a un effet marginal sur l'interaction de la protéine ScRPB11.

I.2. Analyse comparative de l'homodimérisation et de l'hétérodimérisation de HsRPB11a avec HsRPB3

Nous avons décidé de nous focaliser sur la sous-unité HsRPB11a et d'examiner l'effet de plusieurs mutations qui affectent cette sous-unité lorsqu'elle est

fusionnée au domaine de liaison à l'ADN de LexA sur les interactions HsRPB11a/HsRPB11a et HsRPB11a/HsRPB3 (figure 52). Nous avons comparé deux séries de fusions VP16 qui ne diffèrent que par la présence ou l'absence d'un peptide de liaison entre le domaine d'activation de VP16 et la sous-unité testée, et sont codées par pVP16₍₁₎ et pVP16₍₂₎, respectivement.

L'expression des protéines de fusion LexA-HsRPB11a seule, soit de type sauvage, soit mutées, induisent une activité β -gal faible (figure 52, N° 1-6). Encore une fois, en présence d'une fusion VP16-HsRPB11a, l'activité β -gal est fortement augmentée (figure 52 N° 7 et 19) ainsi qu'en présence de la fusion VP16-HsRPB3 (figure 52, N° 13 et 25).

L'expression d'un double mutant de LexA-HsRPB11a dans le motif α (E38A et D39A) n'induit aucune activité β -gal (figure 52, comparer N° 2 et 1). En présence d'une fusion VP16-HsRPB11a, l'activité β -gal est très proche de celle observée en présence de la protéine sauvage LexA-HsRPB11a (figure 52, comparer N° 7 et 8, 19 et 20). Au contraire, en présence de la fusion VP16-HsRPB3, une diminution de l'activité β -gal est observée lorsqu'on compare le type sauvage avec le mutant de LexA-HsRPB11a (figure 52 comparer N° 13 et 14, 25 et 26). La double mutation (E38A et D39A) dans le motif α n'a d'effet que sur l'hétérodimérisation.

L'expression du triple mutant (E38A, D39A et L42A) de LexA-HsRPB11a supprime le bruit de fond (figure 52, N° 3). En présence d'une fusion VP16-HsRPB11a, l'activité β -gal reste très faible par rapport à la protéine sauvage de LexA-HsRPB11a (figure 52, comparer N° 7 et 9, 19 et 21). Ceci est aussi vrai en présence de la fusion VP16-HsRPB3 (figure 52, comparer N° 13 et 15, 25 et 27). Une triple mutation dans le motif α affecte donc fortement à la fois l'homo- et l'hétérodimérisation.

L'expression d'un mutant de délétion de HsRPB11a dépourvu des résidus 106-117 ne modifie pas significativement le bruit de fond (figure 52, N° 4). En présence d'une fusion VP16-HsRPB11a, l'activité β -gal est très proche de celle observée avec la protéine sauvage de LexA-HsRPB11a (figure 52, comparer N° 7 et 10, 19 et 22). De même, en présence d'une fusion VP16-HsRPB3, l'activité β -gal est très proche de celle observée en présence de la protéine sauvage LexA-HsRPB11a (figure 52, comparer N° 13 et 16, 25 et 28). Cette délétion n'a aucun effet sur ces interactions.

La formation de l'hétérodimère HsRPB3/HsRPB11a apparaît plus efficace que la formation de l'homodimère HsRPB11a.

Le remplacement de deux résidus (E38A et D39A) dans le motif α par des alanines n'a pas affecté significativement l'homodimérisation de HsRPB11a, alors que la formation de l'hétérodimère HsRPB11a/HsRPB3 est fortement affectée.

Au contraire, le remplacement de trois résidus (E38A, D39A et L42A) dans le motif α abolit à la fois homo- et hétérodimérisation.

La partie C-terminale (79-117) de HsRPB11a joue un rôle important pour l'homo- comme pour l'hétérodimérisation tandis que la partie (1-79) est plus particulièrement critique pour l'homodimérisation.

I.3. Analyse de l'homodimérisation par FRET dans les cellules humaines

Nous avons démontré l'homodimérisation de HsRPB11a dans la levure en utilisant la technique du double-hybride. Nous avons voulu examiner cette homodimérisation dans les cellules humaines par la méthode du FRET (*Förster / Fluorescence Resonance Energy Transfer*) (Forster, 1948). La technique de transfert d'énergie, FRET a été beaucoup utilisée pour démontrer l'existence d'interactions protéines-protéines. La molécule « donneur » porte un fluorochrome dont la longueur d'onde d'émission correspond à la longueur d'onde d'excitation du fluorochrome placé sur la molécule « accepteur », l'interaction entre les deux molécules permettant le rapprochement des deux fluorochromes. Si les paramètres de distance et d'orientation (géométrie) entre les deux fluorochromes sont corrects, un transfert d'énergie pourra se produire et conduire généralement en une émission de fluorescence de la molécule accepteur.

Dans cette expérience, la protéine HsRPB11a est exprimée sous forme de protéine de fusion avec un tag protéique (SNAP-tag de BioLabs) soit en position N-terminale, soit en position C-terminale. Le SNAP-tag est un mutant de l'O⁶-alkylguanine-DNA-alkyltransférase humaine (h-AGT), une protéine de 20 kDa de la machinerie de réparation de l'ADN. Ce tag protéique lie de façon covalente un fluorophore de synthèse de faible taille. Les fluorophores utilisés sont les suivants : la TMR (*TetraMethylRhodamine*) Star (jaune-orange) avec une longueur d'onde d'excitation maximale à 554 nm et une longueur d'onde d'émission maximale à

580 nm, le SNAP-Cell 505 (vert) avec une longueur d'onde d'excitation maximale à 504 nm et une longueur d'onde d'émission maximale à 532 nm. L'expression de la protéine de fusion SNAP-HsRPB11a ou HsRPB11a-SNAP a été vérifiée dans les cellules HeLa par Western-Blot en utilisant d'un anticorps spécifique dirigé contre la sous-unité HsRPB11a (figure 53). Une bande d'environ 35 kDa caractéristique de la fusion SNAP-HsRPB11a ou HsRPB11a-SNAP a été identifiée.

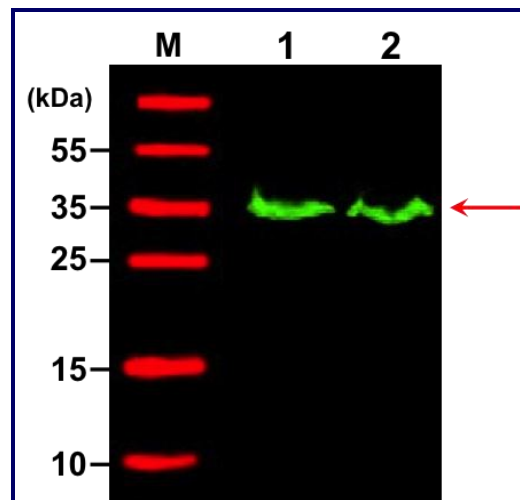


Figure 53 : Analyse de l'expression de la protéine de fusion par western blot.

Les cellules HeLa ont été transfectées par le plasmide pSNAP-HsRPB11a à l'aide d'un agent de transfection (TransIT de Mirus). 24h après la transfection, les cellules sont lysées et les protéines solubilisées dans un tampon de lyse « RIPA ». Les protéines sont déposées sur gel en gradient de polyacrylamide (SDS-PAGE 10-20 %). Les protéines sont électrotransférées sur une membrane de nitrocellulose. La membrane est incubée avec l'anticorps primaire dirigé contre la sous-unité HsRPB11a. La membrane est ensuite révélée après incubation en présence de l'anticorps secondaire couplé à une molécule fluorescente par le système d'imagerie Odyssey®. Les bandes correspondant à la protéine de fusion dans la partie N-terminale (1) et C-terminale (2) sont montrées avec une flèche rouge. M : marqueur de poids moléculaires en kDa.

Le test de marquage des protéines de fusion par des fluorophores (TMR et SNAP-Cell 505) (10 μ M) a été effectué dans les cellules HeLa selon le protocole conseillé par le fabricant (SNAP-Cell Starter Kit de BioLabs). Un marquage satisfaisant a été obtenu avec le fluorophore TMR. En revanche un rendement de fluorescence très faible a été observé avec le fluorophore SNAP-Cell 505. Plusieurs essais d'optimisation de marquage ont été effectués en augmentant la concentration du fluorophore et aussi le temps d'incubation. Aucune amélioration de marquage n'a été observée.

Nous avons testé un autre fluorophore vert (SNAP-Cell Oregon Green) et la fluorescence obtenue a toujours été faible. Ces fluorophores verts ont été testés dans une lignée cellulaire HEK-293 (*Human Embryonic Kidney*). Nous avons observé un signal de fluorescence plus élevé par rapport aux cellules HeLa mais cela s'est avéré insuffisant pour l'étude de FRET.

Nous avons finalement choisi d'utiliser de la eGFP (*Enhanced Green Fluorescent Protein*) (avec une longueur d'onde d'excitation maximale à 488 nm et une longueur d'onde d'émission maximale à 509 nm) et de la TMR comme couple de fluorophores donneur-accepteur respectivement (figure 54).

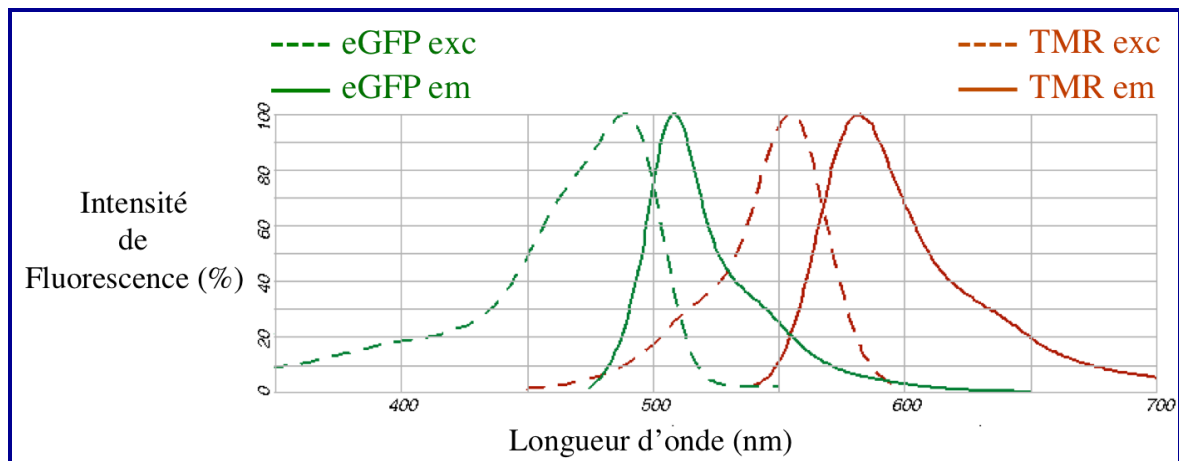


Figure 54 : Couple de FRET eGFP/TMR.

Les spectres d'absorption et d'émission de la eGFP et de la TMR sont présentés avec la longueur d'onde d'excitation du laser (488 nm).

I.3.1. Observation des protéines surexprimées au microscope

La localisation des différentes protéines de fusion a été examinée dans les cellules HeLa. Ces cellules ont été transfectées avec des vecteurs qui expriment les protéines fusionnées avec la eGFP et le domaine SNAP. Après 24h de transfection, les protéines de fusion (SNAP-tag) ont été marquées avec la TMR. La fixation des cellules a été effectuée par le paraformaldéhyde à 4%. Les échantillons sont observés au microscope à fluorescence (figure 55).

Une localisation nucléaire et cytoplasmique a été observée avec la protéine HsRPB11a fusionnée au domaine SNAP ou à la eGFP (figure 55 A N° 1, N° 4). La mutation du motif α (E38A, D39A, L42A) n'affecte pas cette localisation (figure 55 A N° 2). Nous observons la même localisation pour la protéine HsRPB3 fusionnée à la eGFP (figure 55 N° 3).

La co-expression des protéines de fusion SNAP-HsRPB11a et eGFP-HsRPB3 révèle une localisation des deux protéines préférentiellement nucléaire (figure 55 B N° 5). Les deux protéines paraissent co-localiser.

En revanche, la coexpression des protéines de fusion SNAP-HsRPB11a (E38A, D39A, L42A) et eGFP-HsRPB3 présente une co-localisation principalement cytoplasmique (figure 55 B N° 6). Ces mutations dans le motif α paraissent empêcher l'import nucléaire de ces protéines de fusion et/ou piéger les deux protéines dans le cytoplasme.

La co-expression des protéines de fusion SNAP-HsRPB11a et eGFP-HsRPB11a ne semble pas affecter notablement la distribution de ces protéines dans la cellule (figure 55 B piste 7).

Par contre, la co-expression des protéines de fusion SNAP-HsRPB11a (E38A, D39A, L42A) et eGFP-HsRPB11a a un effet dramatique (figure 55 B piste 8). Les deux fusions se concentrent et semblent co-localiser en un site cytoplasmique particulier. Le fait que la fusion eGFP-hRPB11a qui n'est pas mutée voie sa localisation modifiée par la présence de la protéine mutée est difficile à expliquer sans une interaction entre les deux.

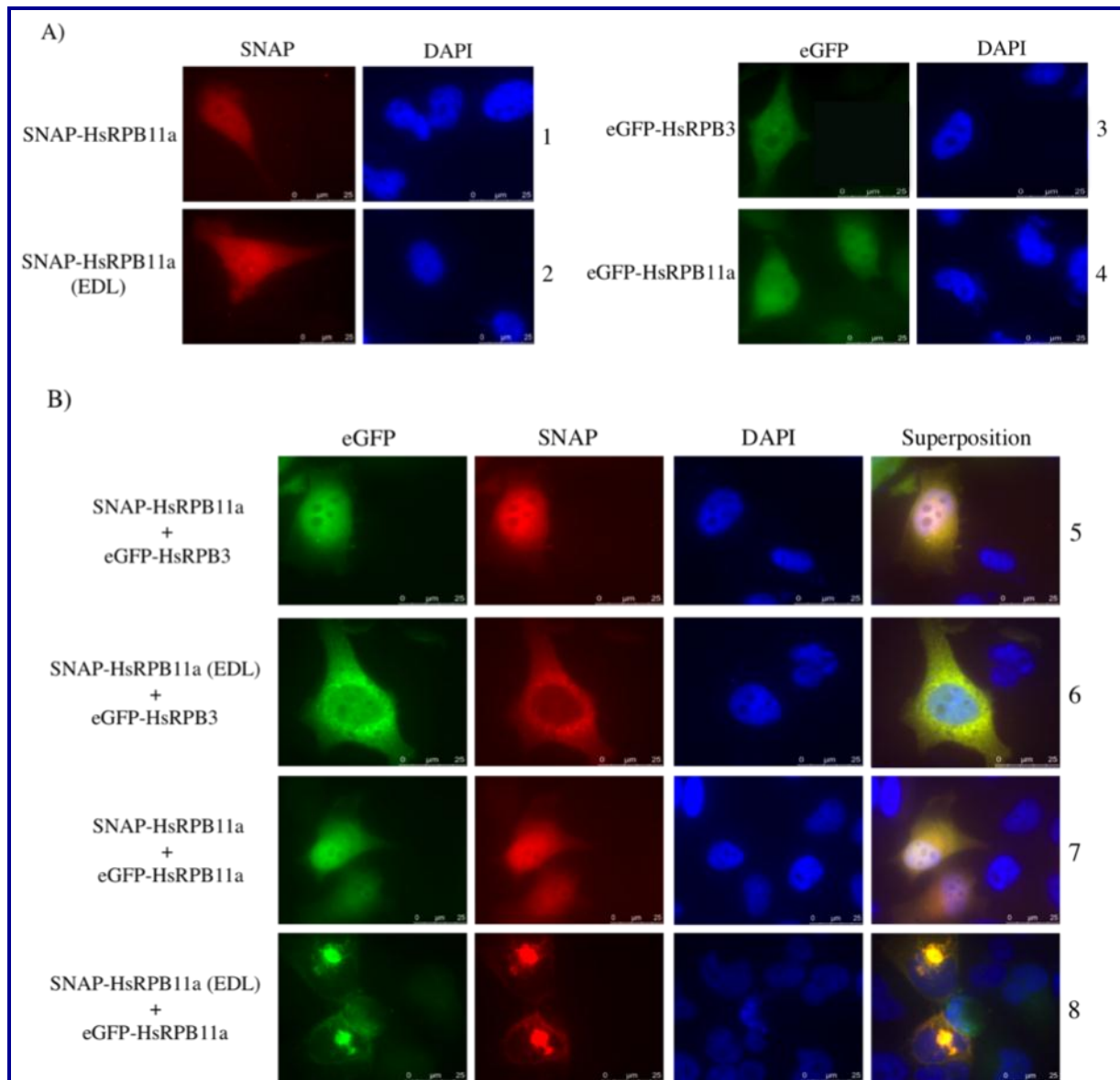


Figure 55 : Localisation subcellulaire des protéines de fusion.

Les différentes protéines de fusion (eGFP et SNAP) ont été surexprimées dans les cellules HeLa. Après 24h, les protéines de fusion (SNAP-tag) ont été marquées avec la TMR. La fixation des cellules a été effectuée par le paraformaldéhyde à 4%. Les échantillons sont observés au microscope à fluorescence. La fluorescence verte et rouge correspond à la eGFP et le SNAP (TMR) respectivement. Les noyaux sont colorés au DAPI en bleu. A) les numéros 1 et 2 correspondent respectivement à la surexpression de la protéine HsRBP11a sauvage ou mutant (EDL) fusionnée avec SNAP-tag. Les numéros 3 et 4 correspondent respectivement à la surexpression de la protéine HsRBP11a ou HsRBP3 fusionnée à la eGFP. B) les numéros 5 à 8 présentent la surexpression des différentes combinaisons des protéines de fusion.

I.3.2. FRET par photoblanchiment (pbFRET)

Une propriété intéressante des fluorophores qui a été exploitée est le photoblanchiment. En effet, quand on expose un fluorophore à une grande quantité d'énergie, il perd ses propriétés de fluorescence. De plus, il est possible de photoblanchir sélectivement l'un des deux fluorophore, si leurs spectres d'excitation sont bien séparés. Lors de phénomènes de FRET, le donneur va transférer une partie de son énergie pour exciter l'accepteur de façon non radiative. En détruisant sélectivement cet accepteur, on aura donc augmentation de l'intensité du donneur et diminution de l'intensité de l'accepteur. La mesure de l'intensité d'émission du donneur avant et après photoblanchiment de l'accepteur nous donnera une estimation de ce transfert d'énergie.

Nous avons effectué l'expérience du FRET par photoblanchiment en microscopie confocale. Dans cette expérience nous avons observé une photoconversion de l'accepteur (TMR) après le blanchiment. Cela signifie que l'accepteur seul émettait, dans le canal du donneur lorsque l'on a fait le blanchiment de l'accepteur. Nos tentatives de mesure du FRET ont donc échoué, car le spectre d'émission du donneur a été contaminé par l'accepteur.

I.4. Analyse de l'homodimérisation par co-immunoprécipitation

Afin de vérifier l'homodimérisation de HsRPB11a dans le contexte de cellules humaines, nous avons entrepris des expériences de co-immunoprécipitations. Des cellules HeLa ont été co-transfectées avec différentes combinaisons de vecteurs d'expression : GFP-HsRPB11a et HsRPB3 ; GFP-HsRPB11a et HsRPB11a ; HsRPB11a et GFP. Les protéines ont été immunoprécipitées à l'aide d'un anticorps anti-GFP. Les protéines précipitées ont été éluées dans du tampon de dénaturation (Laemmli) et séparées par électrophorèse sur gel SDS-PAGE. Ces protéines ont ensuite été transférées sur une membrane de nitrocellulose puis révélées avec des anticorps dirigés contre les protéines HsRPB11a et HsRPB3 (figure 56). Les protéines surexprimées étaient présentes avant la précipitation dans les conditions testées (figure 56, pistes 1, 3 et 5).

La révélation par un anticorps de souris dirigé contre la protéine HsRPB11a a confirmé la présence de la protéine de fusion GFP-HsRPB11a précipitée par l'anticorps anti-GFP (figure 56, pistes 2 et 4). La révélation par des anticorps de lapin

dirigés contre la protéine HsRPB3 confirme la présence attendue de HsRPB3 dans le matériel co-précipité par la fusion GFP-HsRPB11a (figure 56, piste 2). Cette interaction constitue notre contrôle positif.

Une bande faible mais clairement détectable est visible ; elle correspond à la protéine HsRPB11a co-précipitée avec la protéine de fusion GFP-HsRPB11a mais pas avec la GFP seule (figure 56, piste 4 et 6). Ceci confirme la capacité de l'homodimérisation de la protéine HsRPB11a dans les cellules humaines.

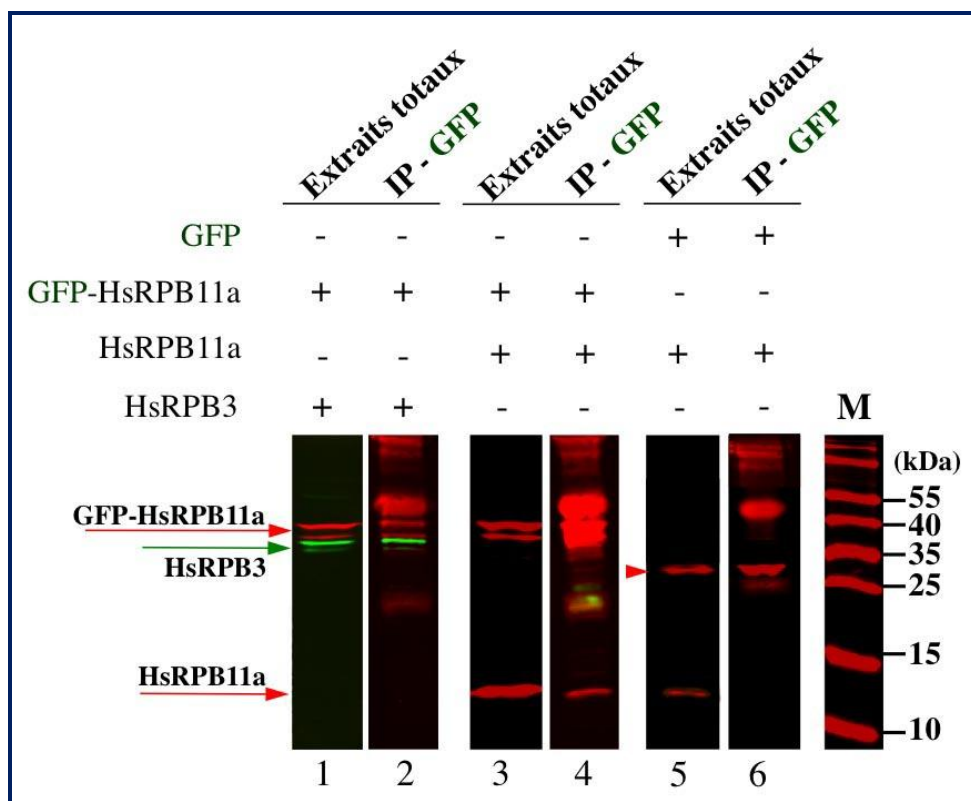


Figure 56 : Analyse de l'homodimérisation par co-immunoprécipitation.

Les cellules HeLa sont co-transfectées pendant 24h avec les plasmides permettant l'expression des protéines GFP, GFP-HsRPB11a, HsRPB11a et HsRPB3 ainsi qu'indiqué dans le haut de la figure. Les protéines totales sont ensuite extraites. Les lysats cellulaires obtenus sont soumis à une immunoprécipitation à l'aide des billes de protéine G-sépharose couplées à l'anticorps anti-GFP. Les billes sont lavées 3 fois avec le tampon phosphate (PBS). Les protéines isolées sont séparées sur un gel en gradient de polyacrylamide (SDS-PAGE 10-20 %). Les protéines sont électrotransférées sur une membrane de nitrocellulose. La membrane est incubée avec des anticorps anti-HsRPB3 (signal vert) (pistes 1 et 2), anti-HsRPB11a (signal rouge) (pistes 1 à 6) et anti-GFP (pistes 5 et 6). La présence des anticorps secondaires couplés à une molécule fluorescente sur la membrane est vérifiée par imagerie Odyssey®. Les cellules transfectées avec le vecteur GFP / HsRPB11a sont utilisées comme contrôle négatif (pistes 5 et 6). Les bandes correspondant à la GFP sont montrées avec un triangle rouge. M : marqueur de poids moléculaires en kDa.

II. Discussion et perspectives

Détection de l'interaction Hs11a/Hs11a

Nos expériences de double-hybrides dans la levure ont montré que l'isoforme humaine HsRPB11a a la capacité de former des homodimères. Ceci est cohérent avec le comportement chromatographique sur gel exclusion de la protéine HsRPB11a surproduite qui se comporte comme un dimère, et avec les images préliminaires déduites de l'analyse des cristaux obtenus (voir chapitre 1). La présence de domaines fusionnés en N-terminal de HsRPB11a ne semble donc pas gêner cette interaction. A l'appui de cette conclusion, mentionnons ici un résultat non publié : il a été observé au laboratoire dans des expériences de complémentation que la fusion N-terminale de ScRPB11 est pleinement fonctionnelle in vivo dans la levure.

Par ailleurs, les expériences de co-immunoprécipitation démontrent qu'en effet, la fusion N-terminale eGFP-HsRPB11a est capable d'entraîner la protéine HsRPB11a, et ceci cette fois-ci dans le contexte d'une cellule humaine en culture. La capacité de cette protéine à homodimériser ne fait guère de doute.

Toutes ces expériences ont cependant en commun une surproduction significative des protéines observées, ce qui est de nature à engendrer certains artéfacts, notamment en favorisant des interactions potentiellement faibles aux concentrations physiologiques des protéines étudiées.

Comparaison des interactions RPB11/RPB11 et RPB11/RPB3

Il nous est néanmoins possible d'en analyser les modalités par mutagenèse ciblée de l'un ou l'autre des partenaires de l'interaction. Le système des double-hybrides se prête le mieux à ce type d'analyse. Nous avons analysé en parallèle l'effet des mutations affectant HsRPB11a sur l'hétérodimère HsRPB11a/HsRPB3.

Nous avons observé que trois résidus conservés (E38, D39 et L42) dans le motif α sont importants pour la dimérisation de HsRPB11a, alors que la mutation des résidus E38A et D39A ne semble pas l'affecter. L'intégrité de la Leu 42 est donc critique pour le phénomène d'homodimérisation.

En revanche, l'interaction entre HsRPB11a et HsRPB3 est fortement affectée aussi bien par la mutation des deux résidus acides E38, D39 que par la triple

mutation E38A, D39A, L42A (voir plus haut). La figure 57 présente une comparaison de l'interaction entre ScRPB11 et ScRPB3 structure établie dans le contexte de l'ARNP II complète de la levure et de l'interaction potentielle entre HsRPB11a avec elle-même qui n'est qu'hypothétique. Nous n'avons certes pas de garantie que l'interface entre HsRPB11a et HsRPB3 soit identique à celle de ses homologues de levure, mais le fait que HsRPB11b α (très semblable tout de même à HsRPB11a) puisse remplacer ScRPB11 plaide pour une certaine conservation de cette interface.

