

**UNIVERSITÉ LOUIS PASTEUR
STRASBOURG I
ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ**

THÈSE

présentée
pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ LOUIS PASTEUR
Discipline : Sciences du vivant
Domaine : Aspects Moléculaires et Cellulaires de la Biologie

par

Sophie JAEGER

**Étude de la maturation de l'extrémité 3' non
traduite et de la traduction de l'ARN messager
codant pour l'histone H4**

Soutenue le 25 Novembre 2005 devant la commission d'examen :

M.	Gilbert ERIANI
M.	Eric WESTHOF
Mme	Christiane BRANLANT
Mme	Annie CAHOUR
M	Richard GIEGE

Directeur de Thèse
Rapporteur interne
Rapporteur externe
Rapporteur externe
Examinateur

<i>Abréviations</i>	7
<i>INTRODUCTION</i>	9
<i>1 Les histones, des composantes essentielles du nucléosome</i>	11
<i>2 Fonctions du nucléosome</i>	15
2.1 Compaction de l'ADN	15
2.2 Régulation de l'expression du génome, information épigénétique	15
2.2.1 Ajout de modifications post-traductionnelles sur les histones.....	16
2.2.1.1 Les différents types de modification des histones.....	17
2.2.1.2 Régulation de l'ajout et du retrait des modifications touchant les histones	18
2.2.2 Modifications de la structure et composition du nucléosome.....	18
2.2.2.1 Remodelage de la structure du nucléosome dépendant de l'ATP	18
2.2.2.2 Incorporation de variants d'histones.....	20
2.2.2.2.1 Caractéristiques des variants d'histones.....	20
2.2.2.2.2 Fonctions des variants d'histones, quelques exemples.....	21
2.2.2.2.3 Régulation de l'expression des variants d'histones	26
2.2.2.2.4 Incorporation des variants dans les nucléosomes	26
2.3 Transmission et modulation de l'information épigénétique	27
2.3.1 Le code histone.....	27
2.3.2 Héritage de l'information épigénétique	27
2.3.3 Dynamique de la chromatine	28
<i>3 Synthèse des histones majeures et régulation de leur expression</i>	29
3.1 Nécessité d'une régulation de la production des histones	29
3.2 Régulation de la synthèse d'histones dans la levure et les plantes	31
3.3 Expression et régulation des gènes réplication-dépendants chez les métazoaires	33
3.3.1 Transcription.....	34
3.3.2 Maturation de l'extrémité 3'	37
3.3.2.1 Les différents acteurs de la réaction de maturation.....	38
3.3.2.1.1 La snRNP U7.....	38
3.3.2.1.2 La protéine ZFP100.....	41
3.3.2.1.3 Le complexe HBP:tige-boucle sur les ARN pré-messagers d'histones	41
3.3.2.1.3.1 Les partenaires	41
3.3.2.1.3.2 Rôle du complexe HBP:tige-boucle dans la maturation en 3'	44
3.3.2.1.4 Le site de coupure	48
3.3.2.2 Lsm11 est-elle responsable de la coupure ?	49
3.3.3 Export de l'ARNm mature du noyau vers le cytoplasme	49

3.3.4	Traduction.....	51
3.3.4.1	Rôle de la protéine HBP dans la traduction.....	51
3.3.4.2	Des études très controversées.....	54
3.3.5	Stabilité des ARNm d'histones.....	55
4	Article 1.....	57
	AVANT-PROPOS DE LA THÈSE	67
	CHAPITRE 1.....	71
1	Introduction	73
2	La méthode du triple hybride	74
3	Résultats.....	75
4	Discussion	76
5	Article 2.....	78
6	Article 3.....	86
	CHAPITRE 2.....	95
1	Introduction	97
2	Résultats.....	98
3	Discussion	105
4	Article 4.....	107
	CHAPITRE 3.....	129
1	Introduction	131
2	Résultats.....	133
2.1	L'ARNm de l'histone H4 présente un mécanisme de traduction original.....	133
2.2	Étude de la structure secondaire de l'ARNm de l'histone H4	135
2.3	Identification des éléments structuraux de l'ARNm impliqués dans la traduction	138
2.4	Sondage de l'ARNm en présence de petits ARNs anti-sens.....	139
2.5	Identification de nucléotides impliqués dans la traduction par mutagenèse dirigée	140
2.6	Sondage de la structure des ARNm mutés	141
2.7	Recrutement du ribosome.....	141

2.8	Recrutement de deux ribosomes, cartographie du deuxième site de recrutement.....	142
3	<i>Discussion</i>	144
4	<i>Article 5</i>	149
	<i>CONCLUSION ET PERSPECTIVES</i>	187
	<i>MATÉRIEL ET MÉTHODES</i>	193
1	<i>Organismes et vecteurs utilisés</i>	193
1.1	Souche de <i>Saccharomyces cerevisiae</i>	193
1.2	Souche d' <i>Escherichia coli</i>	193
1.3	Vecteurs.....	193
2	<i>Culture, transformation et manipulation des souches de levures et de bactéries</i>	195
2.1	Culture des cellules	195
2.2	Transformation d' <i>E. coli</i> par électroporation.....	197
2.3	Transformation de <i>S. cerevisiae</i>	197
2.3.1	Transformation à l'acétate de lithium.....	197
2.3.2	Transformation à haute efficacité	197
2.4	Utilisation du triple hybride	198
2.5	Test en gouttes	199
3	<i>Tampons et solutions couramment utilisés</i>	199
4	<i>Méthodes relatives à l'étude de l'ADN</i>	200
4.1	<i>Techniques générales de manipulation de l'ADN</i>	200
4.1.1	Détermination de la concentration de l'ADN.....	200
4.1.2	Précipitation de l'ADN.....	200
4.1.3	Extraction des protéines au phénol	200
4.1.4	Digestion et modification enzymatique de l'ADN.....	200
4.1.4.1	Hydrolyse par les enzymes de restriction	200
4.1.4.2	Déphosphorylation des extrémités 5' des vecteurs linéarisés.....	201
4.1.4.3	Ligation de fragments d'ADN	201
4.1.5	Séparation et purification de fragments d'ADN.....	201
4.1.5.1	Séparation sur gel d'agarose	201
4.1.5.2	Élution des fragments d'ADN séparés sur gel d'agarose.....	201
4.1.6	Séquençage automatique	202

4.2	Extraction et purification de l'ADN	202
4.2.1	Minipréparation d'ADN plasmidique par la méthode de la lyse alcaline	202
4.2.2	Préparation à grande échelle d'ADN plasmidique	202
4.2.3	Préparation d'ADN plasmidique de levure	202
4.2.4	Mutagenèse aléatoire de l'ADN par traitement à l'hydroxylamine.....	203
5	Méthodes relatives à l'étude de l'ARN	203
5.1	Techniques générales de manipulation de l'ARN.....	203
5.1.1	Détermination de la concentration de l'ARN.....	203
5.1.2	Précipitation de l'ARN	203
5.2	Séparation des molécules d'ARN sur gel de polyacrylamide dénaturant	203
5.3	Préparation des molécules d'ARN par transcription <i>in vitro</i>	204
5.3.1	Préparation de l'ADN matrice pour la transcription par l'ARN polymérase du phage T7	204
5.3.1.1	Clonage d'une matrice par assemblage d'oligonucléotides.....	205
5.3.1.2	Préparation d'une matrice constituée d'oligonucléotides	206
5.3.1.3	Obtention d'une matrice à transcrire par amplification par PCR	206
5.3.2	Transcription par la T7 ARN polymérase	207
5.3.3	Purification des transcrits	207
5.4	Marquage radioactif de l'ARN	207
5.4.1	Marquage de l'extrémité 5'	207
5.4.1.1	Déphosphorylation des extrémités 5' des transcrits par la BAP.....	208
5.4.1.2	Marquage de l'extrémité 5' par la T4 polynucléotide kinase	208
5.4.2	Marquage de l'extrémité 3' par la T4 ARN ligase.....	208
5.5	Purification des transcrits marqués	209
6	Méthodes relatives à la manipulation de protéines.....	209
6.1.1	Concentration des protéines	209
6.1.2	Gels de retardement	210
7	Méthodes relatives à l'étude de la structure en solution des ARNs et de l'empreinte de la protéine HBP	211
7.1	Principe	211
7.2	Les différents sondages.....	212
7.2.1	Sondage enzymatique.....	212
7.2.2	Sondage au plomb	212
7.2.3	Sondage chimique.....	213
7.2.4	Empreinte en solution de la protéine HBP sur les molécules d'ARN	215
7.2.5	Sondage en solution de l'ARNm bloqué par des ARNs anti-sens.....	215

7.2.6	Digestion à la RNase H des duplexes ARNm-ARN anti-sens.....	215
7.3	Détection des coupures ou modifications obtenues au cours des sondages en solution	216
7.3.1	Détection directe des coupures sur un ARN radioactif.....	216
7.3.2	Détection indirecte des coupures ou modifications par transcription inverse d'un ARN non marqué	217
7.3.2.1	Marquage de l'oligonucléotide amorce pour la détection indirecte	218
7.3.2.2	Élongation de l'amorce marquée par transcription inverse	219
7.3.2.3	Séquençage par transcription inverse	219
7.3.2.4	Hydrolyse alcaline de l'ARN matrice.....	220
Références.....		221

Abréviations

3' UTR	:	« 3' UnTranslated Region »
3AT	:	3-aminotriazole
A _{700nm}	:	absorbance à 700 nm
Amp	:	ampicilline
BSA	:	albumine de sérum bovin
C-terminal	:	carboxy-terminal
DNase	:	désoxyribonucléase
dNTP	:	déoxyribonucléoside triphosphate
DTE	:	dithioéritol
DTT	:	dithiothréitol
eau milliQ	:	eau désionisée et filtrée (Millipore)
HBP	:	« Hairpin Binding Protein »
HDE	:	« Histone Downstream Element »
HEPES	:	acide N-2 hydroxy éthyl pipérazine N'2 éthane sulfonique
IPTG	:	isopropyl-1-thio-β-D-galactopyranoside
N-terminal	:	amino-terminal
PCR	:	« Polymerase Chain Reaction »
p/v	:	poids à volume
RBD	:	« RNA Binding Domain »
RNase	:	ribonucléase
rpm	:	rotation par minute
SDS	:	dodécyl sulfate de sodium
SLBP	:	« Stem Loop Binding Protein »
TEMED	:	N,N,N,N'-tétraméthyl éthylène diamine
Tris	:	tris-(hydroxyméthyl) aminométhane
U	:	unité
UV	:	ultra violet
v/v	:	volume à volume
X-Gal	:	5-bromo-4-chloro-3-indoyl-β-D-galactoside

INTRODUCTION

1 LES HISTONES, DES COMPOSANTES ESSENTIELLES DU NUCLÉOSOME

La chromatine est formée par l'enroulement d'une molécule d'ADN, support de l'information génétique, autour de protéines appelées histones. Les histones sont de petites protéines d'environ 100 à 130 résidus particulièrement riches en acides aminés basiques. Elles font partie des protéines les plus conservées au cours de l'évolution et les plus lentes à évoluer (Kamakaka & Biggins, 2005). Cinq classes d'histones ont été décrites, ce sont les histones H1, H2a, H2b, H3 et H4. Chez les métazoaires, les histones peuvent également se répartir en deux familles. La première contient les histones appelées majeures. Leur expression est massive et leur incorporation à l'ADN s'effectue uniquement au cours de la phase S du cycle cellulaire lors de la réplication de l'ADN. Les gènes codants pour ces histones sont qualifiés de réplication-dépendants. Ces gènes se regroupent au sein de trois « clusters » localisés sur les chromosomes 1 et 6 chez l'homme. La seconde famille de gènes d'histones correspond à des variants d'histones qualifiés de réplication-indépendants ou histones de réparation. Leurs séquences diffèrent parfois très peu des histones majeures et leur expression s'effectue à un niveau basal tout au long du cycle cellulaire. Ces histones sont codées par des gènes uniques qui se répartissent dans le génome. Ils sont spécifiques de tissus ou ont des fonctions particulières.

Les histones sont des protéines essentielles. En effet, des expériences d'ARN interférence chez le nématode *Caenorhabditis elegans* ont démontré que la déplétion d'une ou plusieurs histones conduisent à des phénotypes de létalité embryonnaire, les embryons s'arrêtant de croître très tôt au stade 50 cellules. L'absence d'histones engendre également des défauts de condensation des chromosomes parfois accompagnés de ponts inter-chromosomiques. Une inhibition de l'expression des histones à un stade post-embryonnaire entraîne outre la stérilité, un phénotype de type « multivulve » pour les vers atteignant le stade adulte (Kodama *et al.*, 2002; Pettitt *et al.*, 2002).

Les histones forment le nucléosome :

Les histones sont les principaux constituants d'un complexe macromoléculaire appelé nucléosome. La structure cristallographique du nucléosome a été résolue à 1,9 Å de résolution

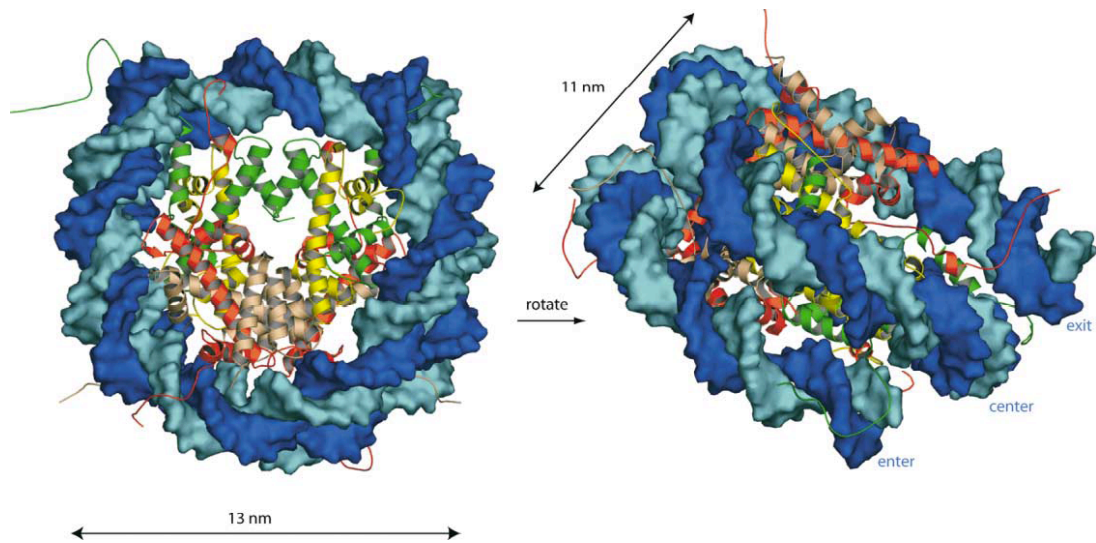
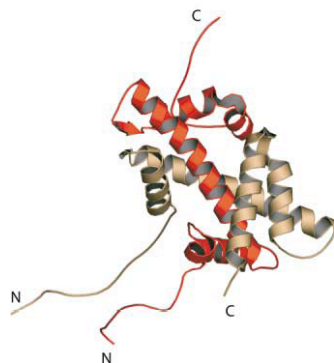


Figure 1 : Vue générale de la structure tridimensionnelle du nucléosome. Les 2 brins d'ADN se distinguent par des teintes de bleu différentes. L'ADN effectue 1,7 tour autour d'un octamère constitué de deux molécules de chacune des histones : H2a (en rouge), H2b (en rose), H3 (en vert) et H4 (en jaune). L'histone H1 est absente du modèle. Représentation d'après Khorasanizadeh (2004).

A H3-H4 tetramer



B H2A-H2B dimer



C Linker Histone

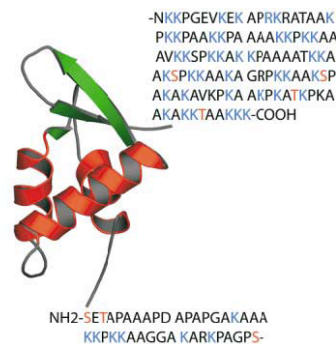


Figure 2 : Vues tridimensionnelles détaillées des paires d'histones composant l'octamère et de l'histone H1. **(A)** Tétramère d'histones H3 (en vert) et H4 (en jaune). **(B)** Dimère d'histones H2a (en rouge) et H2b (en orange). **(C)** Histone H1, appelée aussi histone linker. Les extrémités N- et C-terminales sont désordonnées et contiennent de nombreux résidus lysines et serines. Représentation d'après Khorasanizadeh (2004).

en associant des histones de xénope produites dans *Escherichia coli* à de l'ADN répétitif de satellite alpha (Davey *et al.*, 2002; Richmond & Davey, 2003; Khorasanizadeh, 2004)(**Figure 1**). Les histones s'assemblent en octamère contenant deux exemplaires de chacune des histones H2a, H2b, H3 et H4. Ces protéines présentent toutes une partie globulaire centrale appelée « histone-fold » qui est formée de 3 hélices α qui leur permettent d'interagir avec l'ADN. Au sein de l'octamère, les protéines se regroupent en tétramère de type H3-H4 et dimères H2a-H2b associant leur domaine « histone-fold » en motif de type « hand-shake » (**Figure 2**). L'octamère ainsi formé est ensuite entouré d'une portion d'ADN de 146 paires de bases qui fait environ 1,7 tour autour de ce noyau protéique. Enfin, une histone de type H1 interagit à l'extérieur avec les régions entrantes et sortantes de l'ADN afin de retenir l'ADN autour du noyau octamérique. Pour cela, l'histone H1 possède un domaine globulaire central contenant un motif de liaison à l'ADN de type « winged helix-turn-helix », motif composé d'un repliement de type $\alpha\beta\alpha\alpha\beta\beta$ (Khorasanizadeh, 2004).

Le nucléosome représente l'entité de base de la chromatine. La résolution de la structure du nucléosome a permis de mettre en évidence des interactions entre histones mais également avec l'ADN. C'est au niveau du sillon mineur de l'ADN que se concentrent les interactions avec les histones. Des interactions directes sont aussi observées entre les différentes histones. H3 forme des contacts directs avec H2a, tandis que les deux dimères H2a-H2b se positionnent de façon à ce que les histones H2a puissent interagir ensemble. La protéine H1, appelée aussi « linker », a une fonction d'organisation supérieure en intervenant dans la compaction des nucléosomes adjacents (Khorasanizadeh, 2004).