Le motif α N-terminal de RPB11 se trouve proche d'une petite hélice alpha qui interagit de façon anti-parallèle avec celle de son partenaire. La Leu 42 est en particulier localisée dans le premier tour d'hélice et directement impliquée dans l'interaction avec l'hélice alpha de ScRPB3. Les résidus E38 et D39 sont localisés juste devant cette hélice et leur contribution semble critique pour l'interaction. Si notre modèle de l'homodimère se révèle juste, les deux hélices alpha de HsRPB11a interagiraient différemment, écartant de l'interface les résidus E38 et D39 qui ne seraient donc plus impliqués dans l'interaction (figure 57). Il y aurait donc une certaine compatibilité entre nos observations expérimentales et le modèle provisoire de l'homodimère de HsRPB11a.

L'absence de la grande hélice C-terminale (79-117) de HsRPB11a empêche toute interaction, que ce soit avec HsRPB3 ou avec HsRPB11a. Cette hélice (79-117) isolée est capable d'interagir avec la sous-unité HsRPB3 mais pas du tout avec HsRPB11a. L'analyse de nos modèles (figure 57) permet peut-être de rendre compte de cette observation : on voit clairement des feuillets β localisés dans le domaine N-terminal de ScRPB3 interagir avec l'hélice C-terminale de ScRPB11, contribuant à la stabiliser. Un tel domaine n'existe pas dans la protéine HsRPB11a, l'interaction entre HsRPB11a et l'hélice C-terminale isolée ne peut avoir la même stabilité.

La protéine HsRPB11a n'interagit donc pas de la même façon avec elle-même et avec la sous-unité HsRPB3.

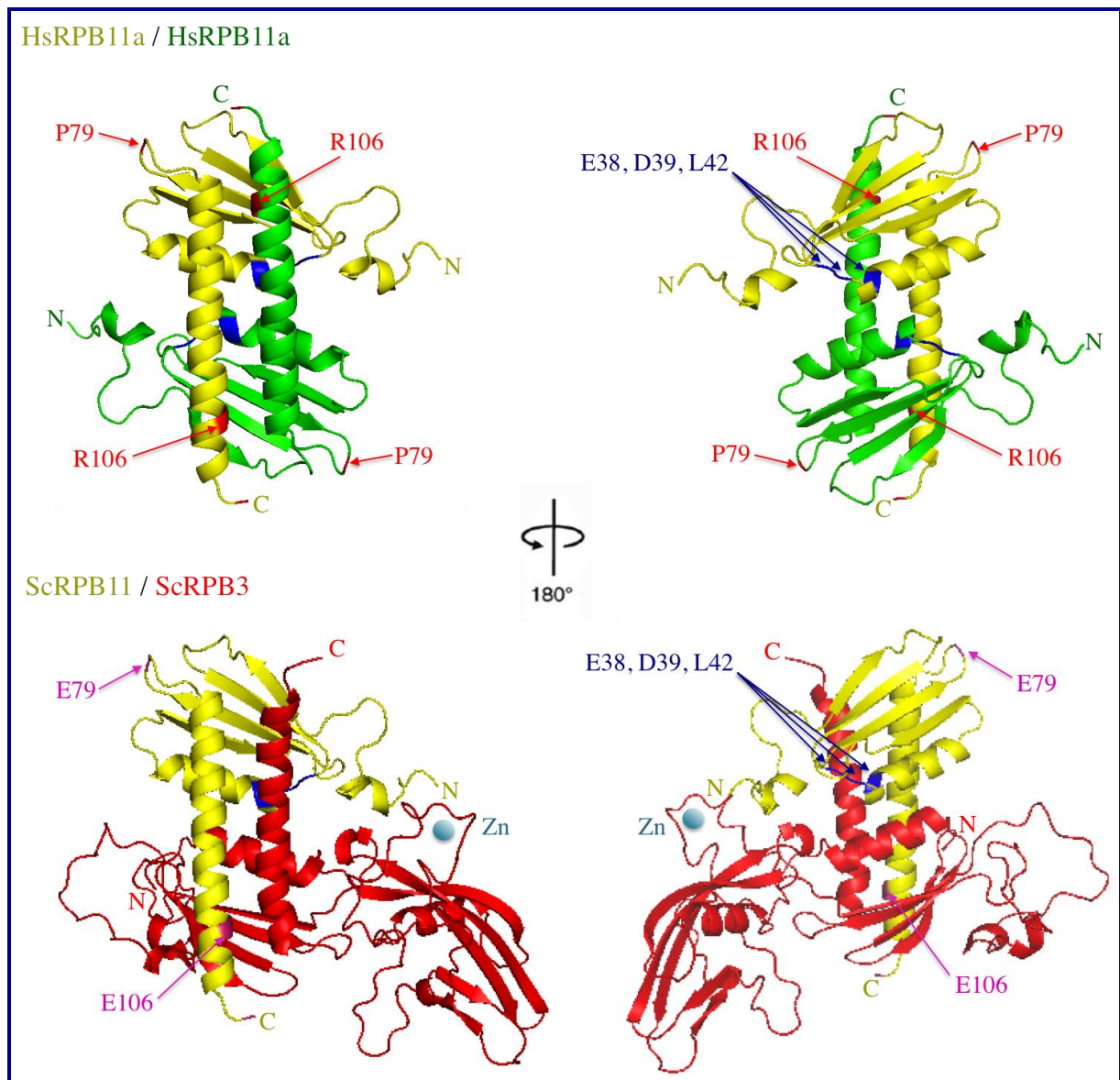


Figure 57 : Comparaison des interactions HsRPB11a/HsRPB11a et ScRPB3/ScRPB11.

Représentation en rubans de la structure de l'homodimère HsRPB11a et l'hétérodimère ScRPB3/ScRPB11. Les résidus conservés dans le motif α (E38, D39, L42) sont en bleu. Les résidus 79 et 106 à partir desquels les mutants de délétions de la partie C-terminale de HsRPB11a sont construits sont montrés. L'atome de zinc est représenté par une sphère bleu clair.

Signification physiologique ?

Le fait que deux interactions soient possibles pour HsRPB11a dans la cellule pose la question de leur signification physiologique. Il faut tout d'abord relever que

cette homodimérisation n'est pas observée pour l'orthologue de levure ScRPB11 dans des conditions expérimentales en tout point comparables. Il s'agirait donc d'une propriété particulière aux protéines humaines. En effet, l'isoforme HsRPB11b α présente également cette propriété de pouvoir interagir avec elle-même dans un test de double-hybride. On peut imaginer qu'il s'agit d'une interaction peu spécifique qui ne se manifeste que pour solubiliser la protéine dans des conditions de surexpression. Pourtant, les deux isoformes humaines n'interagissent pratiquement pas entre elles : ces interactions homodimériques permettent donc de distinguer les isoformes, une propriété qui mérite d'être notée.

Pour le reste, la durée de vie *in cellulo* de tels homodimères est inconnue. Ils pourraient constituer de simples intermédiaires d'assemblage. On peut aussi imaginer que leur formation peut être éventuellement régulée, mais il se pose la question préalable de l'existence des isoformes HsRPB11b α *in vivo* dans des cellules humaines.

Des expériences d'immunofluorescence nous ont permis d'observer que des fusions N-terminales de HsRPB11a se répartissent dans l'ensemble de la cellule de même que les fusions de HsRPB3. Leur co-expression semble entraîner une augmentation de la concentration de RPB3 dans le noyau, ce qui matérialiserait leur interaction (voir figure 55). Surtout, la co-expression de HsRPB11a mutée dans son motif α (E38A, D39A et L42A) empêche cette relocalisation et piège HsRPB3 dans le cytoplasme (figure 55).

Cette mutation a un effet encore plus spectaculaire sur la colocalisation des fusions eGFP-HsRPB11a sauvage et de la fusion SNAP-HsRPB11a mutée, puisque les deux protéines se retrouvent piégées dans une structure cytoplasmique qui pourrait être l'agrésome. Cet effet paraît suggérer que les deux protéines interagiraient dans ce contexte, contrairement à ce que les données de double-hybride semblaient indiquer. Une séquestration de ce type expliquerait à elle seule l'absence de dimère dans le noyau qui laisserait conclure à une absence d'interaction. Il reste à démontrer si cet effet existe ou non dans la levure.

Quoiqu'il en soit, cette colocalisation et relocalisation des deux fusions HsRPB11a mutées ou non sont un argument supplémentaire en faveur de l'existence d'une interaction homodimérique dans le contexte d'une cellule humaine.

Chapitre 3 : Recherche de l'isoforme HsRPB11b α

I. Stratégie

Nous avons mis au point une approche utilisant les siRNA dirigés contre le messager HsRPB11a afin de réduire au maximum la quantité de cette protéine dans la cellule. Nous avons essayé de mettre en évidence la présence de la protéine HsRPB11b α par immunoprécipitation et Western-Blot en utilisant des anticorps spécifiques dirigés contre HsRPB11b α .

I.1. Détection de l'ARNm des différentes isoformes

Les deux ARN messager des isoformes HsRPB11a et HsRPB11b α ont été mis en évidence par RT-PCR. Les ARN totaux sont purifiés à partir d'une lignée de cellules HeLa par extraction au Trizol (Invitrogen). La reverse transcription a été amorcée à l'aide d'un oligonucléotide oligo dT. Des amorces spécifiques des isoformes humaines de RPB11 ont été synthétisées. Elles ont été choisies à partir des séquences génomiques publiées dans GenBank (NM-006234, NM-032959) afin d'amplifier spécifiquement l'ADNc de chacune des isoformes (figure 58). Les caractéristiques choisies pour les primers sont une taille d'amplicon comprise entre 75 et 130 pb, un Tm proche de 60°C et une taille de primer comprise entre 20 et 30 pb (Tableau 10). L'amplification de l'ADNc de l'hypoxanthine guanine phosphoribosyl transférase (HPRT) a été utilisée comme contrôle interne.

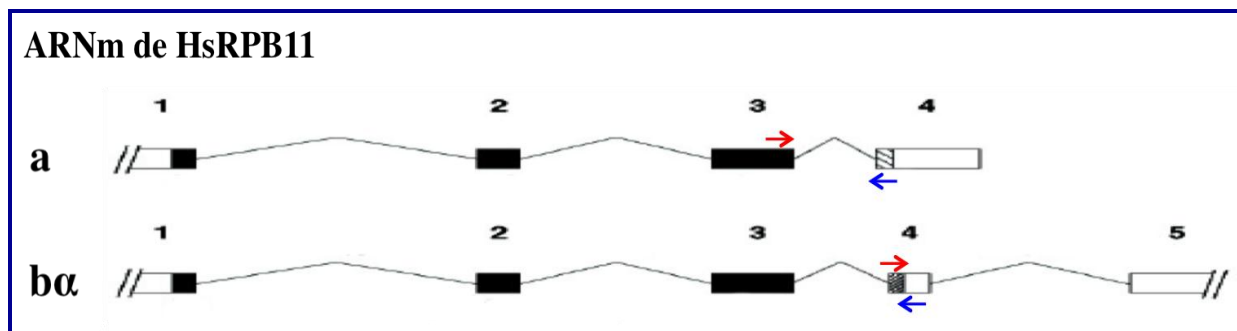


Figure 58 : Positions des oligonucléotides utilisés sur l'ARN messager des isoforme HsRPB11.

	Oligonucléotides sens	Oligonucléotides Anti-sens
<i>HsRPB11a</i>	5' -ATCCGAGTGCAGACCACGCCGGAC-3'	5' -TCTTTTATGGCCACCCGAAAGCGCRCCT-3'
<i>HsRPB11bα</i>	5' -ACGTGCCTGCTTCCCTTCG-3'	5' -TGGGGCACATAGCTGATGAG-3'
<i>HPRT</i>	5' -GACTAATTATGGACAGGACTGAACG-3'	5' -CCAGCAGGTCAGCAAAGAATT-3'

Tableau 10 : Liste des couples d'oligonucléotides utilisés en RT-PCR.

Nous avons identifié les messagers correspondant aux deux isoformes de HsRPB11 (figure 59).

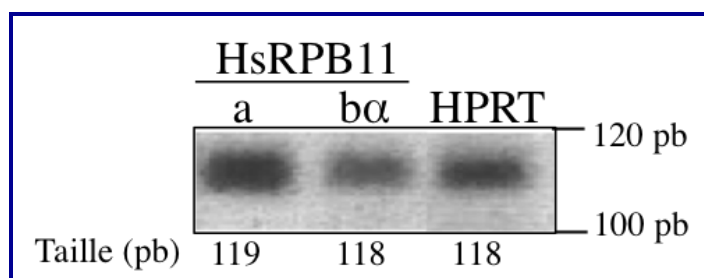


Figure 59 : Identification des ARN messagers correspondant aux deux isoformes de HsRPB11 par RT-PCR.

Les séquences codantes des deux isoformes de RPB11 ont été amplifiées et analysées sur un gel d'agarose de 2,5 %. L'amplification de l'ADNc de l'HPRT a été utilisée comme contrôle interne. Les tailles de tous les fragments amplifiés sont indiquées en paires de bases (pb).

I. 2. Analyse des effets de différents siRNAs par qRT-PCR

Deux séquences siRNA « 19 » et « 20 » (respectivement si10819 et si10820 commandé chez Applied Biosystems) ont été utilisées pour le gène POLR2j1 (Tableau 11). Ils sont dirigés contre la partie 3' non codante spécifique de HsRPB11a. Une séquence siRNA « 54 » (si54728 de Applied Biosystems) dirigée contre l'exon 2 du gène POLR2j a été utilisée comme contrôle, car il devrait agir sur les messagers de toutes les isoformes de RPB11 (Tableau 11 et figure 60).

siRNA	sens	Anti-sens
si10819	CGGUGACUUCGCAAGCAAAtt	UUUGCUUGCGAAGUCACCGct
si10820	GCAUAACUCGAAACACCUUtt	AAGGUGUUUCGAGUUAUGCag
si54728	CCCAAUGCCUGUUUAUUCAtt	UGAAUAAACAGGCAUUGGGta

Tableau 11 : Liste des séquences des siRNA utilisés

Les cellules HeLa sont transfectées par les différentes concentrations de siRNA (10 à 35 nmol) en utilisant l'agent de transfection HiPerFect de QIAGEN. Un siRNA portant un fluorochrome été utilisé en parallèle en tant que contrôle de transfection. Les cellules sont récoltées après 48 et 72 heures de transfection. Les ARN extraits et purifiés sur colonne (QIAGEN) sont rétrotranscrits en ADNc.

Nous avons mesuré la quantité d'ARNm du gène cible par PCR quantitative (qRT-PCR) afin d'évaluer le pourcentage d'inhibition de l'expression. L'amplification de l'ADNc de l'HPRT a été utilisée comme contrôle interne. Le niveau d'expression

des ARNm a été normalisé par rapport à celui de l'HPRT. Les données d'expression de RT-PCR sont analysées avec le logiciel StepOne de Applied Biosystems.

Le même niveau d'inhibition de l'expression des ARNm a été observé à partir d'une concentration de 20 nmol des siRNA. Malgré une bonne efficacité de transfection (> 70%), un très faible niveau d'inhibition de l'ARNm de HsRPB11a (25 à 30 %) a été obtenu pour les siRNA « 19 » et « 20 » respectivement ou ensemble « 19/20 » (figure 60).

Dans les mêmes conditions le siRNA « 54 » s'est révélé plus efficace (70 à 80% d'inhibition de l'ARNm, 2 à 3 jours après la transfection), mais aucun phénotype majeur sur la croissance cellulaire n'est observable.

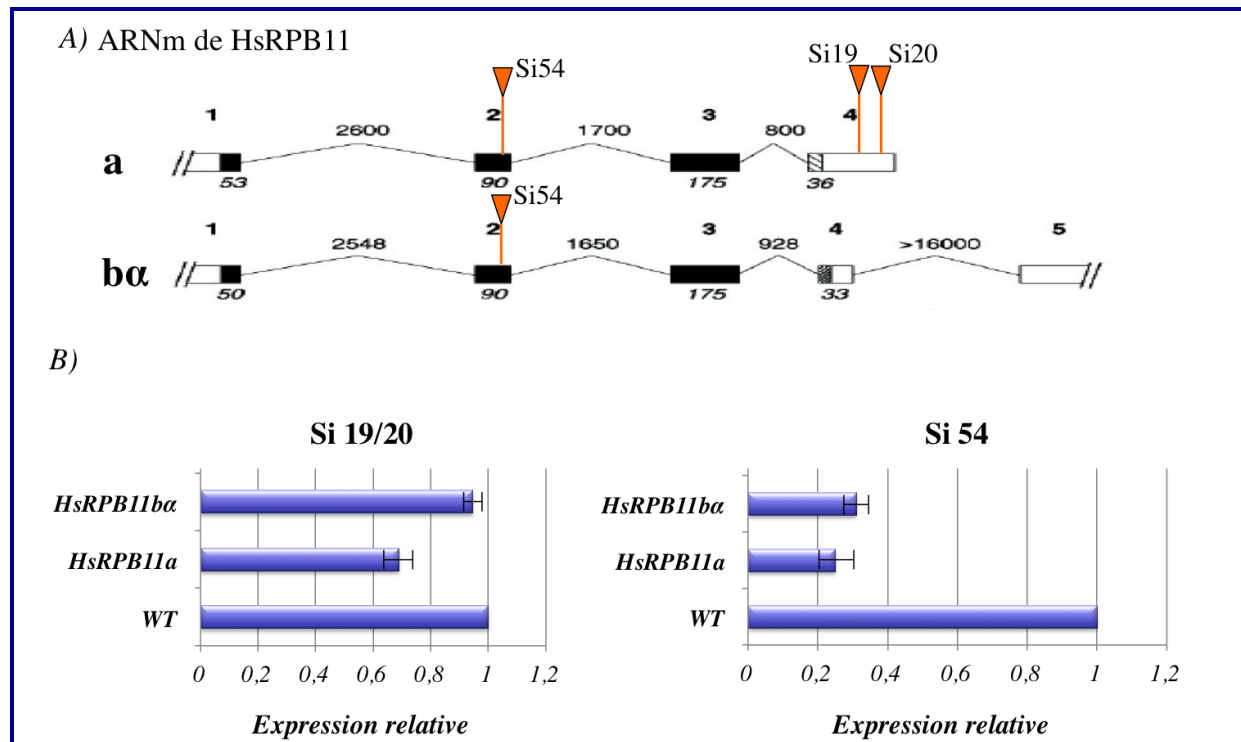


Figure 60 : Analyse de la fonctionnalité des siRNA dirigés contre le gène POLR2j1 par qRT-PCR.

Les qRT-PCR ont été réalisées à partir des ADNc extraits des cellules HeLa transfectées par deux siRNA 19 et 20 dirigés contre la partie 3' non codante spécifiques de HsRPB11a et un siRNA 54 dirigés contre l'exon 2 du gène POLR2j. A) La position des siRNA est montrée sur les ARNm des isoformes HsRPB11. B) les niveaux d'expression des ARNm obtenus.

Les essais d'extraction des protéines sont réalisés à partir de ces cellules pour étudier l'expression de la protéine HsRPB11a endogène et aussi des expériences d'immunoprécipitation et d'immunofluorescence ont été effectuées, mais la limite de sensibilité de nos anticorps n'a pas permis de détecter la protéine endogène.

II. Discussion et perspectives

L'analyse des interactions protéine-protéine par la technique du double-hybride a montré que la protéine HsRPB11b α se comporte comme la protéine HsRPB11a qui est une protéine fonctionnelle dans le complexe de l'ARNP II. Cependant, la présence effective de HsRPB11b α dans la cellule est une question que nous nous posons de façon récurrente depuis quelques années dans notre équipe.

Nous avons réussi à détecter les ARNm correspondant aux deux isoformes dans les cellules HeLa. Par une approche de siRNA nous avons tenté de réduire spécifiquement le messenger codant la protéine HsRPB11a en utilisant deux siRNA « 19 et 20 » dirigés contre la partie non codante de l'exon 4. Malheureusement, ces deux siRNA ne se sont pas avérés efficaces.

Le siRNA « 54 » ciblant l'exon 2 a été beaucoup plus efficace et nous avons obtenu une inhibition de 80% de l'expression des messagers codant les deux sous-unités. L'absence de phénotype majeur sur la croissance cellulaire peut être dû à la présence d'une quantité suffisante de la protéine HsRPB11a dans la cellule. Malheureusement, nous n'avons pas pu examiner l'effet de cette inhibition sur le niveau d'expression de la protéine endogène, car le seul anticorps disponible ne permet de détecter que la protéine surproduite.

Suite à ces résultats négatifs, nous avons entrepris d'établir des lignées cellulaires stables qui expriment soit l'isoforme HsRPB11b α ou l'isoforme HsRPB11a et un shRNA fonctionnel dirigé contre HsRPB11a. Ces lignées nous serviront à établir si les protéines HsRPB11b α s'intègrent effectivement dans les complexes ARNP II (expériences d'immunoprécipitation de l'ARN polymérase II).

Discussion générale

Discussion générale

L'ARN polymérase II (ARNP II) transcrit la quasi-totalité du génome nucléaire eucaryotique. Nombre d'oncogènes agissent au niveau de l'expression des gènes en tant que facteurs de transcription (*c-jun*, *c-fos*, *c-myc* etc...). Le contrôle de l'activité de l'ARNP II est donc crucial dans le développement de nombre de cancers. En outre, cette enzyme rencontre des dommages dans sa matrice d'ADN au cours de la phase d'élongation. Le mécanisme de « Transcription Coupled Repair » permet de catalyser efficacement la réparation de la matrice. Ses défauts sont également responsables de susceptibilités à développer divers cancers (Xeroderma pigmentosum).

Nous nous sommes intéressés à l'ARNP II elle-même. L'ARNP II de levure est constituée de 12 sous-unités et sa structure a été établie par cristallographie à 2,8 Å (Cramer et al., 2001). Chez la levure, chaque sous-unité est codée par un gène unique. Ils sont nommés RPB1 à RPB12, 10 d'entre eux sont essentiels. Dans le génome humain, ils sont nommés POLR2A à L selon la nomenclature internationale (HUGO), et codent les sous-unités classées dans un ordre de tailles décroissantes.

Etudes structurales de l'ARNP II humaine

Les structures de quatre sous-unités isolées de la polymérase humaine ont été établies, soit par cristallographie, soit par RMN (del Rio-Portilla et al., 1999; Kang et al., 2006; Meka et al., 2005). La structure 3D de l'ARNP II humaine complète a été analysée par la cryo-microscopie électronique à une résolution de 22 Å (Kostek et al., 2006). Cette méthode d'analyse a permis de montrer l'existence de plusieurs conformations possibles pour cette enzyme, en l'absence même de tout substrat ou interactant.

Cependant la structure cristallographique de l'enzyme humaine n'a pas encore été élucidée. L'enzyme de la levure *Saccharomyces cerevisiae* est sans doute un bon modèle de l'enzyme humaine ainsi qu'on peut le penser du fait de l'existence de complémentations fonctionnelles de plusieurs sous-unités de l'ARNP II humaine dans la levure. Il demeure que l'homme n'est pas la levure et que cette étude reste à faire. Une équipe américaine (Dr A. Gnatt) a été capable, voici quelques années, de cristalliser des préparations d'ARNP II humaine et a obtenu des cartes de diffraction. Afin de déterminer la phase de leur cristaux, il leur fallait les cartes de diffraction de

sous-unités humaines isolées. Nous avons décidé d'entreprendre l'étude de HsRPB11a que nous savions soluble.

Au cours de ce travail de thèse, nous avons réussi à cristalliser cette protéine HsRPB11a surproduite et purifiée à partir de plusieurs sources (*Anabaena species*, *Escherichia coli*). Des cristaux ont été obtenus dans différentes conditions, soit de la protéine sauvage, soit de divers dérivés. Les meilleurs cristaux diffractent à une résolution de 2,9 Å. Ils semblent néanmoins mâclés et la structure finale n'est pas encore résolue. Ce travail se poursuit.

La détermination de la structure de HsRPB11a sera une contribution aux efforts de la communauté scientifique pour l'élucidation de la structure tridimensionnelle de l'ARN polymérase II humaine qui recèle peut-être encore des surprises.

L'assemblage de l'ARNP II

Le mécanisme par lequel chacune des 12 sous-unités de la polymérase II est assemblée dans le complexe est encore mal connu.

Les sous-unités RPB3 et RPB11 présentent des homologies avec la sous-unité α de l'ARNP d'*E. coli*. L'hétérodimérisation de RPB11 avec RPB3 joue un rôle structural fondamental dans l'assemblage de l'enzyme qui rappelle celui de la sous-unité alpha de la ARNP eubactérienne (Kimura et al., 1997). Il avait par ailleurs été montré que la sous-unité RPB3 est essentiel pour l'assemblage des deux grandes sous-unités via un complexe intermédiaire formé de RPB3 et RPB2 (Kolodziej and Young, 1991). Ce sous complexe stable recruterait la sous-unité RPB1 et constituerait ainsi, comme pour l'enzyme eubactérienne, le premier intermédiaire d'assemblage d'un complexe stable. Il semblait qu'une similitude entre le système eubactérien et eucaryotique se dessinait.

Il restait pourtant à expliquer le recrutement des huit autres sous-unités et savoir comment se transpose ce travail au système humain. Plusieurs travaux récents ont mis en lumière les mécanismes de transport de l'ARNP II assemblée dans le cytoplasme vers le noyau (voir pour revue (Corden, 2011)). Les observations faites indiquent que plusieurs voies semblent exister qui impliquent des protéines différentes, ainsi lwr1, conservée de la levure à l'homme, qui se fixe dans le site de fixation de l'ADN du complexe et exhibe un signal de localisation nucléaire nécessaire pour diriger le complexe vers le noyau. En ce qui concerne le détail de l'assemblage des sous-unités, il reste beaucoup à découvrir, mais l'implication de

plusieurs protéines jouant un rôle d'assemblage a été établie (Boulon et al., 2010). Ainsi, Hsp90 qui stabilise la grande sous-unité HsRPB1 dans le cytoplasme avec d'autres protéines. De façon générale, il semble que deux sous-complexes soient assemblés autour des deux grande sous-unités séparément, d'un côté Les sous-unités 2,3,10,11,12 et de l'autre, 1,4,5,6,7,8,9. On retrouve ici en partie ce qui avait pu être observé dans la levure (voir plus haut). Le détail de l'ordre des interactions entre les sous-unités reste toutefois à établir.

Ainsi que nous l'avons discuté plus haut, nos résultats indiquent que la sous-unité HsRPB11a est capable de former un homodimère, au contraire de son homologue de levure qui en est incapable. Dans quelle mesure la formation de l'homodimère de HsRPB11a peut-elle intervenir dans l'assemblage ? Il est sans doute bien moins stable que l'hétérodimère HsRPB3/HsRPB11a, mais pourrait être stabilisé *in vivo* dans des conditions qui restent à établir. Dans l'hypothèse où d'autres isoformes pourraient être présentes dans la cellule humaine, ce pourrait être une étape de régulation dans le processus d'assemblage de l'ARNP II.

Existence de plusieurs isoformes de RPB11 chez l'homme

Les gènes qui codent les sous-unités de l'ARNP II humaine sont uniques sauf un : il existe dans le génome humain d'une famille de 4 gènes (POLR2J1 à 4) susceptibles de coder des isoformes de la sous-unité RPB11 de l'ARNP II (Grandemange et al., 2001). Le gène POLR2J1 code la sous-unité HsRPB11a dont la séquence peptidique est identique à son orthologue de souris qui, lui, est unique. Cette protéine est retrouvée dans le complexe ARNP II fonctionnel.