Assemblage des histones pour former le nucléosome :

Les histones majeures ou réplication-dépendantes sont incorporées au sein du nucléosome uniquement au cours de la phase S du cycle cellulaire de manière couplée à la réplication. L'assemblage du nucléosome s'effectue grâce à des chaperonnes qui accompagnent les histones jusqu'à leur dépôt sur l'ADN. Elles fonctionnent en coordination avec des facteurs de remodelage de la chromatine qui modifient l'organisation dynamique du nucléosome pour le plus souvent aider à l'activation ou la répression de l'expression des gènes. Dans ce cas, les facteurs de remodelage de la chromatine agissent en coordination avec les chaperonnes afin de procéder à un dépôt précis des histones (Sarma & Reinberg, 2005). Selon le modèle généralement admis, l'assemblage du nucléosome se déroule en deux étapes.

Un tétramère d'histones H3 et H4 est d'abord déposé sur l'ADN, puis c'est le tour des deux dimères de type H2a-H2b. Le dépôt des histones H3 et H4 sur le nouvel ADN est induit par le facteur d'assemblage CAF1 (Chromatin Assembly Factor 1) assisté par une chaperonne d'histone ASF1 (Annunziato, 2005). Le couplage avec la réplication se ferait par le recrutement de CAF1 au niveau de la fourche de réplication par le facteur PCNA (Proliferating Cell Nuclear Antigen). Ce complexe protéique joue le rôle d'une pince glissant le long de l'ADN et de facteur de processivité de l'ADN polymérase (Sarma & Reinberg, 2005). Le dépôt des histones H2a et H2b est quant à lui assisté par le facteur NAP1 (Nucleosome Assembly Protein 1) (Adams & Kamakaka, 1999; Mello & Almouzni, 2001). Enfin, l'assemblage du nucléosome se termine par l'addition d'une histone H1 aux brins entrants et sortants de l'ADN entourant l'octamère. Le nucléosome est alors formé.

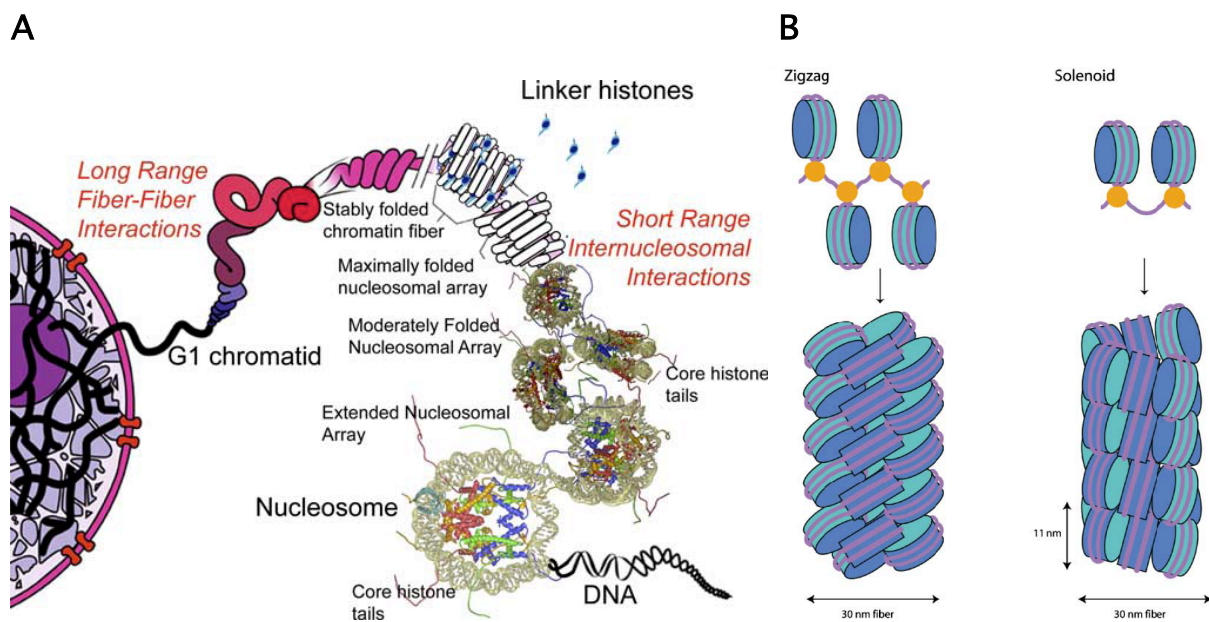


Figure 3 : Fonction du nucléosome dans la compaction de l'ADN. (A) Représentation schématique des différents niveaux de compaction de l'ADN eucaryotique. De longs enchaînements de nucléosomes sont compactés par le biais d'interactions à courte ou à longue distance dans des fibres dont l'architecture n'est pas encore connue. (B) Schéma représentatif des différents modes proposés pour la compaction de l'ADN en fibres de 30 nm. L'octamère d'histones est représenté en bleu, l'histone H1 en orange et l'ADN en violet. Le premier modèle proposé est un empiement en zig-zag de nucléosomes adjacents placés de manière alternative. Dans le second modèle, 6 nucléosomes sont assemblés par tour d'hélice pour former une structure de type solénoïde. D'après Chakravarthy *et al.* (2005) et Khorasanizadeh (2004).

2 FONCTIONS DU NUCLÉOSOME

2.1 Compaction de l'ADN

La première fonction du nucléosome est de compacter le génome (**Figure 3**). Celle-ci est en effet nécessaire pour maintenir les deux mètres linéaires d'ADN génomique dans un noyau de cellule d'environ $1000\ \mu\text{M}^3$ de volume. Le nucléosome représente le premier niveau de compaction de l'ADN, il permet d'en réduire la longueur de 5 fois. La chromatine ainsi formée adopte une structure en « collier de perles » de 10 nm de diamètre. Ensuite, l'interaction entre les histones H1 des nucléosomes mais également des interactions inter-nucléosomes impliquant les queues des autres histones permettent la formation de fibres de 30 nm de diamètre. Ces fibres représentent une matrice pour la plupart des processus nucléaires. À la mitose, la compaction est encore plus poussée grâce à l'action d'enzymes ATP-dépendantes comme la condensine et topoisomérase II. La chromatine se condense alors en chromosomes mitotiques qui pourraient ne plus contenir d'histones H1 pour permettre un super-enroulement des nucléosomes dans les chromosomes (Khorasanizadeh, 2004).

2.2 Régulation de l'expression du génome, information épigénétique

Outre sa fonction essentielle dans la compaction de l'ADN en chromatine, le nucléosome a des fonctions plus complexes : c'est une unité de régulation du génome. En plus d'une régulation au niveau de l'ADN, il existe des signaux additionnels au niveau de la chromatine qui permettent une régulation de l'expression des gènes, de la réplication, de la recombinaison et de la réparation de l'ADN. Ces signaux correspondent à des modifications dans la structure de la chromatine qui ont alors un impact sur l'expression des gènes et le maintien de la stabilité du génome. Ils constituent un message appelé information épigénétique, différent de celui qui est porté par la séquence de l'ADN. Ces modifications de la structure de la chromatine sont induites par des complexes enzymatiques responsables de changements ciblant les histones, l'ADN ou le nucléosome en entier.

2.2.1 Ajout de modifications post-traductionnelles sur les histones

Une première façon de modifier un nucléosome consiste en l'ajout d'une ou plusieurs modifications post-traductionnelles sur une ou plusieurs histones qui le composent. Cet ajout pourra modifier les caractéristiques de charge ou d'encombrement de ce nucléosome, et ainsi opèrera des changements dans les diverses interactions, ce qui aura pour conséquence de renforcer ou rendre moins stables les nucléosomes et donc rendre plus ou moins accessible l'ADN. Ces modifications post-traductionnelles peuvent s'effectuer de manière séquentielle, les unes favorisant parfois l'apparition d'autres, ou en combinaison. Les modifications ciblent habituellement les queues N- et C-terminales des histones, mais elles peuvent également cibler leur domaine globulaire. Chaque histone est susceptible d'être modifiée. La **figure 4** représente l'ensemble des modifications post-traductionnelles décrites à ce jour.

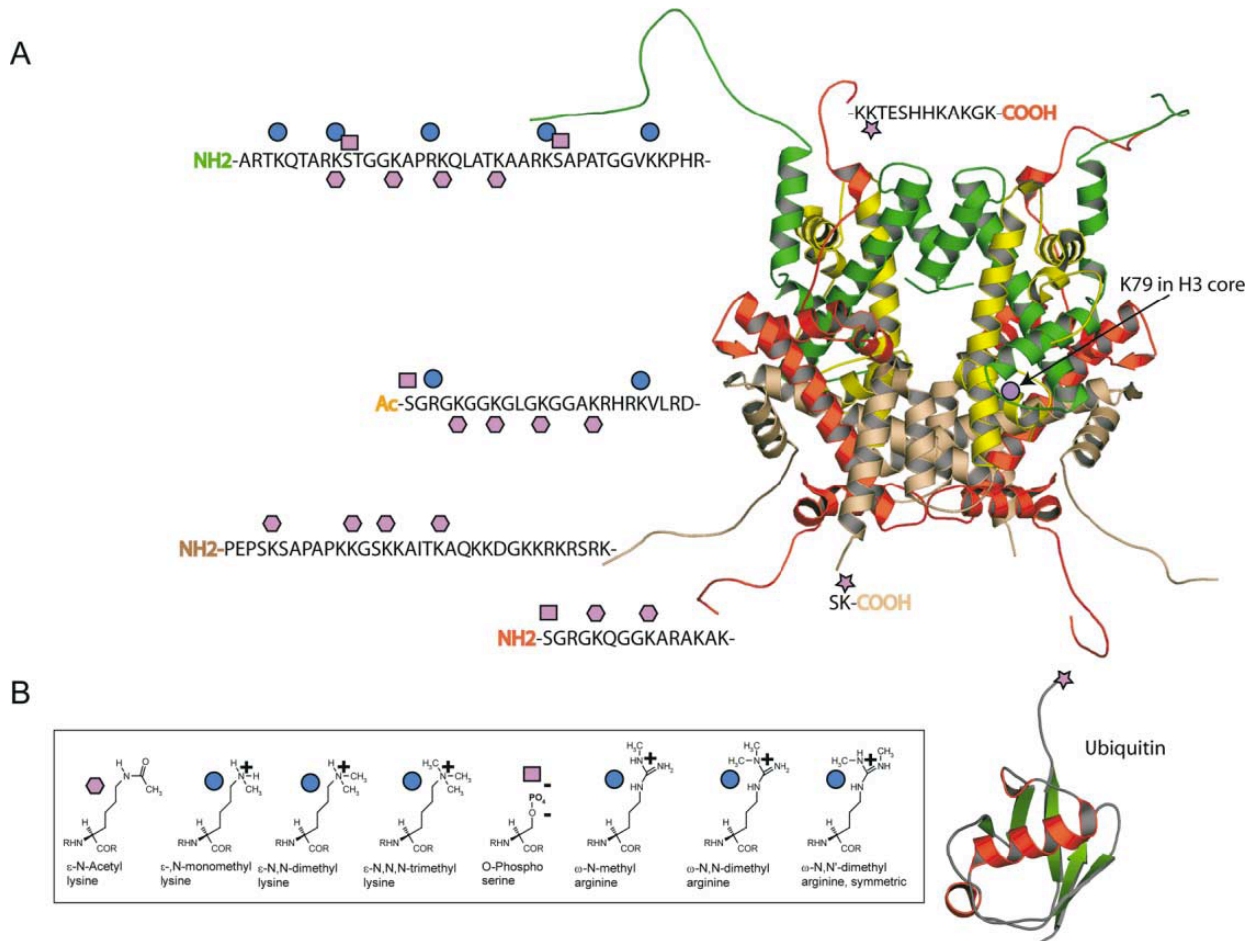


Figure 4 : Représentation des différentes modifications post-traductionnelles retrouvées dans l'octamère d'histones formant le nucléosome. **(A)** Les sites de modifications sont indiqués au niveau des histones H2a (en rouge), H2b (en rose), H3 (en vert) et H4 (en jaune). Par souci de clarté, les modifications ne sont indiquées que sur une seule copie de chaque histone. **(B)** Détail des différentes modifications portées par les résidus des histones. D'après Khorasanizadeh (2004).

2.2.1.1 Les différents types de modification des histones (Khorasanizadeh, 2004)

L'acétylation de lysines entraîne la perte de leur charge positive ce qui induit une diminution de la force d'interaction entre l'histone modifiée et l'ADN. L'accès à la machinerie de transcription est alors facilité ce qui augmente son efficacité. Ces modifications peuvent également être reconnues par des protéines possédant un domaine de fixation qui leur est spécifique appelé bromo-domaine. Ces protéines sont le plus souvent impliquées dans la régulation de la réplication, la réparation ou le dépôt d'histones sur l'ADN.

Les méthylations simples, doubles, ou triples touchent les lysines et les arginines. Concernant les lysines, leur méthylation entraîne soit une inhibition de la transcription par condensation de la chromatine et formation d'hétérochromatine via le recrutement de la protéine HP1 (Heterochromatin Protein 1), soit une activation de la transcription en recrutant des activateurs de transcription qui empêchent alors la fixation de la protéine HP1 et qui maintiennent le niveau de transcription. Une même modification peut donc avoir des effets contraires. Dans le cas des résidus arginines, on note une corrélation entre méthylation et état d'activation de la transcription ; ceci est dû au fait que la méthylation favorise l'acétylation des résidus lysine qui a un effet activateur de la transcription (voir plus-haut).

La phosphorylation des résidus sérine ou thréonine peut être impliquée dans la condensation des chromosomes à la mitose. Dans d'autres cas, elle régule la formation des autres modifications en augmentant l'acétylation des lysines ou en neutralisant l'effet des methyl-lysines.

L'ubiquitylation des lysines est importante pour la méthylation des lysines et l'activation de la transcription.

La sumoylation des lysines engendre une régulation négative de la transcription par le recrutement de complexes de modification.

L'ADP ribosylation des résidus glutamate ou aspartate consiste en l'addition d'un polymère d'ADP-ribose plus ou moins long. La fonction de cette modification n'a jusqu'ici pas été clairement définie. Elle semble influencer l'organisation de la chromatine par décondensation des fibres de chromatine. Elle pourrait limiter l'accès de certaines protéines à l'ADN ou jouer le rôle de signal pour l'établissement d'autres modifications. Certains de ces rôles semblent parfois contradictoires (Ausio *et al.*, 2001).

2.2.1.2 Régulation de l'ajout et du retrait des modifications touchant les histones

L'incorporation des modifications se fait grâce au recrutement des complexes enzymatiques au niveau des histones à modifier. Le recrutement s'effectue à l'aide de facteurs supplémentaires, ou de modifications des histones, ou par l'intermédiaire de petits ARNs non codants. Ces ARNs non codants, ou siARNs, induisent des modifications au niveau de la chromatine qui ont pour conséquence une inactivation de la transcription de la zone ciblée. La chromatine alors silencieuse est appelée hétérochromatine. Sa formation touche le plus souvent des régions particulières du chromosome contenant beaucoup de répétitions tels les centromères, les télomères ou les éléments transposables (Lippman & Martienssen, 2004). Les siARNs guident la méthylation des histones fixées sur ces régions, ce qui a pour conséquence le recrutement de la protéine HP1 impliquée dans la formation de l'hétérochromatine condensée (Peterson & Laniel, 2004).

La plupart des modifications post-traductionnelles ajoutées sur les histones sont réversibles. En effet, certaines enzymes spécifiques sont capables de les retirer, ce qui laisse à la cellule la possibilité d'éteindre les signaux qu'elles représentent (Khorasanizadeh, 2004).

2.2.2 Modifications de la structure et composition du nucléosome

2.2.2.1 Remodelage de la structure du nucléosome dépendant de l'ATP

Les nucléosomes constituent une vraie barrière pour l'accès de l'ADN qui entoure les histones. L'expression et la régulation du message génétique deviennent dès lors très difficiles. Des complexes enzymatiques « SWI-SNF » utilisent l'énergie de l'ATP pour restructurer l'ADN nucléosomal et le rendre accessible. De manière plus générale, ils modulent l'accessibilité de certaines régions de l'ADN et activent ou répriment l'expression des gènes. En déstabilisant les interactions entre l'ADN et l'octamère d'histones, ils facilitent les mouvements de l'un par rapport à l'autre et un remodelage du nucléosome. Ces enzymes modifient l'organisation du nucléosome en induisant par exemple le glissement ou « sliding » d'octamères au niveau de séquences d'ADN adjacentes auparavant libres (**Figure 5A**). Dans ce cas, l'octamère est déplacé par rapport à l'ADN. Ainsi, des régions auparavant emprisonnées dans le nucléosome deviennent accessibles et inversement, des régions

auparavant accessibles deviennent masquées. Le remodelage peut également s'effectuer par la création de segments d'ADN accessibles à la surface même du nucléosome en induisant la formation d'une boucle sortant du nucléosome. Ce phénomène appelé « DNA looping » (Fan *et al.*, 2003) est initié par un glissement du nucléosome puis recapture d'ADN plus loin dans la séquence pour finir l'enroulement autour de l'octamère (**Figure 5B**) (Khorasanizadeh, 2004).