Les trois gènes POLR2J2, 3, 4, sont spécifiques des grands singes, plus exactement des hominins, et n'existent ni chez autres grands singes comme par exemple l'orang-outan et autres petits singes ni chez la souris (Shematorova et al., 2010). La duplication du gène POLR2J2 qui a donné POLR2J3 serait la plus récente (Shematorova et al., 2010). Nous savons que les gènes POLR2J2, 3, 4 sont transcrits et que les niveaux de ces transcrits varient d'un tissu à l'autre, mais nous n'avons pas pu mettre en évidence les protéines correspondantes, même en inhibant l'isoforme HsRPB11a. Ces transcrits sont soumis à des épissages alternatifs, ce qui n'est pas le cas des transcrits de POLR2J1 (Grandemange et al., 2001).

L'isoforme codée par POLR2J2 et 3, HsRPB11b α , est fonctionnelle et constitue une sous-unité viable de l'ARNP II dans la levure (Grandemange et al., 2001).

Nos observations peuvent s'interpréter en supposant que les gènes qui codent HsRPB11b α ont été sélectionnés par l'évolution dans les génomes des grands singes pour produire de protéine avec des propriétés nouvelles. L'expression de cette protéine peut être requise uniquement dans des circonstances particulières, qui restent à identifier. Nous proposons la conjecture qu'il puisse exister un deuxième forme de l'ARNP II humaine qui intègre la sous-unité HsRPB11b α à la place de HsRPB11a, lui conférant une ou plusieurs fonctions spécifiques.

ANNEXE

Analyse génétique des sous-unités de l'ARNP II de *Plasmodium falciparum* dans la levure

Le paludisme est une maladie infectieuse potentiellement mortelle transmise par la piqûre d'un moustique du genre *Anopheles*. Elle est causée par un parasite eucaryote unicellulaire de la famille des Apicomplexes et du genre *Plasmodium*.

Chez l'homme, quatre espèces de plasmodies sont infectieuses : *P. falciparum*, *P. vivax*, *P. ovalae* et *P. malariae*. Parmi ces quatre agents responsables du paludisme identifiés à ce jour, *P. falciparum* est celui qui cause les formes les plus graves de la maladie.

Il existe contre ce parasite un certain nombre de molécules susceptibles de limiter l'infection et ses effets. Cependant l'efficacité de ces antipaludiques est sans cesse remise en cause par l'évolution particulièrement rapide du parasite. Il est devenu urgent de trouver de nouvelles cibles thérapeutiques dans le métabolisme du parasite.

Par ailleurs, la publication de la séquence nucléotidique du génome de *P. falciparum* ainsi que celle de son plastide, l'apicoplaste, et celle de son génome mitochondrial suscite de légitimes espoirs d'identifier de nouvelles cibles parmi les séquences codantes identifiées (Bozdech et al., 2003).

Chez *P. falciparum*, l'organisation des gènes est comparable à ce que l'on connaît de la structure des gènes chez les eucaryotes. De plus, le génome nucléaire de *Plasmodium falciparum* contient des candidats pour l'ensemble des 12 sous-unités de l'ARN polymérase II (Coulson, 2004). Les gènes de *P. falciparum* étant transcrits de façon très spécifique au cours du cycle infectieux, une altération des fonctions transcriptionnelles devrait perturber de manière sensible le développement du parasite (Bozdech et al., 2003).

Dans l'optique de l'identification de cibles nouvelles à des fins thérapeutiques, l'ARNP II de *P. falciparum* nous est apparu comme un candidat stratégiquement intéressant. La mise en évidence d'inhibiteurs spécifiques de l'enzyme du parasite pourrait ainsi être le point de départ de nouveaux antipaludiques contre l'une des plus graves maladies parasitaires humaines.

Nous avons entrepris la caractérisation génétique de l'ARNP II de *P. falciparum* par la méthode de complémentation. Les séquences codantes des 10 petites sous-

unités ont été clonées dans un vecteur d'expression. Nous n'avons pas réussi à cloner les séquences complètes de deux grandes sous-unités dans aucun des vecteurs de clonage utilisés.

Les expériences de complémentation fonctionnelle ont été réalisées pour ces dix sous-unités de l'ARNP II de *P. falciparum* dans les levures. La complémentation des sous-unités de l'ARNP II humaine dans les levures a été effectuée en parallèle comme contrôle.

Nous avons montré que cinq de ces sous-unités sont fonctionnelles dans la levure (PfRPB4, 5, 7, 9 et 12) confirmant que ces protéines sont bien les orthologues de l'enzyme parasite (Tableau 1 de l'article).

Nous avons ainsi construit 5 souches viables de levures exprimant chacune une sous-unité de l'ARNP II de *P. falciparum*. Quatre de ces souches présentent un phénotype, à savoir croissance ralentie par rapport aux souches sauvages correspondantes, et surtout à température élevées. Seule la souche exprimant PfRPB9 est quasiment sans phénotype (figure 1 de l'article).

Sous-unités de l'ARNP II	<i>P. falciparum</i>		<i>Homo sapiens</i>	
	Complémentation	(%) Homologie	Complémentation	(%) Homologie
RPB1	ND	49	+	55
RPB2	ND	55	-	61
RPB3	-	38	-	45
RPB4	+	32	+	31
RPB5	+	43	-	45
RPB6	-	47	+	60
RPB7	+	35	+	43
RPB8	-	36	+	33
RPB9	+	32	+	39
RPB10	-	54	+	73
RPB11	-	36	-	47
RPB12	+	40	+	38

Tableau 1 : Comparaison des résultats de complémentation Homme-levure et *Plasmodium*-Levure. Les sous-unités qui complémentent sont indiquées par un (+) et celles ne complémentant pas par un (-). Les pourcentages d'homologie des protéines sont indiqués entre Homme et levure d'une part et *P. falciparum* et Levure d'autre part (Hazoume et al., 2011).

L' α -amanitine est un inhibiteur sélectif de l'ARNP II à une faible concentration. Une étude comparative du site de fixation de l' α -amanitine sur l'ARNP II humaine, de levure et de *P. falciparum* a été effectuée. Pour cela, nous avons remplacé dans la protéine de levure ScRPB1 soit les 6 acides aminés spécifiques interagissant directement avec l' α -amanitine, soit le domaine complet (176 acides aminés) impliqué dans cette interaction, par les équivalents correspondant aux protéines de *P. falciparum* « plasmodification » ou de la protéine humaine « humanisation » (figure 2 de l'article).

Les expériences de complémentation des différentes chimères de ScRPB1 ont montré que seules les deux constructions où les 6 acides aminés ont été remplacés étaient viables.

Ce travail a été la première tentative de caractérisation génétique de l'ARNP II de *P. falciparum*. Nous avons construit des mutants viables de cette enzyme qui présentent des modifications dans le centre actif au sein des régions de fixation de l' α -amanitine. Les souches de levure obtenues pourront être utilisées pour identifier de nouveaux médicaments antipaludiques potentiels.



Short communication

A genetic analysis of *Plasmodium falciparum* RNA polymerase II subunits in yeastAdonis Hazoume^a, Kambiz Naderi^a, Ermanno Candolfi^b, Claude Kedinger^a, Bruno Chatton^a, Marc Vigneron^{a,*}^a Université de Strasbourg, Institut de Recherche de l'École Supérieure de Biotechnologie de Strasbourg, CNRS-UMR7242, Ecole Supérieure de Biotechnologie de Strasbourg, BP10413, 67412 Strasbourg Illkirch Cedex, France^b Institut de Parasitologie et de Pathologie Tropicale, EA4438, Université de Strasbourg et Hôpitaux Universitaires de Strasbourg, 3 rue Koeberlé, 67000 Strasbourg, France

ARTICLE INFO

Article history:
 Received 30 September 2010
 Received in revised form
 21 December 2010
 Accepted 23 December 2010
 Available online 11 January 2011

Keywords:
 RNA polymerase II
Plasmodium falciparum
 Complementation
 Yeast

ABSTRACT

RNA polymerase II is an essential nuclear multi subunit enzyme that transcribes nearly the whole genome. Its inhibition by the alpha-amanitin toxin leads to cell death. The enzyme of *Plasmodium falciparum* remains poorly characterized. Using a complementation assay in yeast as a genetic test, we demonstrate that five *Plasmodium* putative RNA polymerase subunits are indeed functional *in vivo*. The active site of this enzyme is built from the two largest subunits. Using site directed mutagenesis we were able to modify the active site of the yeast RNA polymerase II so as to introduce *Plasmodium* or human structural motifs. The resulting strains allow the screening of chemical libraries for potential specific inhibitors.

© 2011 Elsevier B.V. All rights reserved.

Plasmodium falciparum (Pf) is causing malaria [1]. Several drugs exhibit antimalarial activity that target a limited set of enzymatic activities (recently among them the artemisinin derivatives). The parasitic genome was sequenced [2]. The RNA polymerase II (RNAPII) is a complex of 12 subunits and transcribes most of the nuclear genome in eukaryotic cells. The genes encoding its subunits are named RPB1–12 in the yeast *Saccharomyces cerevisiae* (Sc), and will be henceforth referred to as Sc1–Sc12. It is the natural target of alpha-amanitin [3], which blocks RNAPII elongation leading to its degradation *in vivo* [4], ultimately resulting in cell death. Both human (Hs) and Pf RNAPII are sensitive to alpha-amanitin [5]. Therefore, a drug that would selectively inhibit the parasite's enzyme might be of therapeutic value.

In this study, we address the functional relevance of the homology-based candidates for Pf RNAPII subunit coding sequences in Sc. The potential subunits were identified in databases (see Table 1 and Sup Table 1). The coding sequences of the 10 smallest subunits were cloned into a pGEN vector allowing their expression under the control of a PGK promoter (see Sup Table 2A). The ability of putative RNAPII subunits of Pf to replace their Sc counterparts was tested using complementation assays. The corresponding Hs subunits were tested in parallel as controls [6,7]. The complementation assay was performed by transforming each yeast strain, defective for a particular RNAPII subunit gene, with a pGEN

derivative allowing the expression of either the same autologous subunit, or its putative Pf and Hs heterologous homologs (for details see Sup. Data, Materials and Methods). The yeast strains are listed in Sup Table 2B.

Pf4 can replace Sc4 *in vivo*. The DD105 yeast strain lacking Sc4 was viable, although growing very slowly (–), (Fig. 1A). When transformed with pGEN Sc4, normal growth was restored (Sc4, Fig. 1A). The expression of both Pf4 and Hs4 efficiently replaced Sc4 as judged from the size of the colonies (Pf4 and Hs4, Fig. 1A) at 30 °C on rich medium. The same efficiencies were observed at 20 °C, while, at 37 °C, neither Pf4 nor Hs4 could replace Sc4 (see Sup. Fig. 2).

Pf5 can replace Sc5 *in vivo*. Sc5 is an essential protein, as confirmed by the inability of the YSF11 strain (see Sup. Table 2B), bearing an interrupted chromosomal RPB5 gene, to grow in its absence (not shown [8]). When transformed with pGEN Sc5, this strain grows (Sc5, Fig. 1B). The expression of Pf5 but not Hs5 in place of Sc5 yielded viable complemented yeast (Sup. Fig. 1: red colonies in Pf5), which showed a reduced growth rate at 30 °C on rich medium (Pf5, Fig. 1B). By contrast, at 20 °C, a nearly WT phenotype was observed, while at 37 °C, Pf5 was essentially inactive (see Sup. Fig. 3).

Pf7 can replace Sc7 *in vivo*. As Sc7 is an essential protein, the D208-4b strain, bearing an interrupted chromosomal RPB7 gene (see Sup. Table 2B), cannot grow in its absence (not shown [9]). D208-4b yeast strain grows when Sc7 expression was restored (Sc7, Fig. 1C). Hs7 replaced Sc7 at 30 °C on rich medium (Hs7, Fig. 1C), Pf7 did not restore a completely wild-type phenotype under the same conditions (Pf7, Fig. 1C). At 20 °C, the difference in colony sizes was

* Corresponding author. Tel.: +33 368 854 782; fax: +33 368 854 770.
 E-mail address: marc.vigneron@unistra.fr (M. Vigneron).

Table 1
Complementation results for Pf and Hs RNAPII subunits.

Subunit	PlasmoDB: ID N	Identity %	Complementation	Subunit	Identity %	Complementation
Pf3	PFI1130c	21	–	Hs3	34	–
Pf4	PFB0245c	16	+	Hs4	19	+
Pf5	PF13_0341	39	+	Hs5	36	–
Pf6	PFC0155c	34	–	Hs6	47	+
Pf7	PF10_0269	33	+	Hs7	43	+
Pf8	PFL0665c	29	–	Hs8	30	+
Pf9	PFA0505c	14	+	Hs9	37	+
Pf10	PF07_0027	51	–	Hs10	71	+
Pf11	PF13_0023	28	–	Hs11a	47	–
Pf12	MAL13P1.213	28	+	Hs12	34	+

considerably reduced (see Sc7, Pf7 and Hs7 Sup. Fig. 4). By contrast, at 37 °C, Pf7 was essentially inactive, although both Sc7 and Hs7 similarly supported yeast growth (see Sup. Fig. 4).

Pf9 can replace Sc9 *in vivo*. YVV9 (lacking Sc9 expression) is a viable, slow-growing strain (–), (Fig. 1D). When transformed with pGEN Sc9, the yeast cells grew much faster (Sc9, Fig. 1D). The expression of Pf9 and Hs9 efficiently replaced their yeast counterpart (Pf9 and Hs9, Fig. 1D) at 30 °C on rich medium. The same was true when growth took place at 20 or 37 °C (see Sup. Fig. 5).

Pf12 can replace Sc12 *in vivo*. As Sc12 is also an essential protein, the YGVSO20 strain, that bears an interrupted chromosomal RPB12 gene (see Sup. Table 2B), cannot grow in its absence (not shown [10]). This yeast strain grows when Sc12 expression is restored (Sc12, Fig. 1E). Both Pf12 and Hs12 replace Sc12 at 30 °C on rich medium (Pf12 and Hs12, Fig. 1E). At 20 °C, there was no difference in colony size (see Sc12, Pf12 and Hs12 Sup. Fig. 6). At 37 °C, neither Pf12 nor Hs12 was fully active, as reflected by the smaller colony size (see Sup. Fig. 6).

Pf3, Pf6, Pf8, Pf10, Pf11 subunits are not able to replace the corresponding Sc subunits. The complementation assays performed with the pGEN vectors expressing Pf3, Pf6, Pf8, Pf11, using the D138-

1, JAY444, YGVSO43, YGVSO72 strains, respectively, did not yield viable colonies (data not shown). The presence of pGEN-encoded mRNAs was verified by RT-PCR, and in all cases the expected transcripts were detected (see Sup. Fig. 7). The assay in the case of Pf10 was the genetic drift. Loosing the balancer plasmid that carries both the ADE2 and the RPB10 genes, yields viable red colonies, provided the Sc10 protein can be replaced by the pGEN encoded candidate (for example Hs10, see Sup. Fig. 1). We analyzed any small colony that may have failed to turn red, and had to finally conclude that Pf10 was not functional in yeast.

We summarize our results in Table 1. We have demonstrated that five of these proteins are functional in the yeast cell: Pf4, Pf5, Pf7, Pf9 and Pf12. This is the first formal evidence that these proteins are *bona fide* functional polymerase subunits. Pf9 is the most functional subunit since even at 37 °C no phenotype was observed. Although dispensable for growth under optimal conditions, the Sc9 subunit is very important for elongation in extreme conditions [11]. When compared to Sc9, the Pf9 protein exhibits a 120-residue C-terminal extension that does not impair its functions in yeast (see Sup. Fig. 5). The homology score (14%, Table 1) is low, demonstrating a poor correlation between this indication and the functionality

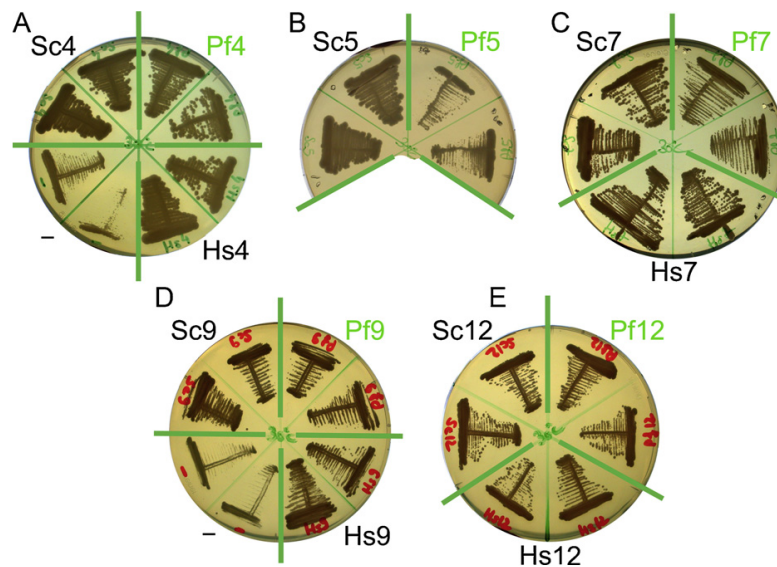


Fig. 1. Five *Plasmodium* subunits are functional in yeast. Yeast strains were either transformed with a pGEN vector allowing the expression of a given subunit, as indicated on the corresponding sector of each Petri dish. (–) Refers to the cases where the subunits are not essential. Two independent viable colonies were streaked on YPGlu rich medium and allowed to grow at 30 °C for two to three days. (A) DD105 strain was transformed or not (–) with pGEN Sc4, pGEN Pf4, pGEN Hs4. (B) YSF11 strain was transformed with pGEN Sc5, pGEN Pf5. (C) D208-4b was transformed with pGEN Sc7, pGEN Pf7, pGEN Hs7. (D) YVV9 strain was transformed or not (–) with pGEN Sc9, pGEN Pf9, pGEN Hs9. (E) YGVSO20 was transformed with pGEN Sc12, pGEN Pf12, pGEN Hs12.

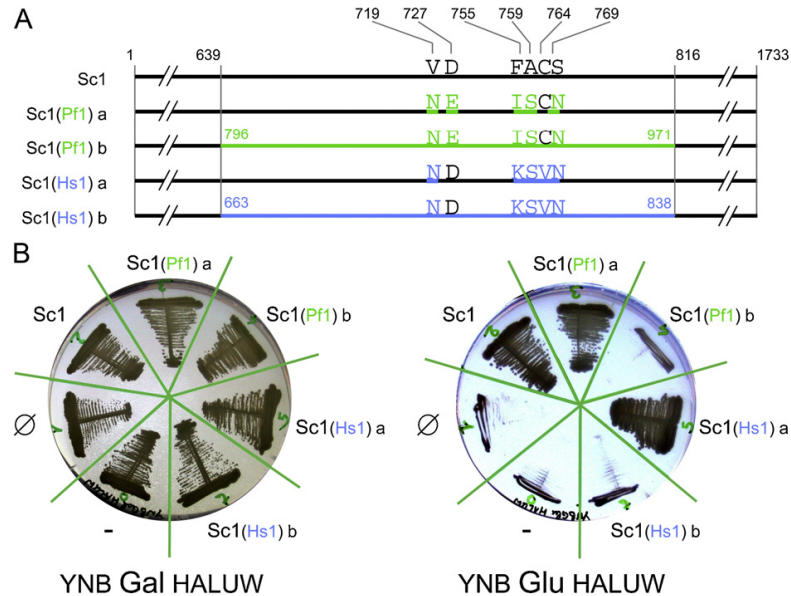


Fig. 2. RNAPII active center modification. (A) Schematic representation of the Sc1 subunit derivatives that were tested. The line on the top represents the wild-type peptidic sequence (Sc1), where the six residues shown with their coordinates have been mutated as indicated below in colours. In Sc1(Pf1)a five residues were mutated. Similarly, five residues were mutated in Sc1(Hs1)a. Sc1(Pf1)b is a chimeric molecule in which a segment spanning residues 640–815 was replaced by a segment spanning the orthologous residues 796–971 from the Pf1 sequence. In Sc1(Hs1)b the same sequence was replaced by residues 663–838 from Hs1. (B) The JAY47 yeast strain was transformed with a pJAY483 expression vector, either empty (\emptyset) or allowing the expression of a Sc1 derivative, as indicated on the sectors of each Petri dish. The untransformed JAY47 was streaked as well (–). A representative colony was streaked on both a supplemented minimal medium as indicated (His, Ade, Leu, Ura, Trp) with either galactose (Gal) or glucose (Glu) as a fermentable carbon source. The yeast cells were allowed to grow at 30°C for four days.

of a protein *in vivo*. Pf4 and Pf7 are fully functional RNAPII subunits in yeast at 20°C. Sc4 and Sc7 form a dimer that is bound to the minimal RNAPII complex, mainly via Sc4, and are involved in mRNA trafficking [12]. Pf4 is poorly conserved (see Table 1 or Sup. Fig. 2), though efficient *in vivo*. Hs4 complemented strain exhibits a strong phenotype at 37°C (see Sup. Fig. 2), which is in contrast with a previously published result [6]. This discrepancy is most probably related to the fact that the incubation temperature was more carefully controlled in the present work, arousing more confidence in the present result, in agreement with others [13]. Pf5 and Pf12 can replace their Sc homologs, and are present within the three types of nuclear RNA polymerase: RNAPI, RNAPII and RNAPIII. The three polymerase activities are simultaneously assayed. The small Pf12 subunit appears to be functionally well conserved, since only at 37°C is a phenotype observed (see Sup. Fig. 6). It can be noticed that Hs12 exhibits a moderate phenotype at 37°C. Strikingly, Pf5 is fully functional at 20°C, but very sensitive to enhanced temperatures (see Sup. Fig. 3). This seems to be a common property, since Pf4, Pf5, Pf7, Pf9 and Pf12 all exhibit nearly wild-type growth at 20°C, in sharp contrast with their respective behaviour at higher temperatures.

Though Pf4 and Pf9 exhibit homology scores as low as 15%, they are functional in distantly related species. By contrast, Pf10 and Pf6, respectively, 51% and 36% homologous to their yeast counterparts are not functional. We investigated the reason behind this observation (data not shown). There are two domains that differ between Sc10 and Pf10 (residues 17–42 and 52–70) (see Sup. Fig. 9). We created chimeric Sc10 proteins in which we exchanged independently both domains. When assayed in complementation, none of these recombinant was viable. At least two major changes are responsible for the lack of function of the Pf10 protein in yeast.

In the case of Pf6, we noticed a 15-residue peptidic insertion within a very conserved domain (Pf6 114–128, Sup. Fig. 10). We

inserted it within the Hs6 protein at the same place. When assayed, the resulting recombinant turned out to be viable, although with a thermosensitive phenotype (data not shown). The presence of this insert within the Pf6 subunit does not explain why it does not work in yeast. It may be worth stressing that Sc5, 6, 8, 10 and 12 are shared by RNAPI, RNAPII and RNAPIII, so that the lack of viability of complemented strains may be due to only one polymerase activity defect.

We were able to amplify both Pf1 and Pf2 coding sequences, but not to clone them in any vector, preventing any complementation test. We decided to focus on the alpha-amanitin binding domain within the RNAPII active center that is mainly located on the largest subunit. We thus constructed the recombinant Sc1 derivatives described in Fig. 2A. The Sc1 residues that are in close contact with the toxin [14] were mutated when different to those present at the same position in either the *Plasmodium* or the human sequence, in order to locally mimic either the *Plasmodium* parasite or its human host RNAPII, generating Sc1(Pf1)a and Sc1(Hs1)a, respectively. Two chimeras were built using unique restriction sites flanking this region (see Fig. 2A and Sup. Fig. 11). A 176-residue sequence of Sc1 was replaced by either the homologous *Plasmodium* or the human sequence, generating Sc1(Pf1)b and Sc1(Hs1)b, respectively.

The ability of these constructs to replace the wild-type Sc1 was tested in the JAY47 strain, under glucose mediated repression of the chromosomal GAL10-RPB1 chimeric gene as shown in Fig. 2B (see Sup. Data, Materials and Methods). Representative colonies of each transformation were streaked on supplemented galactose medium and all are viable (YNB Gal HALUW, Fig. 2B). On glucose medium, very different growths were observed (YNB Glu HALUW, Fig. 2B). The JAY47 strain, whether transformed or not with a pJAY483Δ empty vector (– and \emptyset , Fig. 2B), could not grow. The expression of Sc1 restores growth (Sc1, Fig. 2B). The mutation of five residues

within the RNAPII active center [Sc1(Pf1)a and Sc1(Hs1)a, Fig. 2B] did not prevent growth on glucose medium. In contrast, the swapping of the 176-residue peptide yielded non functional molecules [Sc1(Pf1)b and Sc1(Hs1)b, Fig. 2B] (see also Sup. Fig. 8). This does not formally rule out that the full-length Pf1 or Hs1 proteins may be functional, but strongly argues against this possibility. The edges of the swapped peptide are located in closely homologous regions and only substitutions can be responsible for the phenotype (see Sup. Fig. 11). When assaying the full length Pf1 and Hs1 proteins, a large number of additional changes are likely to add to the phenotype.

The alpha-amanitin binding domain of the active center of RNAPII is shared by both largest subunits, Sc1 and Sc2. The 3D structure of the yeast RNAPII complexed with amanitin has been resolved [14]. Most of the residues involved in the binding of the toxin are located within the largest subunit, in a domain previously genetically identified [15]. Both Sc1(Pf1)a and Sc1(Hs1)a constructs are viable, although Sc1(Pf1)a exhibited a slightly decreased growth rate (data not shown). The side chains of the six residues pointing to the lumen of the active site, this is not unexpected and confirms the validity of the established 3D model and genetic data. The availability of three yeast strains differing only by a few residues (Sc1(Pf1)a and Sc1(Hs1)a and the Sc1 control) within the RNAPII active site will potentially be useful for screening for inhibitors.

This work represents to our knowledge the first attempt to genetically characterize the RNAPII of *Pf*. In addition, we constructed viable mutated enzymes that exhibit modified active sites within the alpha-amanitin binding region. The resulting yeast strains may prove useful to design convenient tests for the identification of new potential antimalarial drugs.

Acknowledgements

We thank Michel Riva and Pierre Thuriaux for gift of materials, Esther Kellenberger, Etienne Weiss, Youra Shpakovski and Marcel Hibert for stimulating discussions.

This work was supported by funds and/or fellowships from the Université de Strasbourg, the Centre National de la Recherche Scientifique, the Ministère de l'Enseignement Supérieur et de la Recherche, the Ligue contre le Cancer – Comités Haut-Rhin et Bas-Rhin, Alsace contre le Cancer, the Ecos-Sud (action A06B02) and the Institut National du Cancer (PLBIO-127).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2010.12.013.

References

- [1] Hay SI, Guerra CA, Gething PW, et al. A World Malaria Map: Plasmodium falciparum endemicity in 2007. *Plos Med* 2009;6:e48.
- [2] Gardner MJ, Hall N, Fung E, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 2002;419:498–511.
- [3] Kedinger C, Gniazdowski M, Mandel J, Gissinger F, Chambon P. Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. *Biochem Biophys Res Commun* 1970;38:165–71.
- [4] Nguyen VT, Giannoni F, Dubois MF, et al. *In vivo* degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. *Nucleic Acids Res* 1996;24:2924–9.
- [5] Militello KT, Patel V, Chessler A-D, et al. RNA polymerase II synthesizes anti-sense RNA in Plasmodium falciparum. *RNA* 2005;11:365–70.
- [6] Schaller S, Grandemange S, Shpakovski GV, Golemis EA, Kedinger C, Vigneron M. Interactions between the full complement of human RNA polymerase II subunits. *FEBS Lett* 1999;461:253–7.
- [7] Shpakovski GV, Acker J, Wintzerith M, Lacroix JF, Thuriaux P, Vigneron M. Four subunits that are shared by the three classes of RNA polymerase are functionally interchangeable between Homo sapiens and Saccharomyces cerevisiae. *Mol Cell Biol* 1995;15:4702–10.
- [8] Woychik NA, Liao SM, Kolodziej PA, Young RA. Subunits shared by eukaryotic nuclear RNA polymerases. *Genes Dev* 1990;4:313–23.
- [9] McKune K, Richards K, Edwards AM, Young RA, Woychik NA. RPB7, one of two dissociable subunits of yeast RNA polymerase II, is essential for cell viability. *Yeast* 1993;9:295–9.
- [10] Treich I, Carles C, Riva M, Sentenac A. RPC10 encodes a new mini subunit shared by yeast nuclear RNA polymerases. *Gene Expr* 1992;2:31–7.
- [11] Van Mullem V, Wery M, Werner M, Vandehaute J, Thuriaux P. The Rpb9 subunit of RNA polymerase II binds transcription factor TFIIE and interferes with the SAGA and elongator histone acetyltransferases. *J Biol Chem* 2002;277:10220–5.
- [12] Choder M. Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. *Trends Biochem Sci* 2004;29:674–81.
- [13] Khazak V, Estojak J, Cho H, et al. Analysis of the interaction of the novel RNA polymerase II (pol II) subunit hsRPB4 with its partner hsRPB7 and with pol II. *Mol Cell Biol* 1998;18:1935–45.
- [14] Bushnell DA, Cramer P, Kornberg RD. Structural basis of transcription: alpha-amanitin-RNA polymerase II cocystal at 2.8 Å resolution. *Proc Natl Acad Sci USA* 2002;99:1218–22.
- [15] Bartolomei MS, Corden JL. Localization of an alpha-amanitin resistance mutation in the gene encoding the largest subunit of mouse RNA polymerase II. *Mol Cell Biol* 1987;7:586–94.

Supplementary Data

Material and Methods.

1. DNA modifying enzymes. Restriction enzymes, ligase, Pfu polymerase, AMV RTase, Taq polymerase were purchased from Fermentas and used according to the supplier's instructions.
2. RNAPII subunits cloning. The various proteins that were studied are listed in Sup Table 1. All *Sc* and *Hs* subunits were cloned previously (see [6] and ref therein). *Pf* RNAPII subunits P13, P110 and P111 coding sequences were obtained by RT-PCR from total RNAs extracted from 3D7 *Plasmodium* strain parasites using the trizol extraction procedure. All others (P14, P15, P16, P17, P18, P19, P12) were reconstituted by synthetic oligonucleotide mediated ligation. In all cases, the codon usage was adapted to the yeast background and flanked with unique restriction sites (AvrII, SpeI, NheI or XbaI). All coding sequences were cloned into either a pJET or pBS-SK vectors, and their sequences established. The two largest subunits (P11, P12) were obtained using Takara Ex Taq amplification from genomic DNA. However, we did not succeed in cloning their complete sequences into any of the vectors used.
3. Yeast plasmids. The various yeast expression plasmids that have been used in the present work are listed in Sup Table 2A. The pJAY483 vector was used for the expression of native *Sc*1 and *Sc*1 derivatives. Pfu-mediated amplification was used to delete the complete *Sc*1 coding sequence of pJAY483 *Sc*1, yielding an empty derivative, pJAY483Δ. The coding sequence of *Sc*1 was modified using oligonucleotide primed Pfu amplification of the pJAY483 *Sc*1 plasmid, yielding pJAY483 *Sc*1(P11)a and pJAY483 *Sc*1(Hs1)a point-mutated sequences (see Table 2A). The pJAY483 *Sc*1(P11)b and pJAY483 *Sc*1(Hs1)b chimeric coding sequences were constructed as follows: the P11 and Hs1 coding sequence fragments were amplified with appropriate oligonucleotides bringing PvuII and XcmI restriction sites, respectively at the 5' and 3' ends of the amplified fragment. They were then cloned into a pJET vector and their sequence was established. The P11 and Hs1 PvuII-XcmI restriction fragments were then ligated into the corresponding sites of the pJAY483 vector, resulting in the plasmids pJAY483 *Sc*1(P11)b and pJAY483 *Sc*1(Hs1)b, respectively. All other complementations were performed using pGEN derivatives. All recombinants allowing the expression of human and yeast subunits were constructed previously and as briefly described in Sup Table 2A ([6],[7]). The coding sequences of the subunits P13 to P12 were similarly inserted into the yeast expression pGEN vector digested with NheI, using their respective flanking restriction sites (see Sup Table 2A).
4. Yeast strains. The various *Sc* strains that have been used in the present work are listed in Sup Table 2B, and have been described as indicated ([16],[17],[18],[19],[20]). They were grown on the appropriate media, usually YPGlu (Yeast extract 1%, Bactopeptone 2%, glucose 2%, agar 1.5%), except for the JAY47 strain, that was grown on YPGal (Yeast extract 1%, Bactopeptone 2%, galactose 2%, agar 1.5%) medium. Alternatively, minimal medium (YNB 0.67%, either glucose or galactose 2%, agar 1.5%) was used, supplemented with the appropriate metabolites: His 50 µg/ml, Adenine 20 µg/ml, Leu 60 µg/ml, Uracile 30 µg/ml, Trp 50 µg/ml, depending on the strain's genotype.
5. Yeast electroporation. The parental yeast strains were transformed by electroporation with either the pJAY483 derivatives in the case of JAY47, or the appropriate pGEN derivatives in the case of all others. The yeast strains were streaked on rich medium and allowed to grow at 30°C for three days. About 20 µl of packed yeasts cell volume per transformation were scraped from the Petri dish in a sterile way and resuspended in 1M Sorbitol, 20 mM HEPES pH 7.5. The cells were washed twice in the same buffer and centrifuged 2 min at 2300 g at room temperature. After resuspension in the same buffer, 250 ng of plasmid were added just prior electroporation. A single pulse at 1.6 kV, 200 Ω, 250 µF was applied using a Biorad GenePulser. The cells were then spread onto Petri dishes containing supplemented minimal medium lacking the Trp, so that only the cells harbouring the TRP1 carrying plasmids can grow. The JAY47 were maintained on galactose since they cannot grow on glucose.
6. Complementation assay. Four cases can be distinguished depending on the genotype of the parental strains.
 - The YVV9 RPB9-deleted yeast strain is viable and lacks any balancer plasmid, so that the transformants were directly tested for growth efficiency in the appropriate conditions.
 - Genetic drift. Strains YFS11 and YGVS017 are auxotroph for adenine (ade2), but contain a balancer plasmid that carries the wild type ADE2 gene. The TRP1 transformants derived from these strains were streaked onto YPGlu Petri dishes and allowed to grow at 30°C. After adenine shortage, the yeast colonies lacking the ADE2 gene turn red (see Sup. Fig. 1). Independent colonies were picked and streaked on different media, so as to ensure that they have the expected genotypes. They were then streaked on YPGlu dishes and tested for growth at various temperatures.
 - Plasmid shuffling. Most yeast strains (DI38-1d, DD105, JAY444, D208-4b, YGVS043, YGVS072, YGVS020) are auxotroph for uracile (ura3) and contain a balancer plasmid carrying the URA3 gene. The enzymatic activity encoded by URA3 is toxic in the presence of Fluoro Orotic Acid. The TRP1 transformants derived from these strains were streaked onto Petri dishes containing supplemented minimal medium including Uracile (12 µg/ml) and Fluoro Orotic Acid (1 mg/ml) and allowed to grow at 30°C. Only yeast cells lacking a functional URA3 gene can grow under these conditions. Independent colonies were picked and streaked on different media, so as to ensure that they have the expected genotypes. They are then streaked on YPGlu dishes and tested for growth at various temperatures.
 - Gal10 promoter mediated control of *Sc*1 expression. The JAY47 strain can only grow on galactose since it was engineered so that the synthesis of *Sc*1 encoding mRNA is under the control of a GAL10 promoter. This promoter is active on galactose, but completely repressed on glucose. The pJAY483 plasmid and its derivatives contain the full RPB1 gene allowing the expression of *Sc*1 or modified proteins on glucose-containing medium. The phenotype of the transformed JAY47 was therefore assayed on glucose-containing medium.

Yeast				Human				Plasmodium					
Subunit	MW (kDa)	NCBI ID N°		Subunit	MW (kDa)	NCBI ID N°	Identity %	Compl.	Subunit	MW (kDa)	PlasmODB: ID N°	Identity %	Compl.
Sc1	192	NP_010141		Hs1	217	NP_000928	48	ND	Pf1	279	PF00805w	30	ND
Sc2	139	NP_014794		Hs2	134	NP_000929	56	-	Pf2	167	PF00715w	45	ND
Sc3	35	NP_012243		Hs3	31	NP_116558	34	-	Pf3	38	PF11130c	21	-
Sc4	25	NP_012395		Hs4	16	NP_004796	19	+	Pf4	15	PF00245c	16	+
Sc5*	25	NP_009712		Hs5	25	NP_002686	36	-	Pf5	24	PF13_0341	39	+
Sc6*	18	NP_015513		Hs6	15	NP_068809	47	+	Pf6	18	PF00155c	34	-
Sc7	19	NP_010692		Hs7	19	NP_002687	43	+	Pf7	20	PF10_0269	33	+
Sc8*	17	NP_014867		Hs8	17	NP_006223	30	+	Pf8	17	PF10665c	29	-
Sc9	14	NP_011445		Hs9	15	NP_006224	37	+	Pf9	27	PF05050c	14	+
Sc10*	8	NP_014853		Hs10	8	NP_066951	71	+	Pf10	8	PF07_0027	51	-
Sc11	14	NP_014638		Hs11a	13	NP_006225	47	-	Pf11	14	PF13_0023	28	-
Sc12*	8	NP_012013		Hs12	7	NP_005025	34	+	Pf12	7	MAL13P1.213	28	+

* : subunits shared by RNAPI , II and III in yeast.

Hazoume Sup. Table 1 : Accession N° and Complementation results.

Yeast plasmids

Name	Non-bacterial genes and cDNAs	Construction
pJAY483 Sc1	ARSCEN6, RPB1, TRP1	pJAY483 [6].
pJAY483 A	ARSCEN6, TRP1	created by deleting Sc1 CDS using Pfu mediated amplification.
pJAY483 Sc1(pT1)a	ARSCEN6, RPB1, TRP1	created by site-directed mutagenesis of RPB1 coding sequence residues V715N, D727E, F755I, A759S, S769N, within pJAY483.
pJAY483 Sc1(pT1)b	ARSCEN6, RPB1, TRP1	created by inserting a PvuII-XcmI engineered fragment of P11 CDS (798-971) into the PvuII-XcmI digested pJAY483.
pJAY483 Sc1(Hs1)a	ARSCEN6, RPB1, TRP1	created by site-directed mutagenesis of RPB1 coding sequence residues V715N, F756K, A759S, C764V, S769N, within pJAY483.
pJAY483 Sc1(Hs1)b	ARSCEN6, RPB1, TRP1	created by inserting a PvuII-XcmI engineered fragment of Hs1 CDS (863-838) into the PvuII-XcmI digested pJAY483.
PGEN	2u TRP1 pPGK	created by inserting AATTCGCTAGCAGCTAGTCTAGAGTCTAGA into Eco RI of pYPGE2 to generate unique EcoRI, NheI and AvrII sites [7].
PGEN Sc3	2u TRP1 pPGK Sc3	created by inserting Sc3 full length CDS (1-318) PCR generated NheI-AvrII fragment into the NheI digested PGEN.
PGEN P13	2u TRP1 pPGK P13	created by inserting P13 full length CDS (1-335) RT-PCR generated NheI-AvrII fragment into the NheI digested PGEN.
PGEN Hs3	2u TRP1 pPGK Hs3	created by inserting Hs3 full length CDS (1-275) PCR generated NheI fragment into the NheI digested PGEN.
PGEN Sc4	2u TRP1 pPGK Sc4	created by inserting Sc4 full length CDS (1-221) PCR generated NheI-AvrII fragment into the NheI digested PGEN.
PGEN P14	2u TRP1 pPGK P14	created by inserting P14 full length CDS (1-132) oligonucleotide-reconstituted NheI-AvrII fragment into the NheI digested PGEN.
PGEN Hs4	2u TRP1 pPGK Hs4	created by inserting Hs4 full length CDS (1-142) PCR generated NheI-SpeI fragment into the AvrII digested PGEN.
PGEN Sc5	2u TRP1 pPGK Sc5	created by inserting Sc5 full length CDS (1-215) PCR generated NheI fragment into the NheI digested PGEN.
PGEN P15	2u TRP1 pPGK P15	created by inserting P15 full length CDS (1-205) oligonucleotide-reconstituted NheI-AvrII fragment into the NheI digested PGEN.
PGEN Hs5	2u TRP1 pPGK Hs5	created by inserting Hs5 full length CDS (1-210) PCR generated NheI fragment into the NheI digested PGEN.
PGEN Sc6	2u TRP1 pPGK Sc6	created by inserting Sc6 full length CDS (1-155) PCR generated NheI fragment into the NheI digested PGEN.
PGEN P16	2u TRP1 pPGK P16	created by inserting P16 full length CDS (1-165) oligonucleotide-reconstituted NheI-AvrII fragment into the NheI digested PGEN.
PGEN Hs6	2u TRP1 pPGK Hs6	created by inserting Hs6 full length CDS (1-127) PCR generated NheI-SpeI fragment into the NheI digested PGEN.
PGEN Sc7	2u TRP1 pPGK Sc7	created by inserting Sc7 full length CDS (1-171) PCR generated NheI-AvrII fragment into the NheI digested PGEN.
PGEN P17	2u TRP1 pPGK P17	created by inserting P17 full length CDS (1-177) oligonucleotide-reconstituted NheI-AvrII fragment into the NheI digested PGEN.
PGEN Hs7	2u TRP1 pPGK Hs7	created by inserting Hs7 full length CDS (1-172) PCR generated NheI-XbaI fragment into the NheI digested PGEN.
PGEN Sc8	2u TRP1 pPGK Sc8	created by inserting Sc8 full length CDS (1-146) PCR generated NheI fragment into the NheI digested PGEN.
PGEN P18	2u TRP1 pPGK P18	created by inserting P18 full length CDS (1-145) oligonucleotide-reconstituted NheI-AvrII fragment into the NheI digested PGEN.
PGEN Hs8	2u TRP1 pPGK Hs8	created by inserting Hs8 full length CDS (1-150) PCR generated SpeI-XbaI fragment into the NheI digested PGEN.
PGEN Sc9	2u TRP1 pPGK Sc9	created by inserting Sc9 full length CDS (1-122) PCR generated NheI-AvrII fragment into the NheI digested PGEN.
PGEN P18	2u TRP1 pPGK P18	created by inserting P18 full length CDS (1-249) oligonucleotide-reconstituted generated AvrII fragment into the NheI digested PGEN.
PGEN Hs9	2u TRP1 pPGK Hs9	created by inserting Hs9 full length CDS (1-125) PCR generated NheI fragment into the NheI digested PGEN.
PGEN Sc10	2u TRP1 pPGK Sc10	created by inserting Sc10 full length CDS (1-70) PCR generated NheI fragment into the NheI digested PGEN.
PGEN P10	2u TRP1 pPGK P10	created by inserting P10 full length CDS (1-69) RT-PCR generated NheI-AvrII fragment into the NheI digested PGEN.
PGEN Hs10	2u TRP1 pPGK Hs10	created by inserting Hs10 full length CDS (1-67) PCR generated NheI-SpeI fragment into the NheI digested PGEN.
PGEN Sc11	2u TRP1 pPGK Sc11	created by inserting Sc11 full length CDS (1-120) PCR generated NheI fragment into the NheI digested PGEN.
PGEN P11	2u TRP1 pPGK P11	created by inserting P11 full length CDS (1-126) RT-PCR generated NheI-AvrII fragment into the NheI digested PGEN.
PGEN Hs11a	2u TRP1 pPGK Hs11a	created by inserting Hs11a full length CDS (1-117) PCR generated NheI fragment into the NheI digested PGEN.
PGEN Sc12	2u TRP1 pPGK Sc12	created by inserting Sc12 full length CDS (1-70) PCR generated NheI-AvrII fragment into the NheI digested PGEN.
PGEN P12	2u TRP1 pPGK P12	created by inserting P12 full length CDS (1-58) oligonucleotide-reconstituted NheI-AvrII fragment into the NheI digested PGEN.
PGEN Hs12	2u TRP1 pPGK Hs12	created by inserting Hs12 full length CDS (1-58) PCR generated NheI-AvrII fragment into the NheI digested PGEN.

Site directed point mutations of the coding sequences inserted in the various plasmids were introduced using the appropriate oligonucleotides (information available on request) and the Pfu mediated amplification.

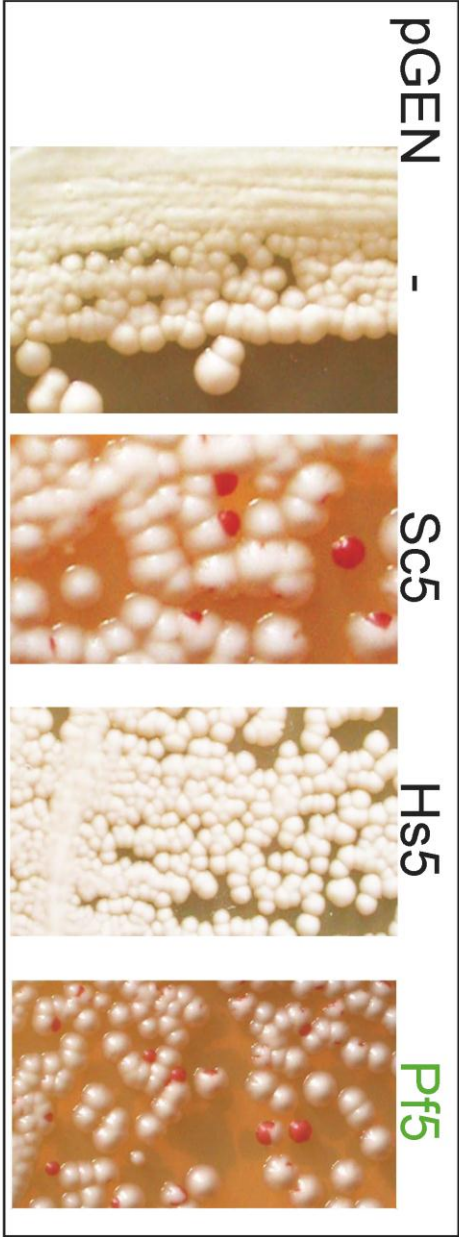
Yeast strains

Name	Tested subunit	Gene	Genotype	Origin
JAY47	Sc1	RPB1	MAT α , <i>ade2-1, Can1-100, his3-1115, leu2-3 112, tpt1-1, ura3-1, pGAL10-RPO21</i> (pUAY483 Sc1(PH)a, TRP1, RPB1)	[16] (present work)
AH01	Sc1	RPB1	MAT α , <i>ade2-1, Can1-100, his3-1115, leu2-3 112, tpt1-1, ura3-1, pGAL10-RPO21</i> (pUAY483 Sc1(Hs1)a, TRP1, RPB1)	(present work)
AH03	Sc1	RPB1	MAT α , <i>ade2-1, Can1-100, his3-1115, leu2-3 112, tpt1-1, ura3-1, pGAL10-RPO21</i> (pUAY483 Sc1, TRP1, RPB1) reconstructed.	(present work)
AH09	Sc1	RPB1	MAT α , <i>ade2-1, Can1-100, his3-1115, leu2-3 112, tpt1-1, ura3-1, pGAL10-RPO21</i> (p124T3 = <i>URA3</i> , CEN, RPB3)	[17]
D138-1d	Sc3	RPB3	MAT α , <i>ade2-1, his3, leu2-3 112, ysc2-Δ201, tpt1-Δ1, ura3-52, rpb5-Δ1::LYS2</i> (p124T3 = <i>URA3</i> , CEN, RPB3)	(P. Thuriaux, personal gift)
DD105	Sc4	RPB4	MAT α , <i>his3-Δ200, leu2* ysc2-201, tpt1Δ1, ura3-52, rpb4-Δ::kanMX4 (2μ <i>URA3</i> p1P1; RPB4)</i>	(present work)
AH59	Sc5	RPB5	MAT α , <i>his3-Δ200, leu2* ysc2-201, tpt1Δ1, ura3-52, rpb5-Δ::URA3 (pASZ11=CEN6 ARSα ADE2 RPB5)</i>	[18]
YFS11	Sc5	RPB5	MAT α , <i>ade2-101, his3-Δ200, leu2-Δ1 ysc2-801a, tpt1-Δ1, ura3-52, rpb5-Δ::URA3 (pASZ11=CEN6 ARSα ADE2 RPB5)</i>	(present work)
AH53	Sc5	RPB5	MAT α , <i>ade2-101, his3-Δ200, leu2-Δ1, ysc2-801a, tpt1-Δ1, ura3-52, rpb5-Δ::URA3 (pGEN P8; 2μm TRP1 P9)</i>	[19]
JAY444	Sc6	RPB6	MAT α , <i>ade2-1, CAN1-100, his3-1115, leu2-3 112, tpt1-1, ura3-1, rpb6::LEU2/pRP026</i> (CEN <i>URA3</i> RPB6)	[19]
D208-4b	Sc7	RPB7	MAT α , <i>ade2-1, his3-Δ200, leu2-Δ1 ou3 112), ysc2-Δ201, tpt1-Δ1, ura3-52, rpb7-Δ1::LEU2</i> (PM1001 2 μ <i>URA3</i> RPB7)	[20]
AH65	Sc7	RPB7	MAT α , <i>ade2-1, his3-Δ200, leu2-Δ1 ou3 112), ysc2-Δ201, tpt1-Δ1, ura3-52, rpb7-Δ1::LEU2</i> (pGEN P7; 2 μ m TRP1 P7)	(present work)
YGVSD043	Sc8	RPB8	MAT α , <i>ade2-1, his3-Δ200, leu2* ysc2* tpt1-Δ63, ura3-52, rpb8-Δ1::LYS2</i> (pSL103; CEN <i>URA3</i> RPB8)	[19]
YVY9	Sc9	RPB9	MAT α , <i>ade2* his3-Δ200, ysc2-Δ201, leu2-3 112, tpt1-Δ63, ura3-52, rpb9-Δ1::HIS3</i>	[11]
AH74	Sc9	RPB9	MAT α , <i>ade2, his3-Δ200, ysc2-Δ201, leu2-3 112, tpt1-Δ63, ura3-52, rpb9-Δ1::HIS3</i> (pGEN P8; 2 μ m TRP1 P9)	(present work)
YGVSO17	Sc10	RPB10	MAT α , <i>ade2-1, his3-Δ200, leu2 ysc2-801, tpt1-Δ63, ura3-52, rpb10-Δ1::HIS3</i> (pRPB10-S; CEN6 ARS α ADE2 RPB10)	[7]
YGVSO72	Sc11	RPB11	MAT α , <i>ade2-101, his3-Δ200, leu2-3 112, ysc2-Δ201, tpt1-Δ63, ura3-52, rpb11-Δ1::HIS3</i> (pRP11/8 = <i>URA3</i> , ARS1, CEN4, RPB11)	[17]
YGVSD20	Sc12	RPB12	MAT α , <i>ade2-1, his3-Δ200, ysc2-801, tpt1-Δ1 ura3-52, rpb10-Δ::HIS3</i> (pFL44-RPC10; 2 μ m <i>URA3</i> RPC10)	[7]
AH36	Sc12	RPB12	MAT α , <i>ade2-1, his3-Δ200, ysc2-801, tpt1-Δ1 ura3-52, rpb10-Δ::HIS3</i> (pGEN PH2; 2 μ m TRP1 PH2)	(present work)

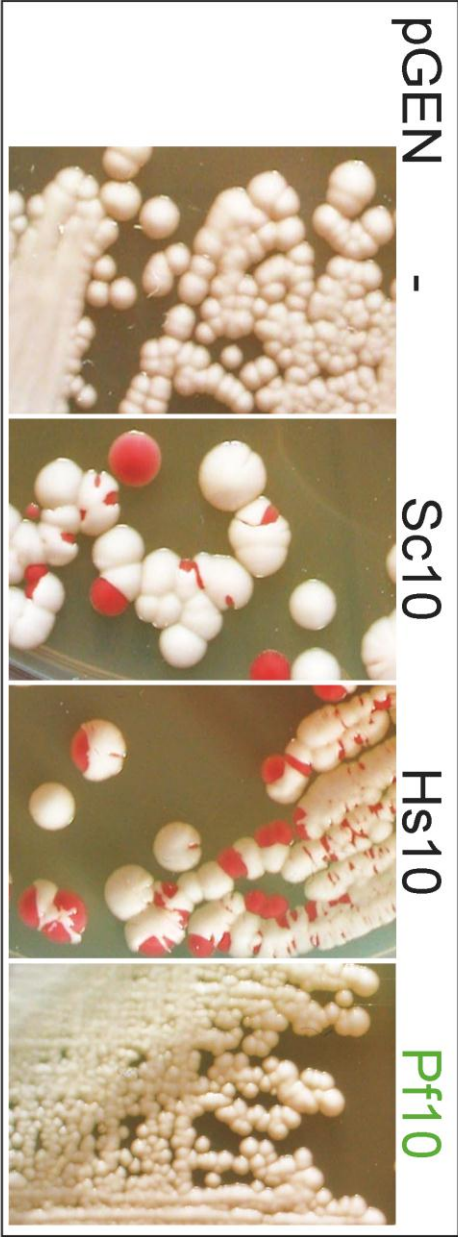
References

- [7] Shpakovski GY, Acker J, Wintzerth M, Lacroix JF, Thuriaux P, Vigneron M. Four subunits that are shared by the three classes of RNA polymerase are functionally interchangeable between Homo sapiens and Saccharomyces cerevisiae. *Molecular and Cellular Biology*. 1996;15:4702-10.
- [16] Archambault J, Debruijn M, Stone J, Friesen JD. Isolation and phenotypic analysis of conditional-lethal, linker-insertion mutations in the gene encoding the largest subunit of RNA polymerase II in Saccharomyces cerevisiae. *Mol Gen Genet*. 1992;232:409-14.
- [17] Berge WJ, Grandemange S, Shpakovski GY, Shematova EK, Kiedinger C, Vigneron M. Distinct regions of RPB11 are required for heterodimerization with RPB3 in human and yeast RNA polymerase II. *Nucleic Acids Research*. 2005;33:3582-90.
- [18] Navarro F, Thuriaux P. In vivo misreading by RNA overdose. *RNA*. 2000;6:103-10.
- [19] Brand JF, Navarro F, Rematier P, Boschetto C, Labarre S, Werner M, et al. Partners of Rpb8p, a small subunit shared by yeast RNA polymerases I, II and III. *Molecular and Cellular Biology*. 2001;21:8056-65.
- [20] Zaros C, Thuriaux P. Rpb25, a conserved RNA polymerase III subunit, is critical for transcription initiation. *Mol Microbiol*. 2005;55:104-14.