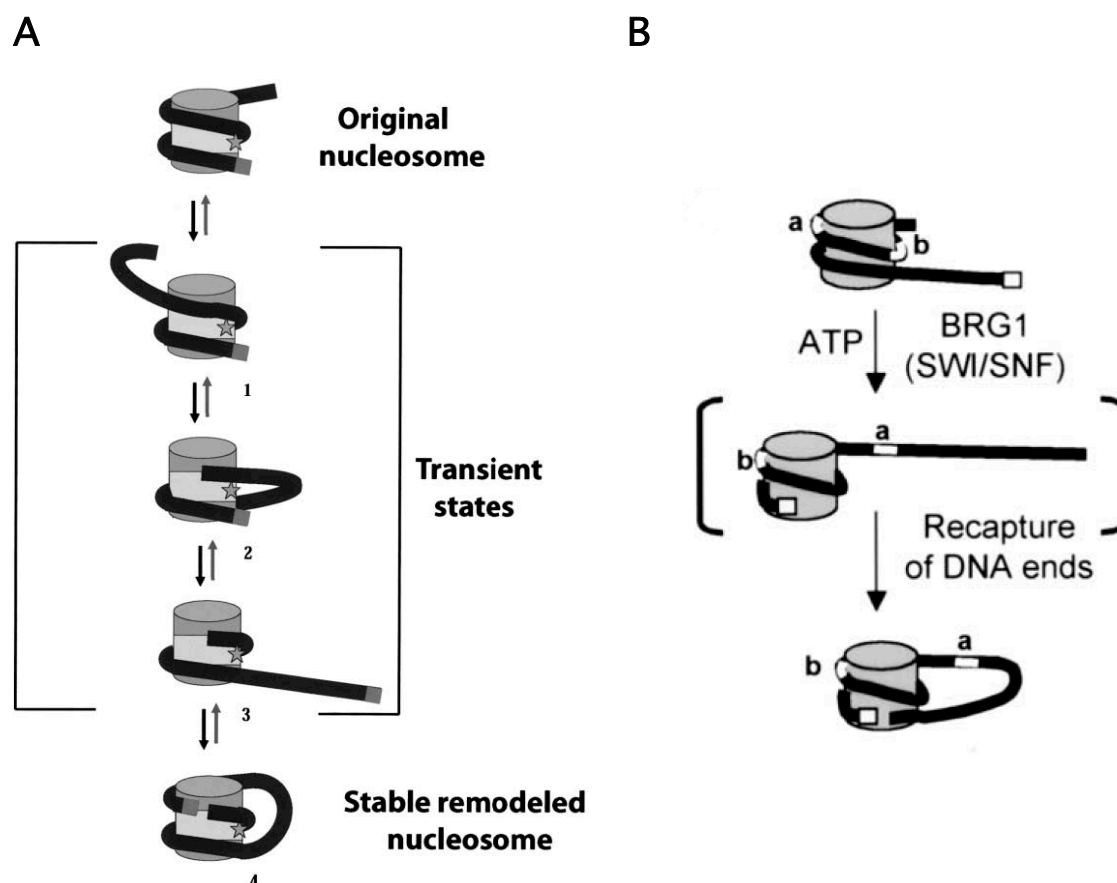


Figure 5 : Remodelage de la structure du nucléosome dépendant de l'ATP.

(A) Modèle de remodelage de la structure du nucléosome par glissement ou « sliding » du nucléosome. Le déroulement de l'ADN du nucléosome par le facteur SWI/SNF est initié à un bord du nucléosome (1). Une boucle instable allant jusqu'à 50 nucléotides est formée transitoirement lorsque l'ADN s'enroule sur l'octamère (2). L'ADN de cette boucle ne peut repartir en arrière car il en est empêché du fait de la fixation du facteur SWI/SNF. Cela résulte en une propagation directionnelle de l'ADN de la boucle et a pour conséquence un glissement du nucléosome (3). L'ADN sorti refixe l'octamère de manière à reconstituer un enroulement fort autour d'un nucléosome alors remodelé. Les dimères H3/H4 sont représentés en gris clair tandis que les dimères H2a/H2b sont en gris foncé. L'étoile symbolise l'axe de la dyade du nucléosome. (B) Modèle de remodelage de la structure du nucléosome par « DNA looping ». Les facteurs BRG1 et SWI/SNF exposent le site (a) de l'ADN nucléosomal par glissement de l'octamère sur l'ADN. L'extrémité de l'ADN exposé refixe ensuite l'octamère de manière à former une boucle stable. Le site (a) de l'ADN devient alors accessible.

D'après Fan *et al.* (2003) et Kassabov *et al.* (2003).

2.2.2.2 Incorporation de variants d'histones

Le dernier type de remaniement que peut subir le nucléosome est la substitution d'histones. Dans certains nucléosomes, une histone majeure va être remplacée par un variant du même type. La substitution d'une histone majeure par un variant mineur présentant des résidus différents entraîne des effets plus ou moins importants. Les interactions du nucléosome avec l'ADN peuvent être modifiées de manière à le stabiliser ou à le déstabiliser. L'incorporation du variant rendra donc le nucléosome plus ou moins mobile ce qui rendra l'ADN plus ou moins accessible à la machinerie de transcription ou à certains facteurs de régulation. La présence du variant peut également changer les résidus exposés à la surface du nucléosome et influencer les modifications post-traductionnelles. Par ces changements de modifications, le nucléosome modifié peut aussi recruter des facteurs de régulation ou ne sera plus en mesure de le faire.

L'incorporation de variants a donc pour conséquence de modifier les signaux codés par la structure du nucléosome et ainsi changer la fonction des nucléosomes, participant de cette manière à la régulation de l'expression des gènes.

2.2.2.2.1 Caractéristiques des variants d'histones

Les variants d'histones, également appelés histones réplication-indépendantes, sont codés par des gènes solitaires dispersés dans le génome. Leur expression se fait à un niveau basal tout au long du cycle cellulaire. Ils peuvent être spécifiques à certains tissus ou posséder des fonctions particulières. Certains présentent des séquences qui divergent significativement de celles des histones majeures, d'autres parfois très peu, comme dans le cas du variant H3.3 qui ne diffère que de 4 résidus de l'histone majeure H3 (Sarma & Reinberg, 2005; Kamakaka & Biggins, 2005), et d'autres pas du tout comme le variant H4-v.1 qui ne contient que des mutations silencieuses au niveau nucléotidique (Poirier *et al.*, 2005). Les variations concernent le plus souvent les domaines N- et C-terminaux et plus rarement le domaine globulaire central contenant le motif « histone fold » (**Figure 6**).

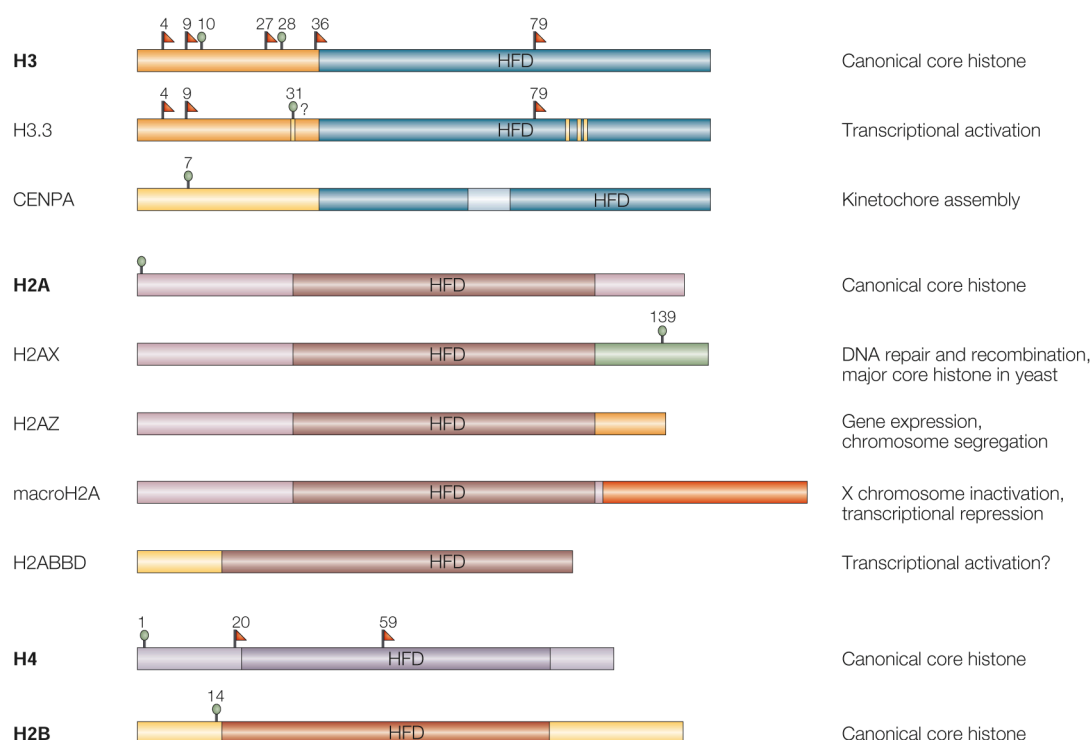


Figure 6 : Représentations linéaires des histones majeures et variants. Les histones contiennent un domaine histone-fold conservé (HFD) ainsi que des extrémités N- et C-terminales sujettes à de nombreuses modifications post-traductionnelles. Les sites majeurs de modifications sont indiqués ; les méthylations de lysines sont représentées par des drapeaux rouges tandis que les phosphorylations de sérines sont symbolisées par des disques verts. Le variant d'histone H3.3 diffère de 4 résidus de l'histone majeure (en jaune). Un de ces résidus semblerait être un site potentiel de phosphorylation. L'histone centromérique CENPA présente une extrémité N-terminale unique. De plus, elle contient une séquence responsable du ciblage de CENPA au niveau des centromères (en bleu ciel). Les variants des histones H2a diffèrent de l'histone majeure au niveau de leur extrémité C-terminale. Les fonctions qui ont été proposées pour les différents variants sont indiquées sur le côté. D'après Sarma & Reinberg (2005).

Malgré les différences de séquence entre histones majeures et variants, le nucléosome contenant le variant ne subit pas un remaniement de structure fondamental qui pourrait conduire à la désorganisation complète de sa structure, seules les fonctions en sont changées (Chakravarthy *et al.*, 2005). Le changement de séquence pourrait par exemple entraîner quelques variations dans la distribution des interactions au sein du nucléosome ou dans l'exposition des résidus, ce qui pourrait avoir comme conséquence d'affecter le profil de modifications et donc la fonction du nucléosome (Chakravarthy *et al.*, 2005).

2.2.2.2 Fonctions des variants d'histones, quelques exemples

Les variants d'histones peuvent tout d'abord agir de façon **indirecte**. En effet, leur incorporation dans le nucléosome implique le départ de l'histone qu'ils remplacent, et par conséquent la perte d'éventuelles modifications, de résidus exposés et d'interactions au sein

du nucléosome que le variant n'est plus capable de reproduire. Dans ce cas, le variant peut entraîner de manière indirecte la perte de la fonction initiale du nucléosome.

Les variants d'histones font de plus en plus l'objet d'études portant sur les fonctions **directes** qu'ils confèrent au nucléosome par les changements structuraux qu'ils induisent (**Tableau 1**).

Variant	Species	Chromatin effect	Function
H1 ^o	Mouse	Chromatin condensation	Transcription repression
H5	Chicken	Chromatin condensation	Transcription repression
SpH1	Sea urchin	Chromatin condensation	Chromatin packaging
H1t	Mouse	Open chromatin	Histone exchange, recombination?
MacroH2A	Vertebrate	Condensed chromatin	X-chromosome inactivation
H2ABbd	Vertebrate	Open chromatin	Transcription activation
H2A.X	Ubiquitous	Condensed chromatin	DNA repair/recombination/transcription repression
H2A.Z	Ubiquitous	Open/closed chromatin	Transcription activation/repression, chromosome segregation
SpH2B	Sea urchin	Chromatin condensation	Chromatin packaging
CenH3	Ubiquitous		Kinetochore formation/function
H3.3	Ubiquitous	Open chromatin	Transcription

The species distribution and likely functions of major histone variants are shown.

Tableau 1 : Tableau récapitulatif des différents variants majeurs d'histones connus. Leur distribution en fonction de l'espèce, l'effet de leur incorporation au niveau de la chromatine et leurs probables fonctions sont indiqués (Kamakaka & Biggins, 2005).

- Régulation de la transcription des gènes

L'incorporation de certains variants au niveau de certains gènes permet de réguler leur transcription. Les variants capables d'activer la transcription vont pouvoir, grâce à leur séquence ou leurs modifications post-traductionnelles déstabiliser le nucléosome et ainsi permettre une meilleure ouverture du promoteur pour faciliter l'entrée de l'ARN polymérase. Ainsi, le variant H3.3 est fréquemment retrouvé dans les loci actifs en transcription car il est riche en modifications post-traductionnelles qui favorisent la transcription active (Sarma & Reinberg, 2005). Dans d'autres cas, l'apport de variants entraînera plutôt une répression de l'expression des gènes par condensation de la chromatine ou recrutement de répresseurs de la transcription (Kamakaka & Biggins, 2005).

- Établissement de l'hétérochromatine

Certains variants sont localisés dans des régions données du génome car ils y ont un rôle bien précis. L'histone H2AZ est notamment retrouvée aux alentours des zones d'hétérochromatine chez la levure car elle participe à l'établissement de barrières dont la

fonction est de bloquer la dispersion du phénomène d'hétérochromatinisation (Kamakaka & Biggins, 2005).

- Inactivation du chromosome X

On retrouve en grande quantité le variant MacroH2A dans les chromosomes X inactivés de femelles de mammifères. Il participe à l'inactivation de ces chromosomes en bloquant l'accès des facteurs de transcription et en interférant avec l'activité des facteurs de remodelage de la chromatine. Il semblerait que le recrutement de MacroH2A s'effectue par l'intermédiaire d'un petit ARN appelé Xist (Sarma & Reinberg, 2005).

- Stabilité du génome

L'incorporation de variants d'histone permet de garantir, dans certains cas, une certaine stabilité du génome, notamment en contribuant au bon fonctionnement de mécanismes impliquant la molécule d'ADN comme la réplication, la mitose, la recombinaison ou encore en assurant un bon déroulement de la réparation des lésions (Kamakaka & Biggins, 2005).

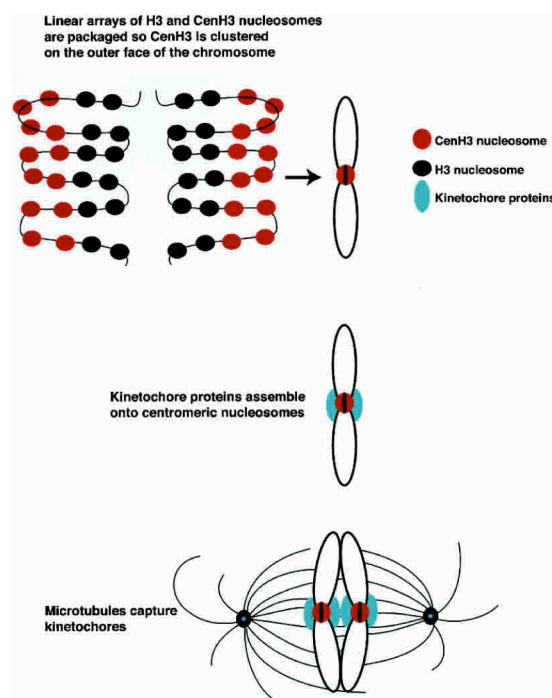


Figure 7 : Fonction du variant de l'histone H3 CenH3 ou CENPA dans la ségrégation des chromosomes lors de la mitose. Les histones H3 et CenH3 se fixent à l'ADN centromérique de manière à ce que, lors d'une compaction ultérieure, les nucléosomes contenant les variants CenH3 soient exposés sur la face externe du chromosome. Ils vont faciliter la fixation de protéines qui forment la structure du kinétochore. Les microtubules qui émanent des pôles de la cellule vont capturer les kinétochores et engendrer la ségrégation chromosomique. Selon Kamakaka & Biggins (2005).

- Fidélité de la ségrégation des chromosomes lors de la mitose

L'histone centromérique CENPA (ou CenH3), correspondant à un variant de H3, participe à la formation du centromère. Elle est requise lors de la mitose pour que la ségrégation des chromosomes s'effectue de manière précise (**Figure 7**) (Kamakaka & Biggins, 2005). Des études ont également démontré que le variant H2AZ encadrant l'hétérochromatine centromérique pouvait interagir avec des protéines impliquées dans la ségrégation des chromosomes, la réplication et le « checkpoint control » (Kamakaka & Biggins, 2005).

- Efficacité de la réparation de l'ADN

Le variant H2AX est impliqué dans la réparation de l'ADN lors de la cassure d'une chromatide ou cassure double brin. Aux alentours de la zone endommagée, les variants d'histones H2AX seront préférentiellement phosphorylés, ce qui permet de recruter au niveau de la cassure la cohésine, une protéine impliquée dans la réparation de l'ADN. Il s'agit en fait d'un complexe protéique qui a pour fonction le maintien des chromatides sœurs entre elles en vue de faciliter la réparation (Lowndes & Toh, 2005).

- Différenciation cellulaire

Dans une cellule, l'incorporation de variants d'histones au niveau de la chromatine permet de modifier le profil d'expression des gènes, ce qui peut conduire à modifier son type cellulaire. Certains variants sont amenés à remplacer des histones majeures préexistantes pendant le développement et la différenciation mais également lors de la transformation en cellules cancéreuses (Govin *et al.*, 2004). Le cas de la spermatogenèse chez les mammifères est un exemple de processus de différenciation cellulaire où s'effectuent les plus importants remaniements de la chromatine. Les histones sont progressivement remplacées par de petites protéines basiques, les protamines, en même temps que se déroule une synthèse massive de variants d'histones incluant des variants spécifiques des tissus testiculaires. En parallèle s'effectuent des modifications post-traductionnelles spécifiques à chaque étape du processus de différenciation (**Figure 8**) (Govin *et al.*, 2004). Le rôle présumé des variants tissus-spécifiques est la condensation et la réorganisation du génome (Govin *et al.*, 2004).