YFS11



YGVS017

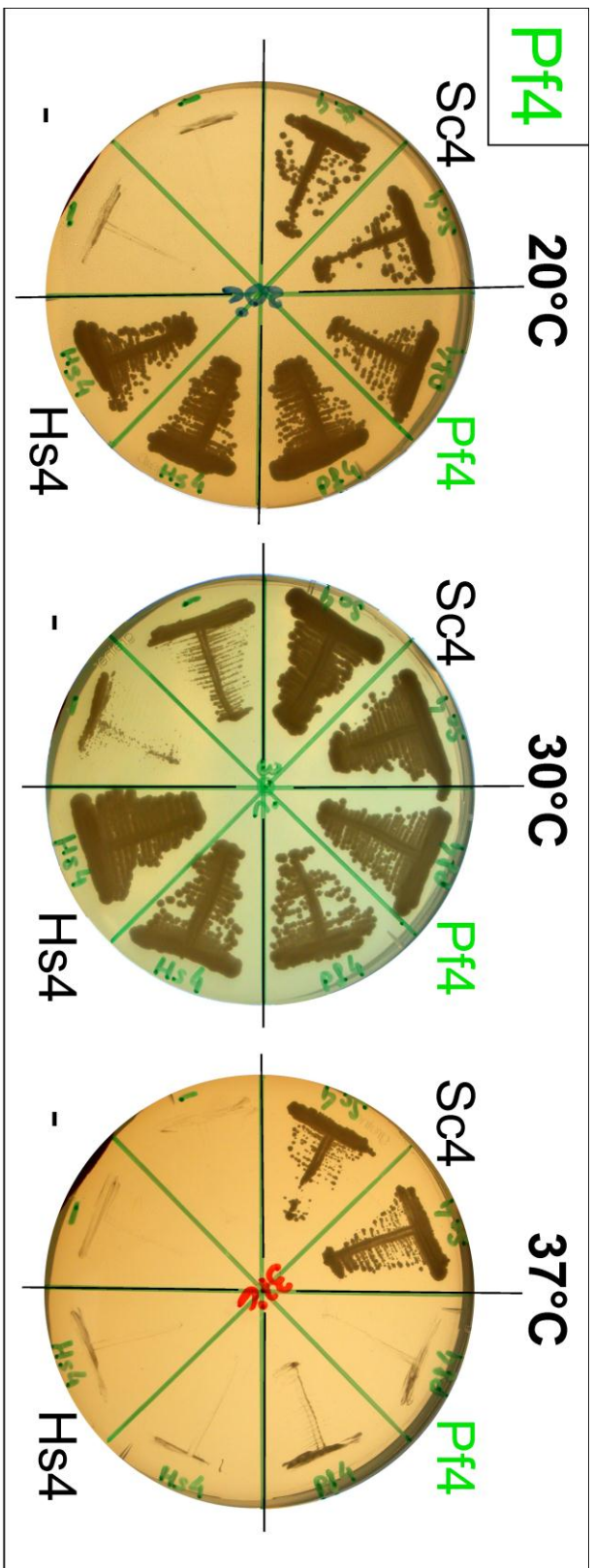


Hazoume Sup. Fig. 1 Genetic drift experiments.

Top panel. YSF11 strain was transformed with the pGEN vector, either empty (-) or allowing the expression of Sc5, Hs5 or Pf5, as indicated. After selection under the appropriate conditions, the clones were streaked and allowed to grow on a YPGlu rich medium at 30°C, for 7 days.

Bottom panel. YGVS017 strain was transformed with the pGEN, either empty (-) or allowing the expression of Sc10, Hs10 or Pf10, as indicated. After selection under the appropriate conditions, the clones were streaked and allowed to grow on a YPGlu rich medium at 30°C, for 7 days.

Pictures were taken at an enlargement of 2 fold. Red cells correspond to complemented yeasts.



Hs4 (1-34) -----MAAGGSDPRRAGDVEEDASQLIFPKFEETAETLLN-----LSEVFMKTLN^YTAR
Sc4 (1-75) MNVSTSTFQTRRRRLKVEEENAAATLQLGQEFQLKQINHQGEEELIALNLS^EEARLVIK^EALVERRA^FKRSQK
Pfl4 (1-26) -----MANNINGDINKLDLGPDFKNC^LCLN-----

Hs4 (35-70) -----SEVHMLLEHRKQONESA^EED^EEQE-----
Sc4 (76-150) KHKKHKLHENANDETTAV^EDEDD^LLEDDV^NADDD^DFMH^SETREK^ELESID^VLL^EQTTG^GNNK^DLK^NT^MQ^YLTN
Pfl4 (27-60) -----CELQLLILGDQLRLT^SKR^NEE-----

Hs4 (71-142) FSRFKNRETIASVRSLL^LLQK-K^LAK^FELAC^LLAN^LCP^ETAE^ESKA^LL^IPS^SLEGR^FEDE^ELQ^QIL^LDD^IIQ^TKRS^FQY
Sc4 (151-221) FSRFRDQETVCAV^IQLL^LKST-GL^HPF^FEVAQ^LGS^LACD^TAD^EAK^TL^IPS^SL^NNK^IS^DDEL^ERL^KEL^SN^LE^TLY-
Pfl4 (61-132) FAAILKNRSSIVDIR^TNLER^ICGD^LHE^VE^IAM^LVN^LLP^KTI^LE^ARY^LIP^SL^IR-L^NDE^TL^NS^IL^LE^HL^ISY^KMY^VS

Hs4	*	Sc4	% Identity =	19.2	(43/223)
Pfl4	*	Sc4	% Identity =	13.5	(30/223)

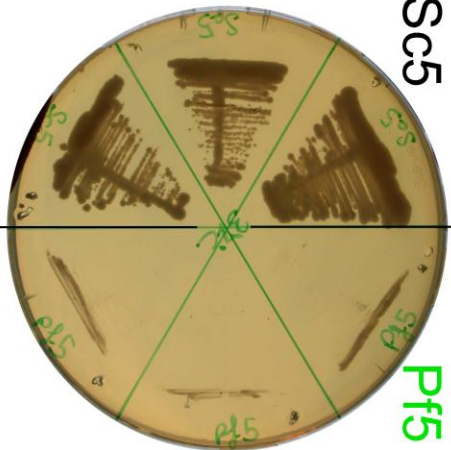
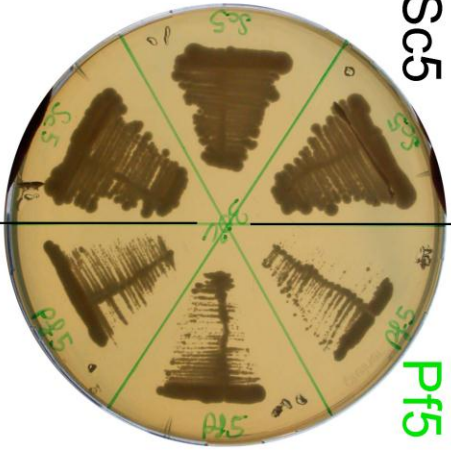
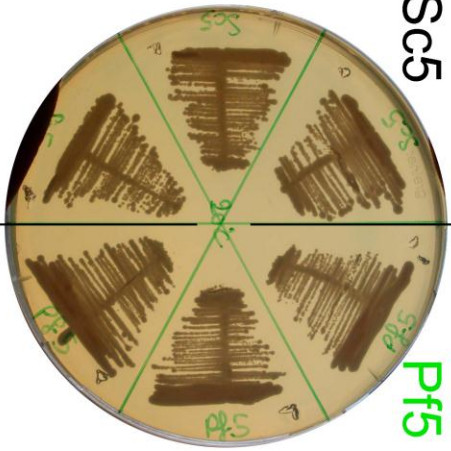
Hazoume Sup. Fig.2 Pfl4 complementation.
 Top : DD105 strain was transformed with the pGEN vector, either empty (-) or allowing the expression of Sc4, Pfl4 or Hs4 as in Figure 1A. Two independent colonies were streaked on three YPGlu rich medium Petri dishes. Each was incubated at either 20, 30 or 37°C, as indicated, for 4, 3 and 4 days, respectively.
 Bottom : Clustal alignment of Hs4, Sc4 and Pfl4 peptidic sequences. The homology was highlighted using the Jalview program (Jalview colour code, conservation 66%). The conservation score (identity percentage in the clustal alignment) between both human and yeast, and Plasmodium and yeast sequences is shown below.

Pf5

20°C

30°C

37°C



Sc5

Pf5

Sc5

Pf5

Sc5

Pf5

Hs5 (1-69) --M D D E E E T Y R L W K I R K T I M Q L C H D R G Y L V T Q D E L D Q T L E E F K A Q F G D K P S E G R P R R T D L T V L V A H N D D P T --
 Sc5 (1-71) M D Q E N E R N I S R L W R A F R T V K E M V K D R G Y F I T Q E E L P L E D F K A K Y C D S M G R -- P Q R K M M S F Q A N P T E E S I S K
 Pf5 (1-63) --M E D P --V T R F Y K C R K T C C E M L E D R G Y I M T P R E K L E N F S T F K E M F E D N D R Q --R S K M C I S T H K N D S N --

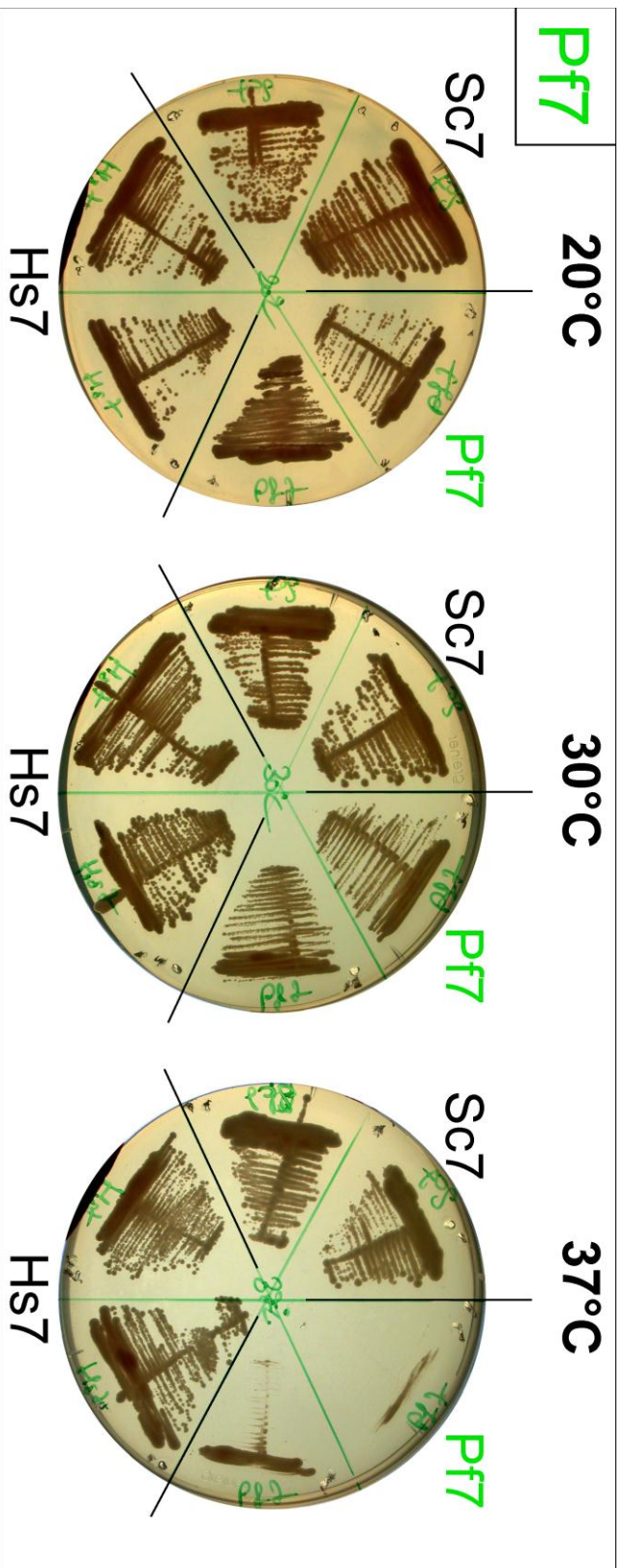
Hs5 (70-137) ---D Q M F V F F P E E P K - V G I K T I K V Y C Q R M Q E E N I T R A L I V V O Q G M T P S A K Q S L V D M A P P K Y I L E Q F L Q Q E L L I
 Sc5 (72-142) F P D M G S L W V E F C D E P S - V G V K T M K T F V I H I Q E K N F Q T G I F V Y Q N N I T P S A M K L V P S I P P - A T I E T F N E A A L V V
 Pf5 (64-132) ---N R I I V Y F A D E V K K T G V K P L R E L T E R M E E K S I Q R A I L V T O N I L T P F A R D A I K E A A P R H I I E N F L D T E L L V

Hs5 (138-210) N I T E H E L V P E H V V M T K E E V T E L L A R Y K L R E N Q L P R I Q A G C D P V A R Y F G I K R G Q V V K I I R P S E T A G R Y I T Y R L V Q
 Sc5 (143-215) N I T H E L V P K H I R L S S D E K R E L L K R Y R L K E S Q L P R I Q R A D P V A L Y L G L K R G E V V K I I R K S E T S G R Y A S Y R I C M
 Pf5 (133-205) N I T K H E L V P R H I P L T S D E K K N L L Q R Y K I M E N K L P R I O D V D P V C R Y F G L S K G Q V V K I I R P S E T A G R Y V T Y R L V V

Hs5	*	Sc5	% Identity = 35.6 (78/219)
Pf5	*	Sc5	% Identity = 38.8 (85/219)

Hazoume Sup. Fig. 3 Pf5 complementation.

Top : YSF11 strain was transformed with pGEN Sc5, pGEN Pf5, as in Figure 1B. Three independent colonies were streaked on three YPGlu rich medium Petri dishes. Each was incubated at either 20, 30 or 37°C, as indicated, for 4 days.
 Bottom : Clustal alignment of Hs5, Sc5 and Pf5 peptidic sequences. The homology was highlighted using the Jalview program (Taylor colour code, conservation 66%). The conservation score (identity percentage in the clustal alignment) between both human and yeast, and Plasmodium and yeast sequences is shown below.



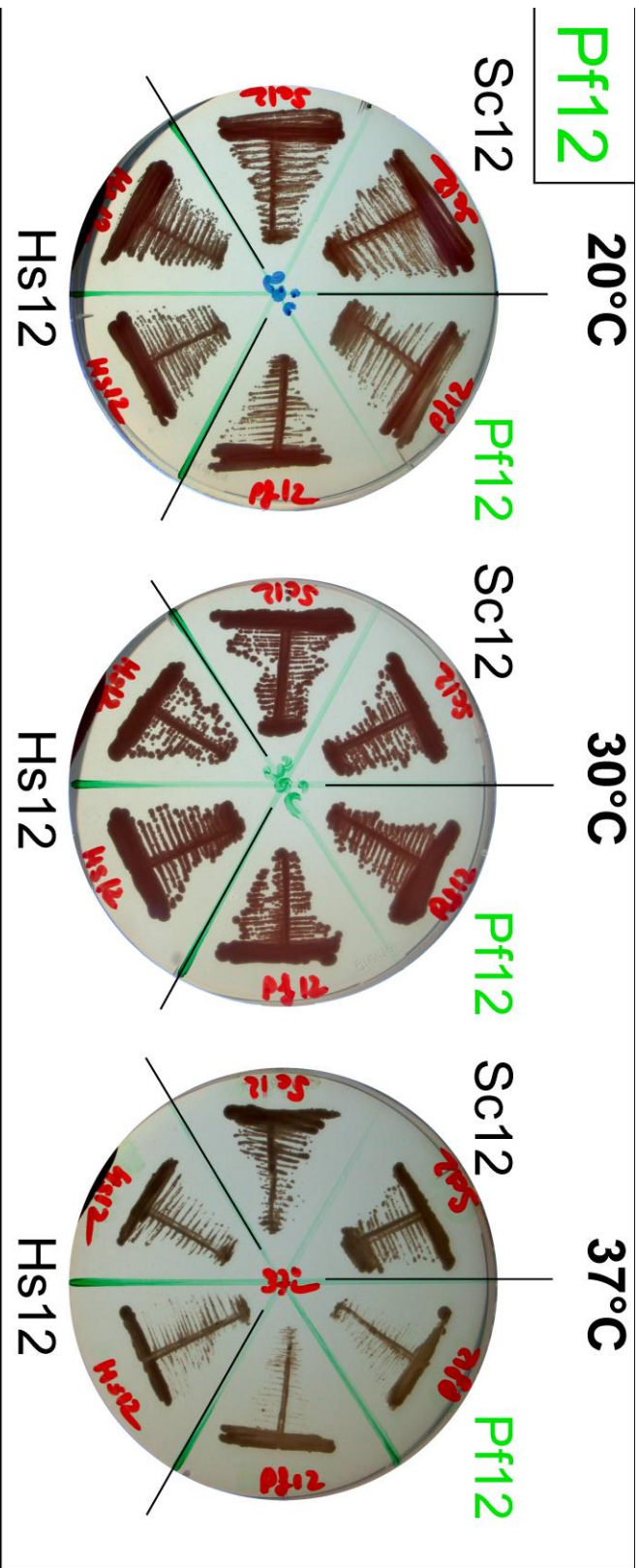
Hs7 (1-88) M F Y H I S L E H E I L L L H P R Y F G P N L L N T V K O K L F T E V E G T C T G K Y G F V I A V T T I D N I G A -- G V I Q P G R G F V L Y P V K V A I V F R P F K G E V V D A V V
 Sc7 (1-90) M F F I K D L S L N I T L H P S F F G P R M K O Y L K T K L L E E V E G S C T G K F C Y I L C V L D Y D N I D I Q R G R I L P T D C S A E F N V W Y R A V V F K P F K G E V V D G T V
 Pf7 (1-88) M Y F V I E E W K N V I I K P S Q L G P R Y Q O Y I E D M L R N S V E G Q C S V K Y G Y V I I C V I R I I H S E P -- G R V Q D G T G M I V V K V K Y Q A I V F K P F K D E V L D A I V

Hs7 (89-172) T Q V N K V G L F T E I G P M S C F I S R H S I P S E M E F D P N S N P P C Y K T M D E D I V I Q Q D D E I K L K I V G T R V D K N D I F A I G S L M D D Y L G L V S
 Sc7 (91-171) V S C S Q H G F E V Q V G P M K V E V T K H L M P Q D L T F N A G S N P P S Y Q S S E D V I T I K S -- R I R V K I E G C I S Q V S S I H A I G S I K E D Y L G A I
 Pf7 (89-177) T D V N K L L G F F A Q A G P L K I E I S R T A I P K Y F E Y S E D L H Y P C F S S G D Y N I K P Q T -- T V R I K I Q G I R Y D L S N M F A I A T I N N E Y L G C C I E S N I L Q V M

Hs7	*	Sc7	% Identity =	42.5	(74/174)
Pf7	*	Sc7	% Identity =	33.0	(59/179)

Hazoume Sup. Fig. 4 Pf7 complementation

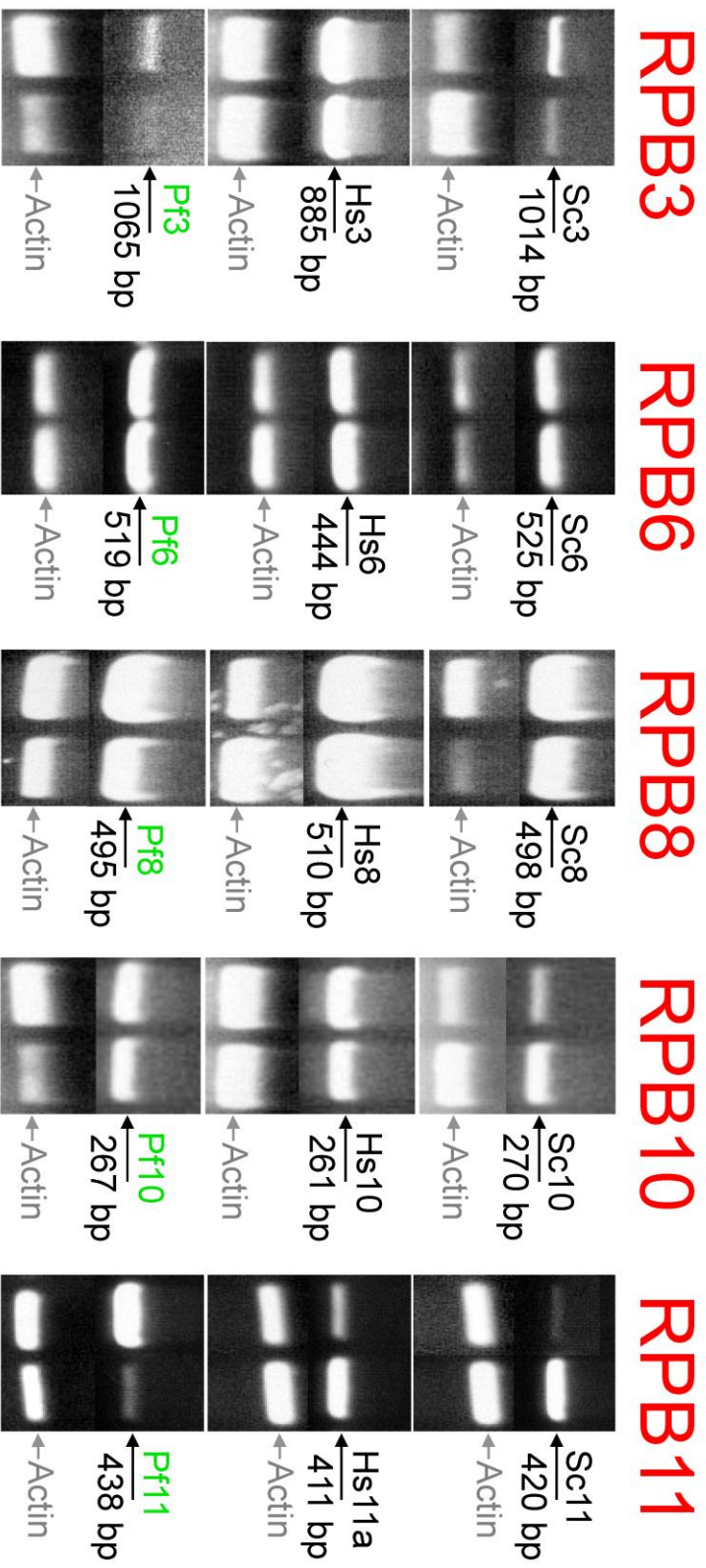
Top : D208-4b strain was transformed with pGEN Sc7, pGEN Pf7, pGEN Hs7 as in Figure 1C. Two independent colonies were streaked on three YPGlu rich medium Petri dishes. Each was incubated at either 20, 30 or 37°C, as indicated, for 5, 2 and 3 days, respectively.
 Bottom : Clustal alignment of Hs7, Sc7 and Pf7 peptidic sequences. The homology was highlighted using the Jalview program (Taylor colour code, conservation 66%). The conservation score (identity percentage in the clustal alignment) between both human and yeast, and Plasmodium and yeast sequences is shown below.



Hs12 (1-58) 1 10 20 30 40 50 60
 -----MDTQK DVQPP -----KQQPMIY ICGE CHTENEI IKS RDP I RCRECGYR I MYK KRTK RL LVV F DAR
 Sc12 (1-70) MSREGFIPTNLD DAAAAGT SQARTATLKY ICAE C SSKLS LSR TDAYR CKD CGHR I LLK ARTK RL VQ FEAR
 Pf12 (1-58) -----MYI REQ DEDIS -----T EPVVY ICGE CGIDTVI PPNASL RCKN CGSK I FFK KRSR R VMQY EAR

Hs12	*	Sc12	% Identity	=	33.8	(24/71)
Pf12	*	Sc12	% Identity	=	28.2	(20/71)

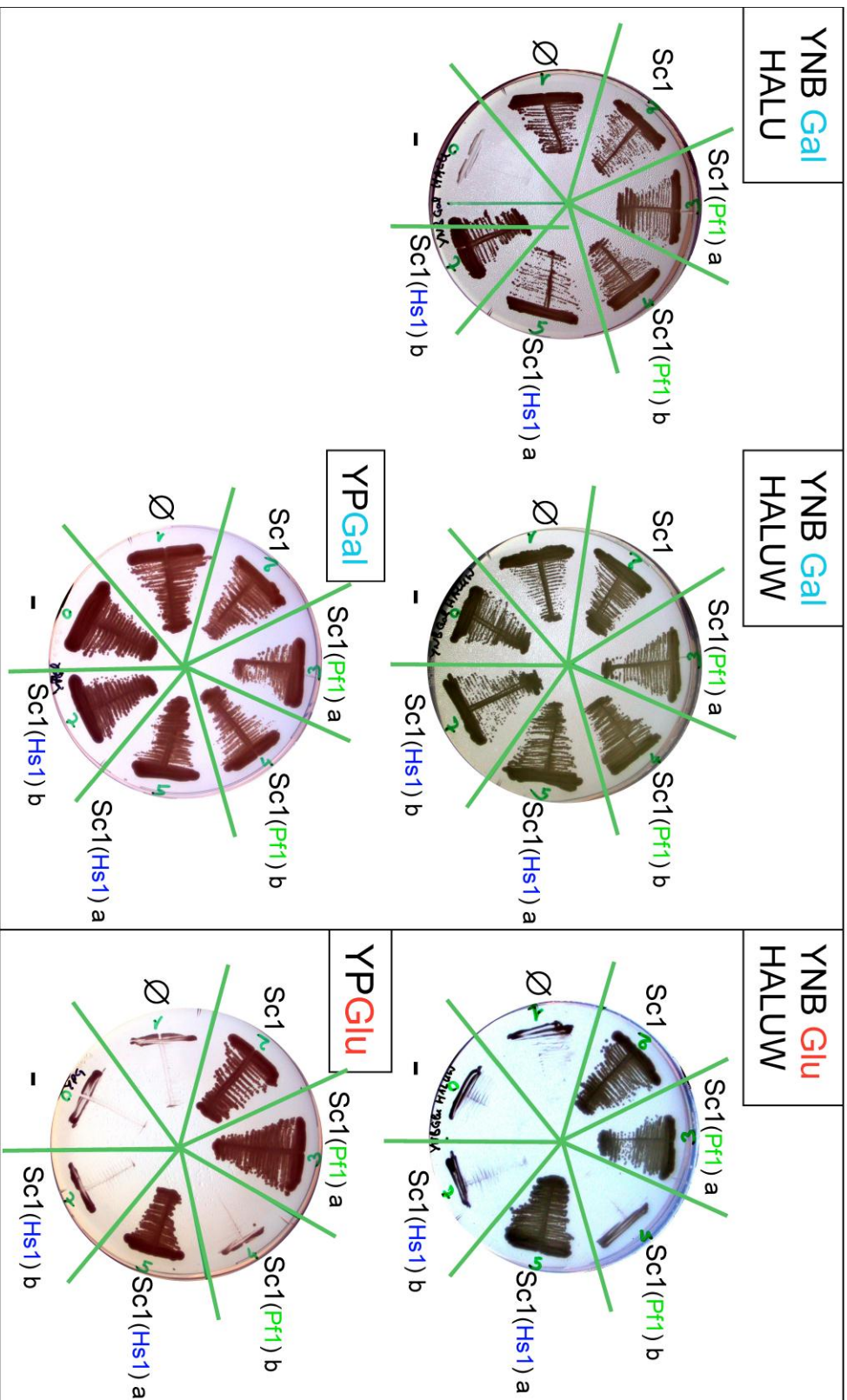
Hazoume Sup. Fig. 6 Pf12 complementation.
 Top : D208-4b strain was transformed with pGEN Sc12, pGEN Pf12, pGEN Hs12, as in Figure 1E. Two independent colonies were streaked on three YPGlu rich medium Petri dishes. Each was incubated at either 20, 30 or 37°C, as indicated, for 2 days.
 Bottom : Clustal alignment of Hs12, Sc12 and Pf12 peptidic sequences. The homology was highlighted using the Jalview program (Taylor colour code, conservation 66%). The conservation score (identity percentage in the clustal alignment) between both human and yeast, and Plasmodium and yeast sequences is shown below.