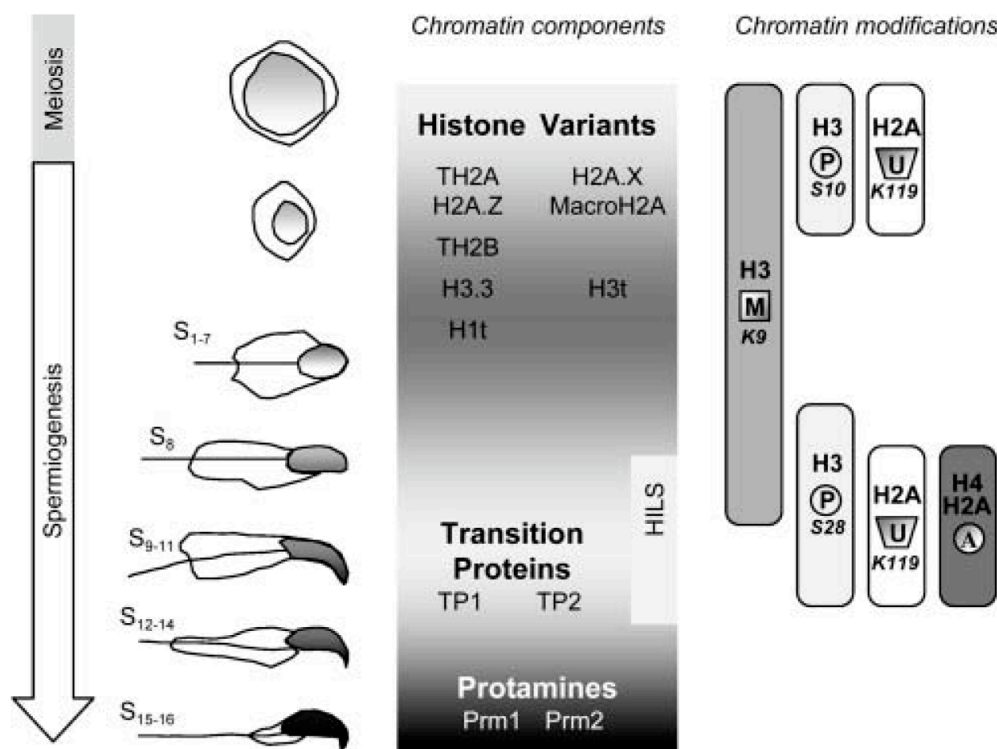


Figure 8 : Variation des constituants de la chromatine au cours de la spermatogénèse. Les composants majeurs et leurs modifications post-traductionnelles sont représentés. Les variants d'histones sont incorporés lors de la méiose sauf pour le variant de l'histone H1 appelé H1LS, dont l'expression est retardée. Au cours de la phase tardive de la spermatogénèse, les histones sont progressivement remplacées par des protéines extrêmement basiques, des protéines de transition et des protamines. Ces modifications conduisent à un remodelage global du noyau qui s'allonge et se compacte pour former le noyau très caractéristique des spermatozoïdes. La distribution temporelle des modifications post-traductionnelles majeures est également représentée (A, acétylation; U, ubiquitination; M, méthylation; P, phosphorylation). D'après Govin *et al.* (2004).

L'incorporation de variants d'histones dans les nucléosomes leur confère des fonctions spécifiques qui régissent l'expression du génome. Ces variants peuvent à leur tour subir des modifications post-traductionnelles qui apportent des signaux supplémentaires et augmentent le degré de complexité du message transmis au travers du nucléosome (**Figure 6**). L'étude de la composition des nucléosomes contenant des variants n'a pour l'instant que permis de mettre en évidence la coexistence de variants de même type. Jamais un variant n'a été trouvé associé à l'histone majeure qui lui correspond. De plus, à ce jour on ne connaît pas de nucléosomes contenant des variants remplaçant deux types d'histones différentes. La distribution des variants variera donc peu au niveau d'un nucléosome particulier, alors qu'elle semblera très variable à l'échelle de la chromatine. En effet, les nucléosomes contenant les variants peuvent se retrouver plus ou moins dilués dans des nucléosomes composés d'histones majeures (Kamakaka & Biggins, 2005).

2.2.2.2.3 Régulation de l'expression des variants d'histones

En général, les variants d'histones sont exprimés à un niveau basal tout au long du cycle cellulaire mais dans certains cas, ils ont acquis une sensibilité à d'autres voies de signalisation comme le développement embryonnaire ou la différenciation cellulaire. Ainsi, comme il l'a été précisé précédemment pour les variants spécifiques aux testicules, une synthèse accrue de ces protéines est induite lors de la spermatogenèse (Govin *et al.*, 2004).

2.2.2.2.4 Incorporation des variants dans les nucléosomes

L'incorporation des histones majeures (réplication-dépendantes) à l'ADN s'effectue uniquement au cours de la phase S du cycle cellulaire, tandis que celle des variants d'histones peut également se dérouler durant les phases G1 et G2 (Annunziato, 2005). L'incorporation des variants peut se produire lors de la réplication, ou de la transcription. Parfois l'incorporation d'histones peut se produire par simple échange en dehors de ces étapes (**Figure 9**). Chaque variant est incorporé au nucléosome par le biais d'un système d'assemblage qui lui est spécifique et différent de celui qui prend en charge les histones majeures. Ainsi, le variant H3.3 a été isolé dans un complexe d'assemblage de la chromatine contenant le facteur HIRA tandis que l'incorporation d'H3.1, l'histone majeure lui correspondant, fait intervenir le facteur CAF1 (voir plus haut) (Kamakaka & Biggins, 2005).

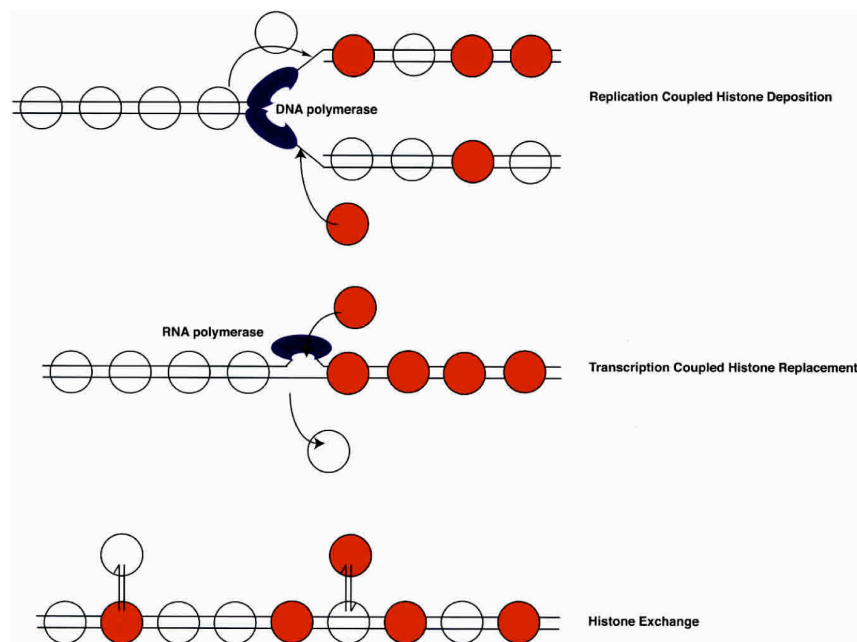


Figure 9 : Représentation des différents modes d'incorporation des nucléosomes contenant des variants d'histones (en rouge) au niveau de la chromatine. Cette incorporation peut être couplée à la réplication de l'ADN (en haut), ou à la transcription (au milieu). Parfois elle s'effectue par simple échange d'histones (en bas). Selon Kamakaka & Biggins, (2005).

2.3 Transmission et modulation de l'information épigénétique

2.3.1 Le code histone

L'idée de l'existence d'un code histone a récemment émergé. Pour chaque fonction se dessine un code histone spécifique. Il résulte de la combinaison de différentes modifications au niveau de la chromatine et a pour conséquence un effet biologique donné. Ainsi, la phosphorylation du variant d'histone H2AX combinée à la méthylation des histones H3 et H4 représenterait un code spécifique caractéristique de la fonction de réparation de l'ADN (Hassa & Hottiger, 2005; Lowndes & Toh, 2005). La spermatogenèse, un événement de différenciation cellulaire faisant intervenir de gros remaniements de chromatine, est également caractérisée par un code histone qui lui est spécifique (Govin *et al.*, 2004).

2.3.2 Héritage de l'information épigénétique

Afin de conserver les signaux de régulation du génome lors de la division cellulaire, il faut transférer la structure de la chromatine au cours des générations (Annunziato, 2005) et donc reconstituer, lors de la synthèse de la chromatine, un nucléosome présentant une structure identique à l'originel, qui portera le même message et la même fonction. Ce nucléosome devra donc présenter les mêmes modifications. L'assemblage des nucléosomes couplé à la réplication débute par le dépôt des histones H3 et H4 puis viennent H2a et H2b sous la forme de deux dimères et enfin H1. Il a été montré que les histones parentales sont ségrégées des deux côtés de la fourche au cours de la réplication. Ainsi, les modifications peuvent être héritées par les deux brins filles. Pour la répartition des histones parentales, il semblerait que les histones se partagent en hétéro-tétramères d'histones H3 et H4 ou octamères d'histones, plutôt que hétéro-dimère H3-H4 ou demi nucléosome (H2a, H2b, H3, H4) (**Figure 10**). L'héritage des modifications post-traductionnelles sur les nucléosomes néo-assemblés se ferait grâce aux enzymes de modifications qui seraient recrutées par les modifications présentes sur les nucléosomes parentaux ségrégés (Annunziato, 2005). L'incorporation des variants d'histones s'effectuant tout au long du cycle cellulaire, le problème de leur héritage sera résolu après la phase S où ils prendront place dans les régions du génome où ils sont requis.