Hazoume Sup. Fig. 7 RT-PCR mediated detection of specific RNAPII-subunit mRNA encoded by pGEN derivatives. The name of each encoded protein is indicated on the right of each panel, together with the size of the corresponding amplified fragment (bp). All pGEN encoded mRNAs were detected using the same vector-specific primers. Two independent colonies were tested for each subunit. The mRNA encoding Actin 1 (NCBI Reference Sequence ID N°: NM_00117992) was amplified as a reference (in gray) from the total RNA extracted from each transformed yeast strain yielding a 407 bp fragment.

Galactose

Glucose



Hazoume Sup. Fig. 8 RNAPII active center modification.
 JAY47 strain was transformed with pJAY483 vector allowing the expression of Sc1 or its derivatives as described in Figure 2. A representative colony was streaked on minimal medium either supplemented with His, Ade, Leu, Ura (HALU) or (as in Fig. 2B) with His, Ade, Leu, Ura, Trp, (HALUW) using galactose as a fermentable carbon source (Gal). Similarly, (as in Fig. 2B), growth on minimal medium supplemented with His, Ade, Leu, Ura, Trp, (HALUW) using glucose as a fermentable carbon source (Glu) is also shown. In addition, rich medium Petri dishes were used for streaking, that contained either galactose (YPGal) or glucose (YPGlu) as a fermentable carbon sources. The Petri dishes were allowed to incubate at 30°C for two days (rich media) or four days (minimal media).

RPB8

Hs8 (1-74) -MAGCI L F E D I F D V K D I D P E G K K F D R V S R L H C E S E S F K - M D L I L D V N I Q I Y P V D L G D K F R L V I A S T L Y E D G T - - - - L D D G E
Sc8 (1-79) -M S N T L F D D I F Q V S E V D P G - - R Y N K V C R I E A A S T T Q D Q C K L T L D I N V E L F P V A A Q S L T V T I A S S L N L E D T P A N D S S A T R S W
Pf8 (1-72) M A S N I L F E D R F V I S S V D N S - - K F E K V S R I K A K A S T G Y D - A E L I L D V H S E L F K V E E K K A I Y L A L Q D K L L N R N D - - - - E K G W

Hs8 (75-150) Y N P T D D R P S R A D Q F E Y V M Y G K V Y R I E G D E T S T E A A T R L S A V S Y G G L L M R L Q G D A N N L H G F E V D S R V Y L M K K L A F - - - -
Sc8 (80-146) R P P Q A C D R S L A D D Y D Y V M Y G T A Y K F E E V S K - - - - D L I A V Y S F G G L L M R L E G N Y R N L N N L Q E N A Y L L I R - - - -
Pf8 (73-145) E Q - - - N E N V P L N N I E Y I M S G R I K F E E L S S - - - - E R R T V Y A S F G G L M M A L T T D K O F I G D L E I D M K I Y L L T K N I D I E R T R E

Hs8 * Sc8 % Identity = 30.1 (47/156)
Pf8 * Sc8 % Identity = 28.8 (45/146)

RPB10

Hs10 (1-67) M I I P V R C F T C C G K I V G N K W E A Y L G L L Q - A E Y T E G D A L D A L G L K R Y C C R R M L L A H V D L I E K L L N Y A P L E K - -
Sc10 (1-70) M I V P V R C F S C G K V V G D K W E S Y L N L L Q E D E L D E G T A L S R L G L K R Y C C R R M I L T H V D L I E K F L R Y N P L E K R D
Pf10 (1-69) M I I P V R C F T C C G K L I G N L W S V Y E K K L E - E G L S K C D A L N E L N L Y R Y C C R R M I L T H A D M M D K L L C Y N I Y E R K L

Hs10 * Sc10 % Identity = 71.4 (50/70)
Pf10 * Sc10 % Identity = 51.4 (36/70)

RPB11

Hs11a (1-60) - - - - - M N A P P A F E S F L L F E G E K K I T I N K D T K V P N A C L F T I N K E D H T L G N I I K S Q L L K D P Q V L F A G
Sc11 (1-60) - - - - - M N A P D R F E L F L L G E G E S K L K I D P D T K A P N A V V I T F E K E D H T L G N L I R A E L L N D R K V L F A A
Pf11 (1-65) M S V P T L S N K P E T V D L L V L A P G E K K V T C T I S D K G - D C N I F V I K L E D H T I G N L I K I Q L C Q D P K V L F A A

Hs11a (61-117) Y K V P H P L E H K I I R V Q T T P D Y S P Q E A F T N A I T D L I S E L S L L E E R F R V R A G P G G A D G V A I K D K Q E C I E
Sc11 (61-120) Y K V E H P F F A R F K L R I Q T T E G Y D P K D A L K N A C N S I I N K L G A L K T N F E T E W N - - L Q T L A A D D A F - - - -
Pf11 (66-126) Y R Q P H P L Q N A I E I T I K P - K G Y A G V K L L S D N V N N I L L S Q V A T L K E N F A K K I Q K Y K E S N S Y Y E D Y - - - -

Hs11 * Sc11 % Identity = 47.5 (57/120)
Pf11 * Sc11 % Identity = 28.1 (36/128)

Hazoume Sup. Fig 9 Peptidic sequences alignment.
Clustal alignments of (Hs8, Sc8, Pf8), (Hs10, Sc10, Pf10) and (Hs11, Sc11, Pf11) peptidic sequences. The homology was highlighted using the Jalview program (Taylor colour code, conservation 66%). The conservation scores (identity percentages in the clustal alignment) between both Hs and Sc, and Pf and Sc sequences are shown below.

RPPB3

Hs3 (1-88) -----MPYANQPTVRAITELTDENVKFIENITDLAVANSRRVFAEVPFAIDWVQIDAKNSVSLVHDEFIAHRLGIPILSDDIVDKLQYSRDD
 Sc3 (1-86) -----MS-EEGPOVKIREASKDNVDFILSNVYDLAMANSRRVMAEPTLAIDSVETNTTTLADEFIHRLGLIPILQSMDS-IEQLYSRDD
 Pf3 (1-91) MTTYNDNTNKHQINITKLTKDELYFTLYNSNSCIANALRRIMSEIPTLAIDVWVNYENTSAFHDEFIAHRLGIPILDSRN-VNNYEFREKCG

Hs3 (89-180) TCEFFCPESCSEFTLDVRCNEDQTRHVTSTRDINSNRPVIVTSTRNRDNDPNDYEQDDILVKKLRKQQLKLRAYAKKGFCKEFAKNNPTA
 Sc3 (87-174) FCEDHCDKCSVVLTLQAFGESSTTNVYSKDLVIVSN---LMGRNI GHP I IQDKEGNGVLLCKLRKQQLKLCVAKKGI AK EHA KGPAA
 Pf3 (92-181) KCKETCSKCTIQYI LEVKCNHSNKIDVSHYDILESLEHER--NVPMP I PHGNKNSVSENAIP I VTL SKNQT LHMKLIATKGI GKMHAKWIPAN

Hs3 (181-229) GVAFEYDPDNAKHTVYKPE---EWPK S-----EYSELEDEDESEQ-APYDPNGKP-----ERF
 Sc3 (175-228) AIEFEYDPWNLKHTDYWEQDSAKEMPQS-----KCEYEDPNEGDPFDYKAKQA-----DTF
 Pf3 (182-272) -VSYRIDHKVLLKHNLIKLSNEHLLLANLNKDCYILNKDTHDDDDNNNNNIQLKLEKMSVMAESSIDLLSELGYKDIKIVYDETMF

Hs3 (230-275) YYNVESGCSLRPETIIVLSALSGLKKLSDSLQTLQSHEIQ-----SDVLTIN-----
 Sc3 (229-318) YMNVESVCSIPVDQVVRGIDTLQKQVASSILLALTMQDQDKVNFASGDNNTASNMIGSNEDVMTGAEDDPYSNASQMGTGSGGYDNW
 Pf3 (273-335) HFHVESVCSIPPEQIQWMAIDILENKLKVLEPQLKSSFSIDEVAKQLKEQCVSLYGIQLDLE-----

Hs3 * Sc3 % Identity = 33.9 (124/366)
 Pf3 * Sc3 % Identity = 21.3 (78/366)

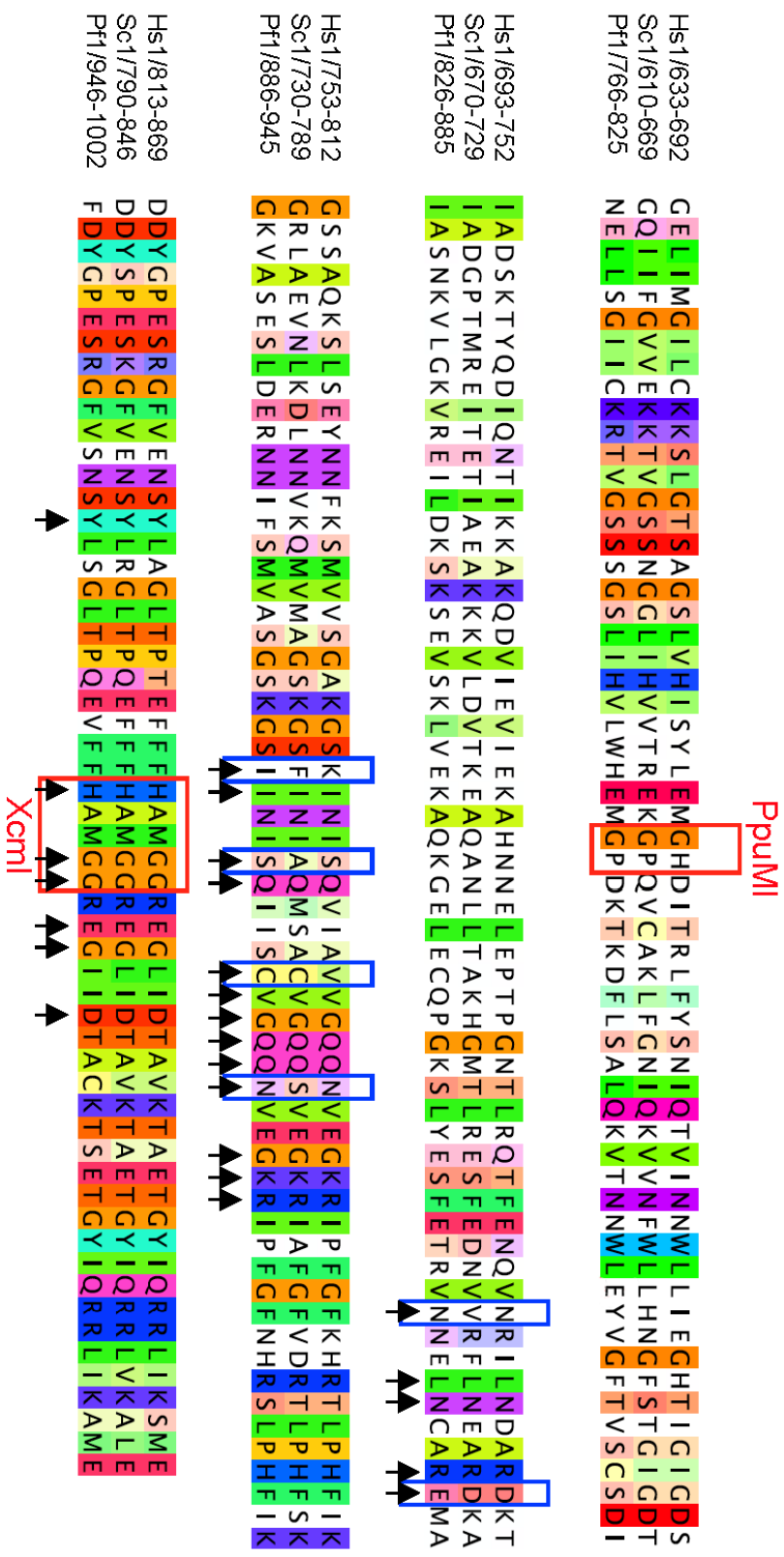
RPPB6

Hs6 (1-57) MSDNEDNF-DGDDDFDVEEDGGLDDL ENAE E EQQENVEILP SCGERPQ-----ANOKRITTPYM
 Sc6 (1-85) MSDYEEAFNDGNENFEDFDVEHFSDSETEYEKPKQFKDGETTDANGKTI V TGGNGP EDFOQHEQIRRKTLEKAI PKDQRAITTPYM
 Pf6 (1-69) MNKYLCLLYFLLDGCFNDNYLNEDEDFMGDEFQGGIDSDDGENYENDIDIITDHIKKDNKSDYE-----NSEANEDNIITSPYL

Hs6 (58-127) TKYERARVLTGTRALQIAMCAPVMVLEEG-----ETDPLLIAMKELKARKIPIIIRRYLPDGSYEDWGVDELITTD
 Sc6 (86-155) TKYERARILGTRALQISMNAPLVFVLDLGG-----ETDPLRIAMKELAEKIPLVIRRYLPDGSFEDWVSEELIVDL
 Pf6 (70-153) TKYERARILGTRALQISMNAPLTIPIETSNDMI NSKNEYDNYLNNP LVI A EKELEA EKSIPIFILLRRYLPNGSYEDWR LDEL I ID-

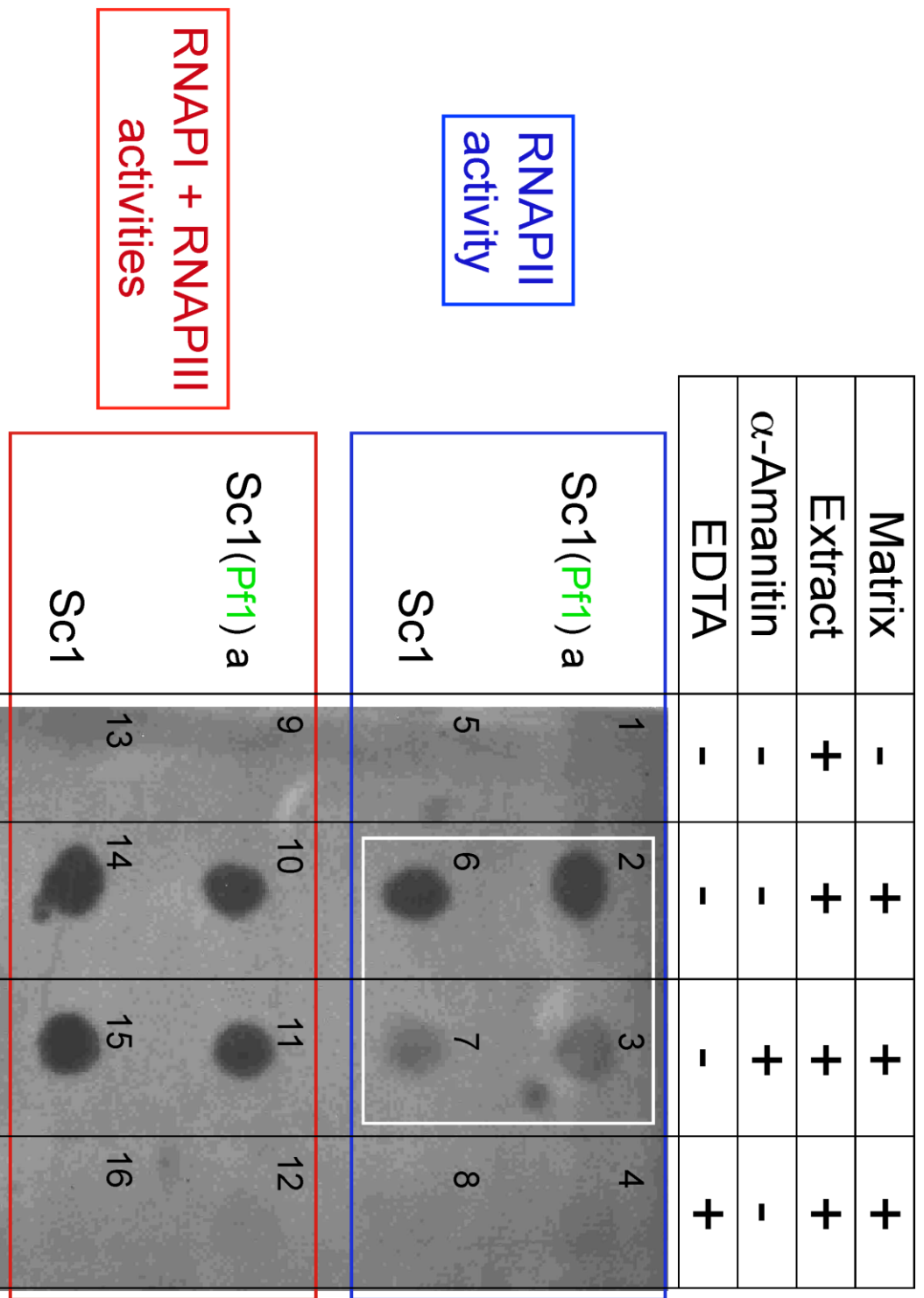
Hs6 * Sc6 % Identity = 46.5 (72/155)
 Pf6 * Sc6 % Identity = 34.1 (58/170)

Hazoume Sup. Fig. 10 Peptidic sequences alignment. Clustal alignments of (Hs3, Sc3, Pf3) and (Hs6, Sc6, Pf6) peptidic sequences. The homology was highlighted using the Jalview program (Taylor colour code, conservation 66%). The conservation scores (identity percentage in the clustal alignment) between both Hs and Sc, and Pf and Sc sequences are shown below.



Arrow heads indicate residues in contact with alpha-Amanitin (Bushnell, PNAS, 2002, see Sup. Fig. 11B)

Hazoume Sup. Fig. 11A. Peptidic sequences alignment.
 Alignment of Hs1, Sc1 and Pt1 alpha-amanitin binding regions. The homology was highlighted using the Jalview program (Taylor colour code, conservation 50%). The residues encoded by the restriction sites (PpuMI, XcmI) used for the swapping are boxed in red. The six residues that were individually mutated are boxed in blue. Fig. 11 bis: data from Bushnell 2002. The residues that were mutated are boxed in blue.



Hazoume Sup. Fig. 12 Alpha-amanitin inhibition of Sc1(Pf1)a RNAPII activity.
 We performed RNA polymerase activity tests as described in Ruet et al. (Eur J Biochem, 1978). Conditions were defined to reveal either the RNAPII alone, or both RNAPI and RNAPIII simultaneously, using specific matrix and conditions. We used extracts either from the control strain (Sc1) or the Sc1(Pf1)a complemented strain. P³² labelled polymerase products are revealed by autoradiography. Clearly, 50 μ g/ml alpha-amanitin similarly inhibits both RNAPII activities (compare 2 and 3, 6 and 7). In contrast RNAPI and III activities are completely insensitive to this toxin (compare 10 and 11, 14 and 15), whereas both activities are abolished in the presence of 12.5 mM EDTA (4, 8, 12, 16). As expected, all these activities are strictly dependent on the presence of the matrix (see 1, 5, 9, 13).

Références bibliographiques

Acker, J., de Graaff, M., Cheynel, I., Khazak, V., Kedinger, C., and Vigneron, M. (1997). Interactions between the human RNA polymerase II subunits. *J Biol Chem* 272, 16815-16821.

Ades, S.E. (2008). Regulation by destruction: design of the sigmaE envelope stress response. *Current opinion in microbiology* 11, 535-540.

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIF kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* 34, 387-393.

Akoulitchev, S., Makela, T.P., Weinberg, R.A., and Reinberg, D. (1995). Requirement for TFIIF kinase activity in transcription by RNA polymerase II. *Nature* 377, 557-560.

Allison, L.A., Wong, J.K., Fitzpatrick, V.D., Moyle, M., and Ingles, C.J. (1988). The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. *Molecular and Cellular Biology* 8, 321-329.

An, J., Yang, T., Huang, Y., Liu, F., Sun, J., Wang, Y., Xu, Q., Wu, D., and Zhou, P. (2011). Strand-specific PCR of UV radiation-damaged genomic DNA revealed an essential role of DNA-PKcs in the transcription-coupled repair. *BMC Biochemistry* 12, 2.

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

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

Austin, B.P., Nallamsetty, S., and Waugh, D.S. (2009). Hexahistidine-tagged maltose-binding protein as a fusion partner for the production of soluble recombinant proteins in *Escherichia coli*. *Methods in molecular biology* 498, 157-172.

Bajic, V.B., Tan, S.L., Christoffels, A., Schonbach, C., Lipovich, L., Yang, L., Hofmann, O., Kruger, A., Hide, W., Kai, C., *et al.* (2006). Mice and men: their promoter properties. *PLoS genetics* 2, e54.

Barboric, M., Nissen, R.M., Kanazawa, S., Jabrane-Ferrat, N., and Peterlin, B.M. (2001). NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Molecular Cell* 8, 327-337.

Bartolomei, M.S., Halden, N.F., Cullen, C.R., and Corden, J.L. (1988). Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Molecular and Cellular Biology* 8, 330-339.

Beckouet, F., Labarre-Mariotte, S., Albert, B., Imazawa, Y., Werner, M., Gadai, O., Nogi, Y., and Thuriaux, P. (2008). Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. *Molecular and Cellular Biology* 28, 1596-1605.

Belikov, S., Astrand, C., Holmqvist, P.H., and Wrangé, O. (2004). Chromatin-mediated restriction of nuclear factor 1/CTF binding in a repressed and hormone-activated promoter in vivo. *Molecular and Cellular Biology* 24, 3036-3047.

Bell, G.I., Valenzuela, P., and Rutter, W.J. (1976). Phosphorylation of yeast RNA polymerases. *Nature* **261**, 429-431.

Bell, G.I., Valenzuela, P., and Rutter, W.J. (1977). Phosphorylation of yeast DNA-dependent RNA polymerases in vivo and in vitro. Isolation of enzymes and identification of phosphorylated subunits. *The Journal of biological chemistry* **252**, 3082-3091.

Benga, W.J., Grandemange, S., Shpakovski, G.V., Shematorova, E.K., Kedinger, C., and Vigneron, M. (2005). Distinct regions of RPB11 are required for heterodimerization with RPB3 in human and yeast RNA polymerase II. *Nucleic Acids Research* **33**, 3582-3590.

Bhattacharya, S., Takada, S., and Jacobson, R.H. (2007). Structural analysis and dimerization potential of the human TAF5 subunit of TFIID. *Proceedings of the National Academy of Sciences* **104**, 1189-1194.

Birse, C.E. (1998). Coupling Termination of Transcription to Messenger RNA Maturation in Yeast. *Science* **280**, 298-301.

Blazek, E., Mittler, G., and Meisterernst, M. (2005). The mediator of RNA polymerase II. *Chromosoma* **113**, 399-408.

Boeing, S., Rigault, C., Heidemann, M., Eick, D., and Meisterernst, M. (2010). RNA Polymerase II C-terminal Heptarepeat Domain Ser-7 Phosphorylation Is Established in a Mediator-dependent Fashion. *Journal of Biological Chemistry* **285**, 188-196.

Bonner, G., Lafer, E.M., and Sousa, R. (1994). The thumb subdomain of T7 RNA polymerase functions to stabilize the ternary complex during processive transcription. *The Journal of biological chemistry* **269**, 25129-25136.

Borukhov, S., and Nudler, E. (2008). RNA polymerase: the vehicle of transcription. *Trends Microbiol* **16**, 126-134.

Boulon, S., Pradet-Balade, B., Verheggen, C., Molle, D., Boireau, S., Georgieva, M., Azzag, K., Robert, M.-C., Ahmad, Y., Neel, H., *et al.* (2010). HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. *Molecular Cell* **39**, 912-924.

Bozdech, Z., Zhu, J., Joachimiak, M.P., Cohen, F.E., Pulliam, B., and Derisi, J.L. (2003). Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol* **4**, R9.

Bregman, D.B., Halaban, R., van Gool, A.J., Henning, K.A., Friedberg, E.C., and Warren, S.L. (1996). UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells. *Proc Natl Acad Sci USA* **93**, 11586-11590.

Briand, J.F., Navarro, F., Rematier, P., Boschiero, C., Labarre, S., Werner, M., Shpakovski, G.V., and Thuriaux, P. (2001). Partners of Rpb8p, a small subunit shared by yeast RNA polymerases I, II and III. *Molecular and Cellular Biology* **21**, 6056-6065.

Brueckner, F., Ortiz, J., and Cramer, P. (2009). A movie of the RNA polymerase nucleotide addition cycle. *Current Opinion in Structural Biology* **19**, 294-299.

Bruno, T., Corbi, N., Di Padova, M., De Angelis, R., Floridi, A., Passananti, C., and Fanciulli, M. (1999). The RNA polymerase II core subunit 11 interacts with keratin 19, a component of the intermediate filament proteins. *FEBS Letters* **453**, 273-277.

Bruno, T., Leonetti, C., Aloe, S., Iacobini, C., Floridi, A., Di Tondo, U., Punturieri, A., and Fanciulli, M. (1998). Levels of expression of hRPB11, a core subassembly subunit of human RNA polymerase II, affect doxorubicin sensitivity and cellular differentiation. *FEBS Letters* **427**, 241-246.

Buratowski, S. (2005). Connections between mRNA 3' end processing and transcription termination. *Current Opinion in Cell Biology* **17**, 257-261.

Buratowski, S., Hahn, S., Sharp, P.A., and Guarente, L. (1988). Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* **334**, 37-42.

Burgess, R.R., Travers, A.A., Dunn, J.J., and Bautz, E.K. (1969). Factor stimulating transcription by RNA polymerase. *Nature* **221**, 43-46.

Burke, T.W., and Kadonaga, J.T. (1997). The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes & Development* **11**, 3020-3031.

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

Bushnell, D.A., and 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.

Butler, J.E.F. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & Development* **16**, 2583-2592.

Campbell, E.A., Westblade, L.F., and Darst, S.A. (2008a). Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. *Current opinion in microbiology* **11**, 121-127.

Campbell, J.M., Lockwood, W.W., Buys, T.P.H., Chari, R., Coe, B.P., Lam, S., and Lam, W.L. (2008b). Integrative genomic and gene expression analysis of chromosome 7 identified novel oncogene loci in non-small cell lung cancer. *Genome* **51**, 1032-1039.

Cermakian, N., Ikeda, T.M., Cedergren, R., and Gray, M.W. (1996). Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Research* **24**, 648-654.

Chalkley, G.E., and Verrijzer, C.P. (1999). DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *The EMBO Journal* **18**, 4835-4845.

Chamberlin, M., and Berg, P. (1962). Deoxyribonucleic acid-directed synthesis of ribonucleic acid by an enzyme from *Escherichia coli*. *Proc Natl Acad Sci U S A* **48**, 81-94.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* *318*, 1780-1782.

Cheetham, G.M., and Steitz, T.A. (2000a). Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr Opin Struct Biol* *10*, 117-123.

Cheetham, G.M., and Steitz, T.A. (2000b). Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Current Opinion in Structural Biology* *10*, 117-123.

Chen, B.-S., and Hampsey, M. (2004). Functional interaction between TFIIB and the Rpb2 subunit of RNA polymerase II: implications for the mechanism of transcription initiation. *Molecular and Cellular Biology* *24*, 3983-3991.

Chen, H.T., and Hahn, S. (2003). Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Molecular Cell* *12*, 437-447.

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

Chilcott, G.S., and Hughes, K.T. (2000). Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar typhimurium and *Escherichia coli*. *Microbiology and molecular biology reviews : MMBR* *64*, 694-708.

Comer, F.I., and Hart, G.W. (2001). Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* *40*, 7845-7852.

Conaway, R.C., and Conaway, J.W. (1993). General initiation factors for RNA polymerase II. *Annu Rev Biochem* *62*, 161-190.

Conaway, R.C., and Conaway, J.W. (2011). Function and regulation of the Mediator complex. *Current Opinion in Genetics & Development* *21*, 225-230.

Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L., and Myers, R.M. (2006). Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Research* *16*, 1-10.

Corbi, N., Bruno, T., De Angelis, R., Di Padova, M., Libri, V., Di Certo, M.G., Spinardi, L., Floridi, A., Fanciulli, M., and Passananti, C. (2005). RNA polymerase II subunit 3 is retained in the cytoplasm by its interaction with HCR, the psoriasis vulgaris candidate gene product. *Journal of Cell Science* *118*, 4253-4260.

Corbi, N., Di Padova, M., De Angelis, R., Bruno, T., Libri, V., Iezzi, S., Floridi, A., Fanciulli, M., and Passananti, C. (2002). The alpha-like RNA polymerase II core subunit 3 (RPB3) is involved in tissue-specific transcription and muscle differentiation via interaction with the myogenic factor myogenin. *FASEB J* *16*, 1639-1641.

Corden, J. (2011). Going Nuclear: Transcribers in Transit. *Molecular Cell* *42*, 143-145.

Corden, J.L., and Patturajan, M. (1997). A CTD function linking transcription to splicing. *Trends Biochem Sci* **22**, 413-416.

Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* **319**, 1791-1792.

Coulson, R.M.R. (2004). Comparative Genomics of Transcriptional Control in the Human Malaria Parasite *Plasmodium falciparum*. *Genome Research* **14**, 1548-1554.

Craighead, J.L., Chang, W.H., and Asturias, F.J. (2002). Structure of yeast RNA polymerase II in solution: implications for enzyme regulation and interaction with promoter DNA. *Structure* **10**, 1117-1125.

Cramer, P. (2000). Architecture of RNA Polymerase II and Implications for the Transcription Mechanism. *Science* **288**, 640-649.

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

Cudney, R., Patel, S., and McPherson, A. (1994). Crystallization of macromolecules in silica gels. *Acta Crystallographica Section D* **50**, 479-483.

Czeko, E., Seizl, M., Augsberger, C., Mielke, T., and Cramer, P. (2011). Iwr1 Directs RNA Polymerase II Nuclear Import. *Molecular Cell* **42**, 261-266.

Das, A., Li, H., Liu, T., and Bellofatto, V. (2006). Biochemical characterization of *Trypanosoma brucei* RNA polymerase II. *Molecular and Biochemical Parasitology* **150**, 201-210.

Davis, J.A., Takagi, Y., Kornberg, R.D., and Asturias, F.A. (2002). Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Molecular Cell* **10**, 409-415.

Davison, B.L., Egly, J.M., Mulvihill, E.R., and Chambon, P. (1983). Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. *Nature* **301**, 680-686.

De Angelis, R., Iezzi, S., Bruno, T., Corbi, N., Di Padova, M., Floridi, A., Fanciulli, M., and Passananti, C. (2003). Functional interaction of the subunit 3 of RNA polymerase II (RPB3) with transcription factor-4 (ATF4). *FEBS Letters* **547**, 15-19.

del Rio-Portilla, F., Gaskell, A., Gilbert, D., Ldias, J.A., and Wagner, G. (1999). Solution structure of the hRPABC14.4 subunit of human RNA polymerases. *Nature structural biology* **6**, 1039-1042.

Desmoucelles, C. (2002). Screening the Yeast "Disruptome" for Mutants Affecting Resistance to the Immunosuppressive Drug, Mycophenolic Acid. *Journal of Biological Chemistry* **277**, 27036-27044.

Desplancq, D., Bernard, C., Sibler, A.-P., Kieffer, B., Miguet, L., Potier, N., Van Dorsselaer, A., and Weiss, E. (2005). Combining inducible protein overexpression with NMR-grade triple isotope labeling in the cyanobacterium *Anabaena* sp. PCC 7120. *BioTechniques* **39**, 405-411.

Desplancq, D., Kieffer, B., Schmidt, K., Posten, C., Forster, A., Oudet, P., Strub, J.M., Van Dorselaer, A., and Weiss, E. (2001). Cost-effective and uniform (13)C- and (15)N-labeling of the 24-kDa N-terminal domain of the Escherichia coli gyrase B by overexpression in the photoautotrophic cyanobacterium Anabaena sp. PCC 7120. *Protein Expression and Purification* 23, 207-217.

Dorjsuren, D., Lin, Y., Wei, W., Yamashita, T., Nomura, T., Hayashi, N., and Murakami, S. (1998). RMP, a novel RNA polymerase II subunit 5-interacting protein, counteracts transactivation by hepatitis B virus X protein. *Molecular and Cellular Biology* 18, 7546-7555.

Dougherty, W.G., Parks, T.D., Cary, S.M., Bazan, J.F., and Fletterick, R.J. (1989). Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. *Virology* 172, 302-310.

Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L.I., Phair, R.D., and Misteli, T. (2002). A kinetic framework for a mammalian RNA polymerase in vivo. *Science* 298, 1623-1626.

Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *The Journal of biological chemistry* 266, 71-75.

Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. *Trends Genet* 24, 280-288.

Egloff, S., Szczepaniak, S.A., Dienstbier, M., Taylor, A., Knight, S., and Murphy, S. (2010). The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *The Journal of biological chemistry* 285, 20564-20569.

Egly, J.M., and Coin, F. (2011). A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor. *DNA repair*.

Eloranta, J.J., Kato, A., Teng, M.S., and Weinzierl, R.O. (1998). In vitro assembly of an archaeal D-L-N RNA polymerase subunit complex reveals a eukaryote-like structural arrangement. *Nucleic Acids Research* 26, 5562-5567.

Fairley, J.A., Evans, R., Hawkes, N.A., and Roberts, S.G. (2002). Core promoter-dependent TFIIIB conformation and a role for TFIIIB conformation in transcription start site selection. *Molecular and Cellular Biology* 22, 6697-6705.

Fan, X., Shi, H., Adelman, K., and Lis, J.T. (2004). Probing TBP interactions in transcription initiation and reinitiation with RNA aptamers that act in distinct modes. *Proc Natl Acad Sci U S A* 101, 6934-6939.

Fanciulli, M., Bruno, T., Cerboni, C., Bonetto, F., Iacobini, C., Frati, L., Piccoli, M., Floridi, A., Santoni, A., and Punturieri, A. (1996). Cloning of a novel human RNA polymerase II subunit downregulated by doxorubicin: new potential mechanisms of drug related toxicity. *FEBS Lett* 384, 48-52.

Fanciulli, M., Bruno, T., Di Padova, M., De Angelis, R., Iezzi, S., Iacobini, C., Floridi, A., and Passananti, C. (2000). Identification of a novel partner of RNA polymerase II subunit 11, Che-1, which interacts with and affects the growth suppression function of Rb. *FASEB J* 14, 904-912.

Fanciulli, M., Bruno, T., Di Padova, M., De Angelis, R., Lovari, S., Floridi, A., and Passananti, C. (1998). The interacting RNA polymerase II subunits, hRPB11 and hRPB3, are coordinately expressed in adult human tissues and down-regulated by doxorubicin. *FEBS Letters* **427**, 236-240.

Fire, A., Samuels, M., and Sharp, P.A. (1984). Interactions between RNA polymerase II, factors, and template leading to accurate transcription. *The Journal of biological chemistry* **259**, 2509-2516.

Fogg, M.J., Alzari, P., Bahar, M., Bertini, I., Betton, J.M., Burmeister, W.P., Cambillau, C., Canard, B., Corrondo, M.A., Coll, M., *et al.* (2006). Application of the use of high-throughput technologies to the determination of protein structures of bacterial and viral pathogens. *Acta crystallographica Section D, Biological crystallography* **62**, 1196-1207.

Forget, D., Lacombe, A.A., Cloutier, P., Al-Khoury, R., Bouchard, A., Lavallee-Adam, M., Faubert, D., Jeronimo, C., Blanchette, M., and Coulombe, B. (2010). The protein interaction network of the human transcription machinery reveals a role for the conserved GTPase RPAP4/GPN1 and microtubule assembly in nuclear import and biogenesis of RNA polymerase II. *Molecular & cellular proteomics : MCP* **9**, 2827-2839.

Forster, T. (1948). Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann Phys* **2**, 55-75.

Frias, J.E., Flores, E., and Herrero, A. (1997). Nitrate assimilation gene cluster from the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *Journal of bacteriology* **179**, 477-486.

Fu, J., Gnatt, A.L., Bushnell, D.A., Jensen, G.J., Thompson, N.E., Burgess, R.R., David, P.R., and Kornberg, R.D. (1999). Yeast RNA polymerase II at 5 Å resolution. *Cell* **98**, 799-810.

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

Gadal, O., Labarre, S., Boschiero, C., and Thuriaux, P. (2002). Hmo1, an HMG-box protein, belongs to the yeast ribosomal DNA transcription system. *The EMBO Journal* **21**, 5498-5507.

Gadal, O., Mariotte-Labarre, S., Chedin, S., Quemeneur, E., Carles, C., Sentenac, A., and Thuriaux, P. (1997). A34.5, a nonessential component of yeast RNA polymerase I, cooperates with subunit A14 and DNA topoisomerase I to produce a functional rRNA synthesis machine. *Molecular and Cellular Biology* **17**, 1787-1795.

Gadal, O., Shpakovski, G.V., and Thuriaux, P. (1999). Mutants in ABC10beta, a conserved subunit shared by all three yeast RNA polymerases, specifically affect RNA polymerase I assembly. *The Journal of biological chemistry* **274**, 8421-8427.

Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B., *et al.* (1979). Organisation and sequences at the 5' end of a cloned complete ovalbumin gene. *Nature* **278**, 428-434.

Garman, E.F., and Schneider, T.R. (1997). *Macromolecular Cryocrystallography*. *Journal of Applied Crystallography* **30**, 211-237.

Gazit, K., Moshonov, S., Elfakess, R., Sharon, M., Mengus, G., Davidson, I., and Dikstein, R. (2009). TAF4/4b-TAF12 displays unique mode of DNA binding and is required for core promoter function of subset of genes. *J Biol Chem*.

Gegonne, A., Weissman, J.D., Lu, H., Zhou, M., Dasgupta, A., Ribble, R., Brady, J.N., and Singer, D.S. (2008). TFIID component TAF7 functionally interacts with both TFIIH and P-TEFb. *Proc Natl Acad Sci U S A* *105*, 5367-5372.

Gershenson, N.I. (2004). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics* *21*, 1295-1300.

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

Ghosh, A., Shuman, S., and Lima, C.D. (2011). Structural Insights to How Mammalian Capping Enzyme Reads the CTD Code. *Molecular Cell*.

Gill, G., Pascal, E., Tseng, Z.H., and Tjian, R. (1994). A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. *Proc Natl Acad Sci U S A* *91*, 192-196.

Gilmour, D.S. (2009). Promoter proximal pausing on genes in metazoans. *Chromosoma* *118*, 1-10.

Glover-Cutter, K., Larochele, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. *Molecular and Cellular Biology* *29*, 5455-5464.

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.

Course, R.L., Ross, W., and Gaal, T. (2000). UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Molecular microbiology* *37*, 687-695.

Grandemange, S., Schaller, S., Yamano, S., Du Manoir, S., Shpakovski, G.V., Mattei, M.G., Kedinger, C., and Vigneron, M. (2001). A human RNA polymerase II subunit is encoded by a recently generated multigene family. *BMC Mol Biol* *2*, 14.

Green, D.W., Ingram, V.M., and Perutz, M.F. (1954). The Structure of Haemoglobin. IV. Sign Determination by the Isomorphous Replacement Method. *Proceedings of the Royal Society of London Series A Mathematical and Physical Sciences* *225*, 287-307.

Greenleaf, A.L., Kelly, J.L., and Lehman, I.R. (1986). Yeast RPO41 gene product is required for transcription and maintenance of the mitochondrial genome. *Proc Natl Acad Sci U S A* *83*, 3391-3394.

Grohmann, D., and Werner, F. (2011). Cycling through transcription with the RNA polymerase F/E (RPB4/7) complex: structure, function and evolution of archaeal RNA polymerase. *Research in Microbiology* *162*, 10-18.

Gronostajski, R.M. (2000). Roles of the NFI/CTF gene family in transcription and development. *Gene* 249, 31-45.

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

Gruber, T.M., and Gross, C.A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual review of microbiology* 57, 441-466.

Guzder, S.N., Sung, P., Bailly, V., Prakash, L., and Prakash, S. (1994). RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* 369, 578-581.

Haeusler, R.A., and Engelke, D.R. (2006). Spatial organization of transcription by RNA polymerase III. *Nucleic Acids Research* 34, 4826-4836.

Hampsey, M. (1998a). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiology and molecular biology reviews : MMBR* 62, 465-503.

Hampsey, M. (1998b). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* 62, 465-503.

Harismendy, O., Gendrel, C.G., Soularue, P., Gidrol, X., Sentenac, A., Werner, M., and Lefebvre, O. (2003). Genome-wide location of yeast RNA polymerase III transcription machinery. *The EMBO journal* 22, 4738-4747.

Harreman, M., Taschner, M., Sigurdsson, S., Anindya, R., Reid, J., Somesh, B., Kong, S.E., Banks, C.A., Conaway, R.C., Conaway, J.W., *et al.* (2009). Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. *Proc Natl Acad Sci U S A* 106, 20705-20710.

Haurie, V., Durrieu-Gaillard, S., Dumay-Odelot, H., Da Silva, D., Rey, C., Prochazkova, M., Roeder, R.G., Besser, D., and Teichmann, M. (2010). Two isoforms of human RNA polymerase III with specific functions in cell growth and transformation. *Proceedings of the National Academy of Sciences* 107, 4176-4181.

Hausner, W., Wettach, J., Hethke, C., and Thomm, M. (1996). Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *The Journal of biological chemistry* 271, 30144-30148.

Hayashi, K., Watanabe, T., Tanaka, A., Furumoto, T., Sato-Tsuchiya, C., Kimura, M., Yokoi, M., Ishihama, A., Hanaoka, F., and Ohkuma, Y. (2005). Studies of *Schizosaccharomyces pombe* TFIIE indicate conformational and functional changes in RNA polymerase II at transcription initiation. *Genes to Cells* 10, 207-224.

Hazoume, A., Naderi, K., Candolfi, E., Kedinger, C., Chatton, B., and Vigneron, M. (2011). A genetic analysis of *Plasmodium falciparum* RNA polymerase II subunits in yeast. *Molecular and biochemical parasitology*.

Hemming, S.A., and Edwards, A.M. (2000). Yeast RNA polymerase II subunit RPB9. Mapping of domains required for transcription elongation. *J Biol Chem* 275, 2288-2294.

Hengartner, C.J., Thompson, C.M., Zhang, J., Chao, D.M., Liao, S.M., Koleske, A.J., Okamura, S., and Young, R.A. (1995). Association of an activator with an RNA polymerase II holoenzyme. *Genes & Development* 9, 897-910.

Hertel, K.J., Lynch, K.W., and Maniatis, T. (1997). Common themes in the function of transcription and splicing enhancers. *Current Opinion in Cell Biology* 9, 350-357.

Hirata, A., Klein, B.J., and Murakami, K.S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451, 851-854.

Hirata, A., and Murakami, K.S. (2009). Archaeal RNA polymerase. *Current Opinion in Structural Biology* 19, 724-731.

Hirose, Y., and Ohkuma, Y. (2007). Phosphorylation of the C-terminal Domain of RNA Polymerase II Plays Central Roles in the Integrated Events of Eucaryotic Gene Expression. *Journal of Biochemistry* 141, 601-608.

Hirtreiter, A., Grohmann, D., and Werner, F. (2010). Molecular mechanisms of RNA polymerase-the F/E (RPB4/7) complex is required for high processivity in vitro. *Nucleic Acids Research* 38, 585-596.

Holstege, F.C., Fiedler, U., and Timmers, H.T. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J* 16, 7468-7480.

Holstege, F.C., van der Vliet, P.C., and Timmers, H.T. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *The EMBO Journal* 15, 1666-1677.

Hsiou, Y., Ding, J., Das, K., Clark, A.D., Jr., Hughes, S.H., and Arnold, E. (1996). Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure* 4, 853-860.

Hu, J., and Bogorad, L. (1990). Maize chloroplast RNA polymerase: the 180-, 120-, and 38-kilodalton polypeptides are encoded in chloroplast genes. *Proc Natl Acad Sci U S A* 87, 1531-1535.

Hu, J., Troxler, R.F., and Bogorad, L. (1991). Maize chloroplast RNA polymerase: the 78-kilodalton polypeptide is encoded by the plastid rpoC1 gene. *Nucleic Acids Research* 19, 3431-3434.

Huang, R.C., Maheshwari, N., and Bonner, J. (1960). Enzymatic synthesis of RNA. *Biochemical and Biophysical Research Communications* 3, 689-694.

Huet, J., Buhler, J.M., Sentenac, A., and Fromageot, P. (1975). Dissociation of two polypeptide chains from yeast RNA polymerase A. *Proc Natl Acad Sci U S A* 72, 3034-3038.

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

Hurwitz, J. (1959). The enzymatic incorporation of ribonucleotides into polydeoxynucleotide material. *The Journal of biological chemistry* 234, 2351-2358.

Ishiguro, A., Nogi, Y., Hisatake, K., Muramatsu, M., and Ishihama, A. (2000). The Rpb6 subunit of fission yeast RNA polymerase II is a contact target of the transcription elongation factor TFIIIS. *Molecular and Cellular Biology* 20, 1263-1270.

Ishihama, A. (1981). Subunit of assembly of *Escherichia coli* RNA polymerase. *Adv Biophys* 14, 1-35.

Iyer, S.P.N., Akimoto, Y., and Hart, G.W. (2003). Identification and cloning of a novel family of coiled-coil domain proteins that interact with O-GlcNAc transferase. *J Biol Chem* 278, 5399-5409.

Jang, S.H., and Jaehning, J.A. (1991). The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. *The Journal of biological chemistry* 266, 22671-22677.

Jasiak, A.J., Hartmann, H., Karakasili, E., Kalocsay, M., Flatley, A., Kremmer, E., Strässer, K., Martin, D.E., Söding, J., and Cramer, P. (2008). Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. *J Biol Chem* 283, 26423-26427.

Jenks, M.H., O'rourke, T.W., and Reines, D. (2008). Properties of an Intergenic Terminator and Start Site Switch That Regulate IMD2 Transcription in Yeast. *Molecular and Cellular Biology* 28, 3883-3893.

Jensen, G.J., Meredith, G., Bushnell, D.A., and Kornberg, R.D. (1998). Structure of wild-type yeast RNA polymerase II and location of Rpb4 and Rpb7. *The EMBO Journal* 17, 2353-2358.

Jeong, S.J., Kim, H.J., Yang, Y.J., Seol, J.H., Jung, B.Y., Han, J.W., Lee, H.W., and Cho, E.J. (2005). Role of RNA polymerase II carboxy terminal domain phosphorylation in DNA damage response. *J Microbiol* 43, 516-522.

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.

Jung, C.R., Choi, S., and Im, D.S. (2007). The NS5A protein of hepatitis C virus represses gene expression of hRPB10alpha, a common subunit of host RNA polymerases, through interferon regulatory factor-1 binding site. *Virus Res* 129, 155-165.

Kaine, B.P., Mehr, I.J., and Woese, C.R. (1994). The sequence, and its evolutionary implications, of a *Thermococcus celer* protein associated with transcription. *Proc Natl Acad Sci U S A* 91, 3854-3856.

Kalogeropoulou, M., Voulgari, A., Kostourou, V., Sandaltzopoulos, R., Dikstein, R., Davidson, I., Tora, L., and Pintzas, A. (2010). TAF4b and Jun/activating protein-1 collaborate to regulate the expression of integrin alpha6 and cancer cell migration properties. *Mol Cancer Res* 8, 554-568.

Kanagaraj, R., Huehn, D., Mackellar, A., Menigatti, M., Zheng, L., Urban, V., Shevelev, I., Greenleaf, A.L., and Janscak, P. (2010). RECQ5 helicase associates with the C-terminal repeat domain of RNA polymerase II during productive elongation phase of transcription. *Nucleic Acids Research* 38, 8131-8140.

Kang, X., Hu, Y., Li, Y., Guo, X., Jiang, X., Lai, L., Xia, B., and Jin, C. (2006). Structural, biochemical, and dynamic characterizations of the hRPB8 subunit of human RNA polymerases. *J Biol Chem* 281, 18216-18226.

Kayukawa, K., Makino, Y., Yogosawa, S., and Tamura, T. (1999). A serine residue in the N-terminal acidic region of rat RPB6, one of the common subunits of RNA polymerases, is exclusively phosphorylated by casein kinase II in vitro. *Gene* 234, 139-147.

Ke, A., and Wolberger, C. (2003). Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera. *Protein science : a publication of the Protein Society* 12, 306-312.

Kedinger, C., Gniazdowski, M., Mandel, J.L., Jr., Gissinger, F., and Chambon, P. (1970). Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. *Biochemical and Biophysical Research Communications* 38, 165-171.

Kelly, W.G., Dahmus, M.E., and Hart, G.W. (1993). RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem* 268, 10416-10424.

Kettenberger, H., Armache, K.-J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIIS. *Molecular Cell* 16, 955-965.

Khazak, V., Estojak, J., Cho, H., Majors, J., Sonoda, G., Testa, J.R., and Golemis, E.A. (1998). Analysis of the interaction of the novel RNA polymerase II (pol II) subunit hsRPB4 with its partner hsRPB7 and with pol II. *Molecular and Cellular Biology* 18, 1935-1945.

Khazak, V., Sadhale, P.P., Woychik, N.A., Brent, R., and Golemis, E.A. (1995). Human RNA polymerase II subunit hsRPB7 functions in yeast and influences stress survival and cell morphology. *Molecular biology of the cell* 6, 759-775.

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

Kim, M., Suh, H., Cho, E.J., and Buratowski, S. (2009). Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *The Journal of biological chemistry* 284, 26421-26426.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., *et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182-187.

Kim, Y., Eom, S.H., Wang, J., Lee, D.S., Suh, S.W., and Steitz, T.A. (1995). Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* 376, 612-616.

Kim, Y., Quartey, P., Li, H., Volkart, L., Hatzos, C., Chang, C., Nocek, B., Cuff, M., Osipiuk, J., Tan, K., *et al.* (2008). Large-scale evaluation of protein reductive methylation for improving protein crystallization. *Nat Methods* 5, 853-854.

Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H., and 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., Ishiguro, A., and Ishihama, A. (1997). RNA polymerase II subunits 2, 3, and 11 form a core subassembly with DNA binding activity. *Journal of Biological Chemistry* **272**, 25851-25855.
- Klauck, E., Typas, A., and Hengge, R. (2007). The sigmaS subunit of RNA polymerase as a signal integrator and network master regulator in the general stress response in *Escherichia coli*. *Sci Prog* **90**, 103-127.
- Kobe, B., Center, R.J., Kemp, B.E., and Pountourios, P. (1999). Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. *Proc Natl Acad Sci U S A* **96**, 4319-4324.
- Koleske, A.J., and Young, R.A. (1994). An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**, 466-469.
- Kolodner, R.D., and Marsischky, G.T. (1999). Eukaryotic DNA mismatch repair. *Current Opinion in Genetics & Development* **9**, 89-96.
- Kolodziej, P.A., and Young, R.A. (1991). Mutations in the three largest subunits of yeast RNA polymerase II that affect enzyme assembly. *Molecular and Cellular Biology* **11**, 4669-4678.
- Komarnitsky, P. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & Development* **14**, 2452-2460.
- Kondo, E., Horii, A., and Fukushige, S. (1999). The human PMS2L proteins do not interact with hMLH1, a major DNA mismatch repair protein. *Journal of Biochemistry* **125**, 818-825.
- Koonin, E.V., Makarova, K.S., and Elkins, J.G. (2007). Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota". *Biol Direct* **2**, 38.
- Kornberg, R. (2005). Mediator and the mechanism of transcriptional activation. *Trends in Biochemical Sciences* **30**, 235-239.
- Kostek, S., Grob, P., Decarlo, S., Lipscomb, J., Garczarek, F., and Nogales, E. (2006). Molecular Architecture and Conformational Flexibility of Human RNA Polymerase II. *Structure* **14**, 1691-1700.
- Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* **106**, 10171-10176.
- Koyama, H., Ito, T., Nakanishi, T., and Sekimizu, K. (2007). Stimulation of RNA polymerase II transcript cleavage activity contributes to maintain transcriptional fidelity in yeast. *Genes to Cells* **12**, 547-559.
- Krapp, S., Kelly, G., Reischl, J., Weinzierl, R.O., and Matthews, S. (1998). Eukaryotic RNA polymerase subunit RPB8 is a new relative of the OB family. *Nature structural biology* **5**, 110-114.

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

Kusser, A.G., Bertero, M.G., Naji, S., Becker, T., Thomm, M., Beckmann, R., and Cramer, P. (2008). Structure of an archaeal RNA polymerase. *J Mol Biol* 376, 303-307.

Kwapisz, M., Beckouet, F., and Thuriaux, P. (2008). Early evolution of eukaryotic DNA-dependent RNA polymerases. *Trends in genetics : TIG* 24, 211-215.

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

Lalo, D., Carles, C., Sentenac, A., and Thuriaux, P. (1993). Interactions between three common subunits of yeast RNA polymerases I and III. *Proc Natl Acad Sci U S A* 90, 5524-5528.

Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. (2006). A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J* 25, 118-128.

Langelier, M.-F. (2005). The highly conserved glutamic acid 791 of Rpb2 is involved in the binding of NTP and Mg(B) in the active center of human RNA polymerase II. *Nucleic Acids Research* 33, 2629-2639.

Langer, D., Hain, J., Thuriaux, P., and Zillig, W. (1995). Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* 92, 5768-5772.

Law, W.J., Cann, K.L., and Hicks, G.G. (2006). TLS, EWS and TAF15: a model for transcriptional integration of gene expression. *Brief Funct Genomic Proteomic* 5, 8-14.

Le, T.T.T. (2005). Mutational Analysis of Human RNA Polymerase II Subunit 5 (RPB5): The Residues Critical for Interactions with TFIIF Subunit RAP30 and Hepatitis B Virus X Protein. *Journal of Biochemistry* 138, 215-224.

Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., and Lewis, B.A. (2005a). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Molecular and Cellular Biology* 25, 9674-9686.

Lee, L.T., Tan-Un, K.C., Lin, M.C., and Chow, B.K. (2005b). Retinoic acid activates human secretin gene expression by Sp proteins and nuclear factor I in neuronal SH-SY5Y cells. *Journal of Neurochemistry* 93, 339-350.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. *Annual review of genetics* 34, 77-137.

Lefèvre, S., Dumay-Odelot, H., El-Ayoubi, L., Budd, A., Legrand, P., Pinaud, N., Teichmann, M., and Fribourg, S. (2011). Structure-function analysis of hRPC62 provides insights into RNA polymerase III transcription initiation. *Nat Struct Mol Biol* 18, 516.

Lei, L., Ren, D., and Burton, Z.F. (1999). The RAP74 subunit of human transcription factor IIF has similar roles in initiation and elongation. *Molecular and Cellular Biology* 19, 8372-8382.

Leber-Mache, S. (1993). The 110-kDa polypeptide of spinach plastid DNA-dependent RNA polymerase: single-subunit enzyme or catalytic core of multimeric enzyme complexes? *Proc Natl Acad Sci U S A* *90*, 5509-5513.

Levine, M. (2011). Paused RNA Polymerase II as a Developmental Checkpoint. *Cell* *145*, 502-511.

Levine, M., and Davidson, E.H. (2005). Gene regulatory networks for development. *Proc Natl Acad Sci U S A* *102*, 4936-4942.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. *Nature* *424*, 147-151.

Lewis, B.A., Kim, T.K., and 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, S., Chen, X., Ruggiero, C., Ding, B., and Smerdon, M.J. (2006a). Modulation of Rad26- and Rpb9-mediated DNA Repair by Different Promoter Elements. *Journal of Biological Chemistry* *281*, 36643-36651.

Li, S., Ding, B., Chen, R., Ruggiero, C., and Chen, X. (2006b). Evidence that the Transcription Elongation Function of Rpb9 Is Involved in Transcription-Coupled DNA Repair in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* *26*, 9430-9441.

Li, S., and Smerdon, M.J. (2004). Dissecting transcription-coupled and global genomic repair in the chromatin of yeast GAL1-10 genes. *The Journal of biological chemistry* *279*, 14418-14426.

Liere, K., Kaden, D., Maliga, P., and Borner, T. (2004). Overexpression of phage-type RNA polymerase RpoTp in tobacco demonstrates its role in chloroplast transcription by recognizing a distinct promoter type. *Nucleic Acids Research* *32*, 1159-1165.

Liere, K., Weihe, A., and Borner, T. (2011). The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. *J Plant Physiol.*

Liljelund, P., Mariotte, S., Buhler, J.M., and Sentenac, A. (1992). Characterization and mutagenesis of the gene encoding the A49 subunit of RNA polymerase A in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* *89*, 9302-9305.

Lim, C.Y. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes & Development* *18*, 1606-1617.

Lin, C., Smith, E.R., Takahashi, H., Lai, K.C., Martin-Brown, S., Florens, L., Washburn, M.P., Conaway, J.W., Conaway, R.C., and Shilatifard, A. (2010). AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Molecular Cell* *37*, 429-437.

Littlefield, O., Korkhin, Y., and Sigler, P.B. (1999). The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci U S A* *96*, 13668-13673.

Liu, Y., Manna, A., Li, R., Martin, W.E., Murphy, R.C., Cheung, A.L., and Zhang, G. (2001). Crystal structure of the SarR protein from *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* **98**, 6877-6882.

Lolli, G. (2009). Binding to DNA of the RNA-polymerase II C-terminal domain allows discrimination between Cdk7 and Cdk9 phosphorylation. *Nucleic Acids Research* **37**, 1260-1268.

Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G., and Choder, M. (2007). The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *The Journal of Cell Biology* **178**, 1133-1143.

Lykke-Andersen, S., and Jensen, T.H. (2007). Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* **89**, 1177-1182.

Mackereth, C.D., Arrowsmith, C.H., Edwards, A.M., and McIntosh, L.P. (2000). Zinc-bundle structure of the essential RNA polymerase subunit RPB10 from *Methanobacterium thermoautotrophicum*. *Proc Natl Acad Sci U S A* **97**, 6316-6321.

Mangus, D.A., Jang, S.H., and Jaehning, J.A. (1994). Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes. *The Journal of biological chemistry* **269**, 26568-26574.

Marco, E., Garcia-Nieto, R., and Gago, F. (2003). Assessment by molecular dynamics simulations of the structural determinants of DNA-binding specificity for transcription factor Sp1. *Journal of Molecular Biology* **328**, 9-32.

Matzke, M., Kanno, T., Huettel, B., Daxinger, L., and Matzke, A.J. (2006). RNA-directed DNA methylation and Pol IVb in *Arabidopsis*. *Cold Spring Harbor symposia on quantitative biology* **71**, 449-459.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357-361.

McKune, K., Moore, P.A., Hull, M.W., and Woychik, N.A. (1995). Six human RNA polymerase subunits functionally substitute for their yeast counterparts. *Molecular and Cellular Biology* **15**, 6895-6900.

McKune, K., and Woychik, N.A. (1994). Functional substitution of an essential yeast RNA polymerase subunit by a highly conserved mammalian counterpart. *Molecular and Cellular Biology* **14**, 4155-4159.

Meinhart, A. (2005). A structural perspective of CTD function. *Genes & Development* **19**, 1401-1415.

Meka, H., Werner, F., Cordell, S.C., Onesti, S., and Brick, P. (2005). Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Research* **33**, 6435-6444.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebricht, R.H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit

RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci USA* **98**, 892-897.

Moras, D. (1993). Polymerases. Two sisters and their cousin. *Nature* **364**, 572-573.

Morris, D.P. (2000). The Splicing Factor, Prp40, Binds the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. *Journal of Biological Chemistry* **275**, 39935-39943.

Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., and Washburn, M.P. (2009). Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol Cell* **34**, 168-178.

Mosrin, C., Riva, M., Beltrame, M., Cassar, E., Sentenac, A., and Thuriaux, P. (1990). The RPC31 gene of *Saccharomyces cerevisiae* encodes a subunit of RNA polymerase C (III) with an acidic tail. *Molecular and Cellular Biology* **10**, 4737-4743.

Moss, T. (2004). At the crossroads of growth control; making ribosomal RNA. *Current Opinion in Genetics & Development* **14**, 210-217.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285-1290.

Murakami, K.S., Masuda, S., and Darst, S.A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**, 1280-1284.

Murakami, K.S., Masuda, S., and Darst, S.A. (2003). Crystallographic analysis of *Thermus aquaticus* RNA polymerase holoenzyme and a holoenzyme/promoter DNA complex. *Methods Enzymol* **370**, 42-53.

Naji, S., Grunberg, S., and Thomm, M. (2007). The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *The Journal of biological chemistry* **282**, 11047-11057.

Nakajima, N., Horikoshi, M., and Roeder, R.G. (1988). Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. *Molecular and Cellular Biology* **8**, 4028-4040.

Nesser, N.K., Peterson, D.O., and 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.

Nguyen, V.T., Giannoni, F., Dubois, M.F., Seo, S.J., Vigneron, M., Kédinger, C., and Bensaude, O. (1996). In vivo degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. *Nucleic Acids Research* **24**, 2924-2929.

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

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

- Oelgeschlager, T. (2002). Regulation of RNA polymerase II activity by CTD phosphorylation and cell cycle control. *J Cell Physiol* 190, 160-169.
- Ohler, U., Liao, G.C., Niemann, H., and Rubin, G.M. (2002). Computational analysis of core promoters in the *Drosophila* genome. *Genome biology* 3, RESEARCH0087.
- Ollis, D.L., Kline, C., and Steitz, T.A. (1985). Domain of *E. coli* DNA polymerase I showing sequence homology to T7 DNA polymerase. *Nature* 313, 818-819.
- Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature reviews Genetics* 12, 283-293.
- Orlicky, S.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. *Journal of Biological Chemistry* 276, 10097-10102.
- Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes & Development* 10, 2657-2683.
- Osheim, Y.N., Proudfoot, N.J., and Beyer, A.L. (1999). EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Molecular Cell* 3, 379-387.
- Oufattole, M., Lin, S.W.-J., Liu, B., Mascarenhas, D., Cohen, P., and Rodgers, B.D. (2006). Ribonucleic acid polymerase II binding subunit 3 (Rpb3), a potential nuclear target of insulin-like growth factor binding protein-3. *Endocrinology* 147, 2138-2146.
- Palancade, B., and Bensaude, O. (2003). Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur J Biochem* 270, 3859-3870.
- Paratkar, S., Deshpande, A.P., Tang, G.Q., and Patel, S.S. (2011). The N-terminal domain of the yeast mitochondrial RNA polymerase regulates multiple steps of transcription. *The Journal of biological chemistry*.
- Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. *Molecular Cell* 23, 297-305.
- Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes & Development* 20, 2922-2936.
- Polyakov, A., Severinova, E., and Darst, S.A. (1995). Three-dimensional structure of *E. coli* core RNA polymerase: promoter binding and elongation conformations of the enzyme. *Cell* 83, 365-373.
- Potter, J.A., Randall, R.E., and Taylor, G.L. (2008). Crystal structure of human IPS-1/MAVS/VISA/Cardif caspase activation recruitment domain. *BMC structural biology* 8, 11.
- Prelich, G. (2002). RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. *Eukaryotic Cell* 1, 153-162.
- Proudfoot, N.J., Furger, A., and 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.
- Pugh, B.F., and Tjian, R. (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes & Development* 5, 1935-1945.
- Qureshi, S.A., Khoo, B., Baumann, P., and Jackson, S.P. (1995). Molecular cloning of the transcription factor TFIIB homolog from *Sulfolobus shibatae*. *Proc Natl Acad Sci U S A* 92, 6077-6081.
- Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., Mccuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. *Cell* 141, 432-445.
- Raskin, C.A., Diaz, G., Joho, K., and McAllister, W.T. (1992). Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *Journal of Molecular Biology* 228, 506-515.
- Ratner, J.N., Balasubramanian, B., Corden, J., Warren, S.L., and Bregman, D.B. (1998). Ultraviolet radiation-induced ubiquitination and proteasomal degradation of the large subunit of RNA polymerase II. Implications for transcription-coupled DNA repair. *J Biol Chem* 273, 5184-5189.
- Ream, T., Haag, J., Wierzbicki, A., Nicora, C., Norbeck, A., Zhu, J., Hagen, G., Guilfoyle, T., Paša-Tolić, L., and Pikaard, C. (2008). Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II. *Molecular Cell*.
- Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.-K., Hagen, G., Guilfoyle, T.J., Pasa-Tolić, L., and Pikaard, C.S. (2009). Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Molecular Cell* 33, 192-203.
- Reich, C., Zeller, M., Milkereit, P., Hausner, W., Cramer, P., Tschochner, H., and Thomm, M. (2009). The archaeal RNA polymerase subunit P and the eukaryotic polymerase subunit Rpb12 are interchangeable in vivo and in vitro. *Molecular Microbiology* 71, 989-1002.
- Ren, B. (2010). Transcription: Enhancers make non-coding RNA. *Nature* 465, 173-174.
- Richardson, J.P. (2003). Loading Rho to terminate transcription. *Cell* 114, 157-159.
- Robinson, M.M., Yatherajam, G., Ranallo, R.T., Bric, A., Paule, M.R., and Stargell, L.A. (2005). Mapping and functional characterization of the TAF11 interaction with TFIIA. *Molecular and Cellular Biology* 25, 945-957.
- Rodeheffer, M.S., and Shadel, G.S. (2003). Multiple interactions involving the amino-terminal domain of yeast mtRNA polymerase determine the efficiency of mitochondrial protein synthesis. *The Journal of biological chemistry* 278, 18695-18701.
- Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224, 234-237.
- Roeder, R.G., and Rutter, W.J. (1970). Specific nucleolar and nucleoplasmic RNA polymerases. *Proc Natl Acad Sci U S A* 65, 675-682.

Rong, M., He, B., McAllister, W.T., and Durbin, R.K. (1998). Promoter specificity determinants of T7 RNA polymerase. *Proc Natl Acad Sci U S A* 95, 515-519.

Rossmann, M.G. (1990). The molecular replacement method. *Acta Crystallogr A* 46 (Pt 2), 73-82.

Ruan, W., Lehmann, E., Thomm, M., Kostrewa, D., and Cramer, P. (2011). Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *Journal of Biological Chemistry*, 1-15.

Rubbi, L., Labarre-Mariotte, S., Chedin, S., and Thuriaux, P. (1999). Functional characterization of ABC10alpha, an essential polypeptide shared by all three forms of eukaryotic DNA-dependent RNA polymerases. *The Journal of biological chemistry* 274, 31485-31492.

Ruet, A., Sentenac, A., Fromageot, P., Winsor, B., and Lacroute, F. (1980). A mutation of the B220 subunit gene affects the structural and functional properties of yeast RNA polymerase B in vitro. *J Biol Chem* 255, 6450-6455.

Runner, V.M., Podolny, V., and Buratowski, S. (2008). The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Molecular and Cellular Biology* 28, 1883-1891.

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

Schaller, S., Grandemange, S., Shpakovski, G.V., Golemis, E.A., Kedinger, C., and Vigneron, M. (1999). Interactions between the full complement of human RNA polymerase II subunits. *FEBS Letters* 461, 253-257.

Schatz, N., Brandlein, S., Ruckl, K., Hensel, F., and Vollmers, H.P. (2010). Diagnostic and therapeutic potential of a human antibody cloned from a cancer patient that binds to a tumor-specific variant of transcription factor TAF15. *Cancer Research* 70, 398-408.

Schlegel, B.P., Green, V.J., Ladias, J.A., and Parvin, J.D. (2000). BRCA1 interaction with RNA polymerase II reveals a role for hRPB2 and hRPB10alpha in activated transcription. *Proc Natl Acad Sci U S A* 97, 3148-3153.

Schubot, F.D., Chen, C.J., Rose, J.P., and Wang, B.C. (2000). Crystallization and preliminary X-ray diffraction analysis of the mitochondrial transcription factor sc-mtTFB from *Saccharomyces cerevisiae*. *Acta crystallographica Section D, Biological crystallography* 56, 902-903.

Schulze, J.M., Kane, C.M., and Ruiz-Manzano, A. (2010). The YEATS domain of Taf14 in *Saccharomyces cerevisiae* has a negative impact on cell growth. *Molecular genetics and genomics* : MGG 283, 365-380.

Schweer, J., Turkeri, H., Kolpack, A., and Link, G. (2010). Role and regulation of plastid sigma factors and their functional interactors during chloroplast transcription - recent lessons from *Arabidopsis thaliana*. *Eur J Cell Biol* 89, 940-946.

Schwer, B., and Shuman, S. (2011). Deciphering the RNA Polymerase II CTD Code in Fission Yeast. *Molecular Cell*.

Selitrennik, M., Duek, L., Lotan, R., and Choder, M. (2006). Nucleocytoplasmic Shuttling of the Rpb4p and Rpb7p Subunits of *Saccharomyces cerevisiae* RNA Polymerase II by Two Pathways. *Eukaryotic Cell* 5, 2092-2103.

Sentenac, A. (1985). Eukaryotic RNA polymerases. *CRC Crit Rev Biochem* 18, 31-90.

Shematorova, E.K., Shpakovskii, D.G., and Shpakovskii, G.V. (2010). [PSM2 and POLR2J gene families as molecular markers of the higher primate evolution]. *Genetika* 46, 1254-1257.

Shilatifard, A., Lane, W.S., Jackson, K.W., Conaway, R.C., and Conaway, J.W. (1996). An RNA polymerase II elongation factor encoded by the human ELL gene. *Science* 271, 1873-1876.

Shpakovski, G., Gadal, O., Labarre-Mariotte, S., Lebedenko, E., Miklos, I., Sakurai, H., Proshkin, S., Van Mullem, V., Ishihama, A., and Thuriaux, P. (2000). Functional conservation of RNA polymerase II in fission and budding yeasts. *J Mol Biol* 295, 1119-1127.

Shpakovski, G.V. (1994). The fission yeast *Schizosaccharomyces pombe* rpb6 gene encodes the common phosphorylated subunit of RNA polymerase and complements a mutation in the corresponding gene of *Saccharomyces cerevisiae*. *Gene* 147, 63-69.

Shpakovski, G.V., Acker, J., Wintzerith, M., Lacroix, J.F., Thuriaux, P., and Vigneron, M. (1995). Four subunits that are shared by the three classes of RNA polymerase are functionally interchangeable between *Homo sapiens* and *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* 15, 4702-4710.

Shukla, A., Natarajan, A., Bhaumik, S., El-Shemy, H.A., and Lightfoot, D. (2009). The interactions of the largest subunit of RNA polymerase II with other cellular proteins: a bioinformatic approach. *Curr Issues Mol Biol* 11 Suppl 1, i65-71.

Sims, R.J. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes & Development* 18, 2437-2468.

Sims, R.J., Rojas, L.A., Beck, D., Bonasio, R., Schuller, R., Drury, W.J., Eick, D., and Reinberg, D. (2011). The C-Terminal Domain of RNA Polymerase II Is Modified by Site-Specific Methylation. *Science* 332, 99-103.

Singh, K., and Modak, M.J. (1998). A unified DNA- and dNTP-binding mode for DNA polymerases. *Trends in Biochemical Sciences* 23, 277-281.

Song, J.J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature structural biology* 10, 1026-1032.

Soppa, J. (1999a). Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box. *Molecular Microbiology* 31, 1589-1592.

Soppa, J. (1999b). Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol* 31, 1295-1305.

Soppa, J. (1999c). Transcription initiation in Archaea: facts, factors and future aspects. *Molecular microbiology* 31, 1295-1305.

Sousa, R. (1996). Structural and mechanistic relationships between nucleic acid polymerases. *Trends in Biochemical Sciences* 21, 186-190.

Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* 364, 593-599.

Spain, M.M., and Govind, C.K. (2011). A role for phosphorylated Pol II CTD in modulating transcription coupled histone dynamics. *transcription* 2, 78-81.

Stefanovsky, V.Y., Pelletier, G., Hannan, R., Gagnon-Kugler, T., Rothblum, L.I., and Moss, T. (2001). An immediate response of ribosomal transcription to growth factor stimulation in mammals is mediated by ERK phosphorylation of UBF. *Molecular Cell* 8, 1063-1073.

Steinmetz, E.J. (1997). Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog? *Cell* 89, 491-494.

Steinmetz, E.J., Ng, S.B.H., Cloute, J.P., and Brow, D.A. (2006). cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. *Molecular and Cellular Biology* 26, 2688-2696.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* 41, 207-234.

Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 185, 60-89.

Sun, Z.W., Tessmer, A., and Hampsey, M. (1996). Functional interaction between TFIIB and the Rpb9 (Ssu73) subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Nucleic Acids Research* 24, 2560-2566.

Suzuki, J.Y., Ytterberg, A.J., Beardslee, T.A., Allison, L.A., Wijk, K.J., and Maliga, P. (2004). Affinity purification of the tobacco plastid RNA polymerase and in vitro reconstitution of the holoenzyme. *The Plant journal : for cell and molecular biology* 40, 164-172.

Taatjes, D.J. (2002). Structure, Function, and Activator-Induced Conformations of the CRSP Coactivator. *Science* 295, 1058-1062.

Taatjes, D.J. (2010). The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends in Biochemical Sciences* 35, 315-322.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., *et al.* (2011). Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors. *Cell* 146, 92-104.

Takramah, D., Seiffert, B.M., Schaller, S., Vigneron, M., and Häcker, G. (2003). Baculovirus P35 interacts with a subunit of human RNA polymerase II and can enhance promoter activity in human cells. *J Gen Virol* 84, 3011-3019.

Tan, Q., Linask, K.L., Ebright, R.H., and Woychik, N.A. (2000). Activation mutants in yeast RNA polymerase II subunit RPB3 provide evidence for a structurally conserved surface required for activation in eukaryotes and bacteria. *Genes & Development* 14, 339-348.

Tan, Q., Prysak, M.H., and Woychik, N.A. (2003). Loss of the Rpb4/Rpb7 subcomplex in a mutant form of the Rpb6 subunit shared by RNA polymerases I, II, and III. *Molecular and Cellular Biology* 23, 3329-3338.

Tanese, N., Pugh, B.F., and Tjian, R. (1991). Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes & Development* 5, 2212-2224.

Teichmann, M., Dieci, G., Pascali, C., and Boldina, G. (2010). General transcription factors and subunits of RNA polymerase III: Paralogs for promoter- and cell type-specific transcription in multicellular eukaryotes. *transcription* 1, 130-135.

Thiru, A., Hodach, M., Eloranta, J.J., Kostourou, V., Weinzierl, R.O., and Matthews, S. (1999). RNA polymerase subunit H features a beta-ribbon motif within a novel fold that is present in archaea and eukaryotes. *Journal of Molecular Biology* 287, 753-760.

Thomas, M.C., and Chiang, C.-M. (2006). The General Transcription Machinery and General Cofactors. *Crit Rev Biochem Mol Biol* 41, 105-178.

Thompson, C.M., Koleske, A.J., Chao, D.M., and Young, R.A. (1993). A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* 73, 1361-1375.

Thompson, M., Haeusler, R.A., Good, P.D., and Engelke, D.R. (2003). Nucleolar clustering of dispersed tRNA genes. *Science* 302, 1399-1401.

Thuillier, V., Stettler, S., Sentenac, A., Thuriaux, P., and Werner, M. (1995). A mutation in the C31 subunit of *Saccharomyces cerevisiae* RNA polymerase III affects transcription initiation. *The EMBO Journal* 14, 351-359.

Todone, F., Weinzierl, R.O., Brick, P., and Onesti, S. (2000). Crystal structure of RPB5, a universal eukaryotic RNA polymerase subunit and transcription factor interaction target. *Proc Natl Acad Sci USA* 97, 6306-6310.

Tombácz, I., Schauer, T., Juhász, I., Komonyi, O., and Boros, I. (2009). The RNA Pol II CTD phosphatase Fcp1 is essential for normal development in *Drosophila melanogaster*. *Gene* 446, 58-67.

Tora, L., and Timmers, H.T.M. (2010). The TATA box regulates TATA-binding protein (TBP) dynamics in vivo. *Trends in Biochemical Sciences* 35, 309-314.

Tsai, F.T., and Sigler, P.B. (2000). Structural basis of preinitiation complex assembly on human pol II promoters. *The EMBO Journal* 19, 25-36.

Újvári, A., and 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., Pal, M., and Luse, D.S. (2011). The functions of TFIIF during initiation and transcript elongation are differentially affected by phosphorylation by casein kinase 2. *The Journal of biological chemistry*.

Ullah, H., Scappini, E.L., Moon, A.F., Williams, L.V., Armstrong, D.L., and Pedersen, L.C. (2008). Structure of a signal transduction regulator, RACK1, from *Arabidopsis thaliana*. *Protein science : a publication of the Protein Society* 17, 1771-1780.

Uptain, S.M., Kane, C.M., and Chamberlin, M.J. (1997). Basic mechanisms of transcript elongation and its regulation. *Annu Rev Biochem* 66, 117-172.

van Ingen, H., van Schaik, F.M., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A.C., Kruijzer, J.A., Liskamp, R.M., Timmers, H.T., and Boelens, R. (2008). Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure* 16, 1245-1256.

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

Verma-Gaur, J., Rao, S.N., Taya, T., and Sadhale, P. (2008). Genomewide recruitment analysis of Rpb4, a subunit of polymerase II in *Saccharomyces cerevisiae*, reveals its involvement in transcription elongation. *Eukaryotic Cell* 7, 1009-1018.

Voutsina, A., Riva, M., Carles, C., and Alexandraki, D. (1999). Sequence divergence of the RNA polymerase shared subunit ABC14.5 (Rpb8) selectively affects RNA polymerase III assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Research* 27, 1047-1055.

Walmacq, C., Kireeva, M., Irvin, J., Nedialkov, Y., Lubkowska, L., Malagon, F., Strathern, J., and Kashlev, M. (2009). RPB9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II. *J Biol Chem*.

Walter, T.S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2006). Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* 14, 1617-1622.

Wang, B., Jones, D.N., Kaine, B.P., and Weiss, M.A. (1998). High-resolution structure of an archaeal zinc ribbon defines a general architectural motif in eukaryotic RNA polymerases. *Structure* 6, 555-569.

Wang, D., Bushnell, D., Westover, K., Kaplan, C., and Kornberg, R. (2006). Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. *Cell* 127, 941-954.

Wei, W. (2001). Direct Interaction between the Subunit RAP30 of Transcription Factor IIF (TFIIF) and RNA Polymerase Subunit 5, Which Contributes to the Association between TFIIF and RNA Polymerase II. *Journal of Biological Chemistry* 276, 12266-12273.

Weiss, S.B., and Gladstone, L. (1959). A MAMMALIAN SYSTEM FOR THE INCORPORATION OF CYTIDINE TRIPHOSPHATE INTO RIBONUCLEIC ACID. *J Am Chem Soc* 81, 4118-4119.

Werner, F. (2007). Structure and function of archaeal RNA polymerases. *Molecular microbiology* 65, 1395-1404.

Werner, F., Eloranta, J.J., and Weinzierl, R.O. (2000). Archaeal RNA polymerase subunits F and P are bona fide homologs of eukaryotic RPB4 and RPB12. *Nucleic Acids Research* 28, 4299-4305.

Werner, F., and Weinzierl, R.O. (2002). A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* 10, 635-646.

West, M.L., and Corden, J.L. (1995). Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140, 1223-1233.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* 119, 481-489.

White, R.J. (2011). Transcription by RNA polymerase III: more complex than we thought. *Nature reviews Genetics*.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135, 635-648.

Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* 41, 630-634.

Woychik, N.A., and Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. *Cell* 108, 453-463.

Woychik, N.A., Lane, W.S., and Young, R.A. (1991). Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *The Journal of biological chemistry* 266, 19053-19055.

Woychik, N.A., Liao, S.M., Kolodziej, P.A., and Young, R.A. (1990). Subunits shared by eukaryotic nuclear RNA polymerases. *Genes & Development* 4, 313-323.

Wu, S.Y., and Chiang, C.M. (1998). Properties of PC4 and an RNA polymerase II complex in directing activated and basal transcription in vitro. *The Journal of biological chemistry* 273, 12492-12498.

Wu, S.Y., Thomas, M.C., Hou, S.Y., Likhite, V., and Chiang, C.M. (1999). Isolation of mouse TFIID and functional characterization of TBP and TFIID in mediating estrogen receptor and chromatin transcription. *The Journal of biological chemistry* 274, 23480-23490.

Wu, W., Nishikawa, H., Hayami, R., Sato, K., Honda, A., Aratani, S., Nakajima, T., Fukuda, M., and Ohta, T. (2007). BRCA1 Ubiquitinates RPB8 in Response to DNA Damage. *Cancer Research* 67, 951-958.

Wu, X., Chang, A., Sudol, M., and Hanes, S.D. (2001). Genetic interactions between the ESS1 prolyl-isomerase and the RSP5 ubiquitin ligase reveal opposing effects on RNA polymerase II function. *Current genetics* 40, 234-242.

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

Yan, M., and Gralla, J.D. (1997). Multiple ATP-dependent steps in RNA polymerase II promoter melting and initiation. *The EMBO Journal* 16, 7457-7467.

Yang, H., Gu, J., Zheng, Q., Li, M., Lian, X., Miao, J., Jiang, J., and Wei, W. (2011). RPB5-mediated Protein Is Required for the Proliferation of Hepatocellular Carcinoma Cells. *Journal of Biological Chemistry* 286, 11865-11874.

Yee, A., Booth, V., Dharamsi, A., Engel, A., Edwards, A.M., and Arrowsmith, C.H. (2000). Solution structure of the RNA polymerase subunit RPB5 from *Methanobacterium thermoautotrophicum*. *Proc Natl Acad Sci U S A* 97, 6311-6315.

Yeo, M. (2003). A Novel RNA Polymerase II C-terminal Domain Phosphatase That Preferentially Dephosphorylates Serine 5. *Journal of Biological Chemistry* 278, 26078-26085.

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

Zawel, L., and Reinberg, D. (1995). Common themes in assembly and function of eukaryotic transcription complexes. *Annu Rev Biochem* 64, 533-561.

Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., and Greenleaf, A.L. (1988). The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. *Proc Natl Acad Sci U S A* 85, 3698-3702.

Zhang, G., and Darst, S.A. (1998). Structure of the *Escherichia coli* RNA polymerase alpha subunit amino-terminal domain. *Science* 281, 262-266.

Zillig, W., Palm, P., and Heil, A. (1976). Function and Reassembly of Subunits of DNA-dependent RNA Polymerase. *RNA Polymerase*.