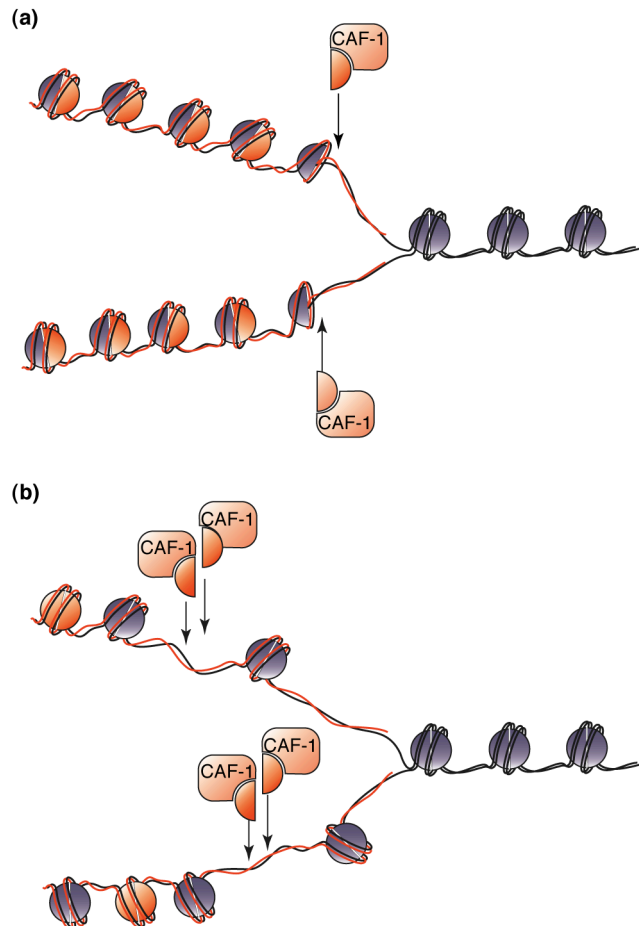


Figure 10 : Deux modèles alternatifs sont proposés pour l'assemblage du nucléosome en aval de la fourche de réplication. **(a)** Dans le modèle d'assemblage semi-conservatif, les octamères se divisent lors de la réplication et chaque moitié (en bleu) est distribuée aux brins filles. Les dimères H2a-H2b ont été omis par souci de clarté. Les chaperonnes d'histones comme le facteur d'assemblage CAF-1 insèrent des histones néo-synthétisées (en rouge) pour compléter l'octamère. **(b)** Dans le mode d'assemblage conservatif, des octamères intacts se répartissent entre les brins filles. Les chaperonnes d'histones forment des octamères complets aux endroits où l'ancien octamère manque. D'après ce modèle, les dimères néo-synthétisés H3-H4 sont insérés par l'action successive de deux complexes CAF-1. D'après Henikoff *et al.* (2004).

2.3.3 Dynamique de la chromatine

Les diverses modifications de la structure de la chromatine constituent des signaux dont la combinaison conduit à une information qualifiée d'épigénétique qui va alors coder des fonctions biologiques comme la régulation de l'expression des gènes, la réparation de l'ADN, la formation du centromère et de l'hétérochromatine.

Mais cette information épigénétique n'est pas délivrée de manière continue par la chromatine. En effet, les gènes régulés sont exprimés de façon discontinue. De même, la réparation d'ADN n'est requise que ponctuellement lors de dommages à l'ADN. Ainsi le flux et la nature des informations codées par la structure de la chromatine est très variable et ceci grâce à une certaine dynamique de la structure de la chromatine et plus particulièrement à une

dynamique de la structure du nucléosome qui « dialogue » avec la machinerie cellulaire. En changeant de structure, il reçoit et émet des informations ce qui provoque globalement un changement des propriétés de la chromatine. La dynamique du nucléosome correspond à un changement en composition, en structure, et un changement de localisation sur l'ADN qui s'effectue en permanence. L'ADN empaqueté dans le nucléosome n'est pas accessible aux protéines qui veulent « lire » ou « copier » la séquence nucléotidique ; par conséquent, une telle structure doit être dynamique et capable de présenter des transitions durant lesquelles le nucléosome est « ouvert » ou « fermé » afin de rendre l'ADN plus ou moins accessible (**Figure 11**). Cette dynamique est engendrée par l'ajout ou le retrait de modifications post-traductionnelles ou de variants d'histones mais également par les facteurs de remodelage de la chromatine qui déstabilisent, déforment ou déplacent le nucléosome.

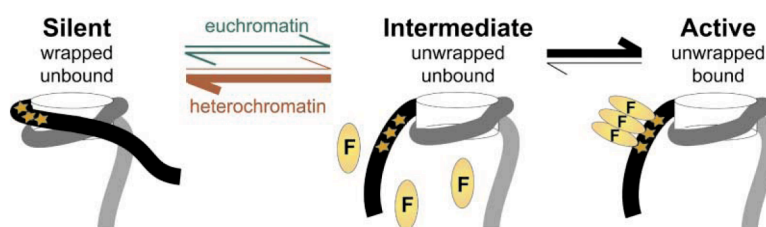


Figure 11 : Modèle d'exposition de l'ADN nucléosomal à un facteur de transcription. Les sites de fixation du facteur de transcription, symbolisés par des étoiles, sont dissimulés du fait de l'enroulement autour de l'octamère d'histones. Le nucléosome formé est alors qualifié de silencieux, il compose l'hétérochromatine. Certains remaniements de la chromatine permettent d'induire le déroulement de l'ADN nucléosomal, ce qui a pour conséquence de rendre accessibles les sites de fixation du facteur de transcription (F). Le nucléosome entre alors dans un état intermédiaire où la portion d'ADN nucléosomal est accessible mais non liée aux facteurs de transcription. Lorsque ces facteurs fixent enfin l'ADN nucléosomal rendu accessible, le nucléosome est qualifié d'actif et constitue l'euchromatine. D'après Ahmad & Henikoff (2002).

3 SYNTHÈSE DES HISTONES MAJEURES ET RÉGULATION DE LEUR EXPRESSION

3.1 Nécessité d'une régulation de la production des histones

Nous avons vu que la chromatine est plus qu'une simple forme de compaction de l'ADN ; de nombreuses autres fonctions liées à ses variations structurales peuvent en découler et jouer des rôles cruciaux pour la cellule. Les histones qui composent la chromatine sont donc des protéines essentielles qui sont à la base même de l'expression du génome et donc du fonctionnement de la cellule. Nous avons jusqu'à présent décrit les variants d'histones

impliqués dans l'information épigénétique transmise par la chromatine ; ceux-ci ne représentent que des histones mineures incorporées dans certains cas particuliers. Au cours de cette thèse nous nous sommes plutôt intéressés aux histones majeures dites « réplication-dépendantes » considérées comme « les briques de base » de la chromatine.

Au cours de la phase S du cycle cellulaire, lorsque l'ADN génomique est répliqué il est impératif pour la cellule de disposer de grandes quantités d'histones afin d'empaqueter l'ADN néo-synthétisé. En effet, la cellule doit orchestrer au cours de la phase S l'incorporation d'environ 30 millions de nucléosomes chez l'homme (Gunjan *et al.*, 2005). Une production massive d'histones est donc nécessaire au cours de cette phase du cycle cellulaire. Cependant, la cellule doit parfaitement contrôler cette production tant du point de vue des quantités produites que du moment où se fait la production car un excès d'histones présente un réel danger pour la cellule. Un équilibre précis doit être trouvé entre les quantités d'histones et d'ADN produites et il faut impérativement veiller à coupler les deux événements. En effet, les histones sont des protéines fortement basiques qui fixent très avidement et de manière non spécifique les acides nucléiques d'où l'importance de maintenir un bon rapport entre histones et ADN. Ainsi, une production des histones en dehors de la phase S induit des arrêts du développement chez la Drosophile (Sullivan *et al.*, 2001). Le problème est délicat car même au cours de la phase S la vitesse de réplication de l'ADN varie. Il faut donc que la cellule soit capable de faire face à ces variations de synthèse et de s'y adapter pour la production d'histones. De même, en présence de lésions, la réplication est ralentie, ce qui nécessite un ralentissement de la production d'histones. Pour résoudre ces problèmes, la cellule a développé des stratégies régulatrices qui couplent la synthèse d'histones avec la progression du cycle cellulaire et la réplication en cours afin que les histones produites soient utilisées pour empaqueter l'ADN et ne restent pas libres dans le cytoplasme (Gunjan *et al.*, 2005).

Un autre danger provient également du non-respect de la stœchiométrie dans la production d'histones qui interfère avec la fidélité de ségrégation des chromosomes à la mitose et la régulation de la transcription (Gunjan *et al.*, 2005). Des expériences réalisées sur la levure ont montré qu'un déséquilibre de l'expression des différentes histones engendrait des problèmes sévères dans la transcription (Clark-Adams *et al.*, 1988). Une disruption de la chromatine est observée lorsque l'on délète ou mute des gènes d'histones (Norris *et al.*, 1988). Enfin certains chromosomes peuvent être perdus lorsque l'on sur-exprime soit le couple H2a-H2b ou les autres histones (Meeks-Wagner & Hartwell, 1986). Chez la levure, il a été récemment proposé que la protéine Rad 53, impliquée dans la réponse aux dommages de

l'ADN, jouerait également un rôle dans un mécanisme de surveillance gérant l'excès d'histones en induisant leur dégradation (Gunjan & Verreault, 2003).

3.2 Régulation de la synthèse d'histones dans la levure et les plantes

La cellule doit produire de façon massive et ciblée au cours de la phase S de son cycle cellulaire, les quantités stœchiométriques d'histones nécessaires et suffisantes pour emballer l'ADN en cours de synthèse. En dehors de la régulation de l'excès d'histones chez la levure par induction de leur dégradation (voir § précédent), les organismes ont développé des mécanismes de régulation transcriptionnels ou post-transcriptionnels pour obtenir les conditions adéquates de production d'histones assurant une bonne division cellulaire. Il en résulte que le taux de synthèse d'histones est majoritairement régulé par le taux des ARN messagers correspondants.

Ainsi, chez la levure, les transcrits d'histones majeures ou réplication-dépendants sont régulés au niveau de leur transcription par les facteurs Hir, un ensemble de protéines qui répriment la transcription des ARNm en dehors de la phase S et lorsque la synthèse d'ADN est arrêtée. Ces mêmes facteurs jouent aussi un rôle dans l'assemblage du nucléosome. Il existe également une régulation au niveau post-transcriptionnel. Celle-ci agirait sur la région 3' non traduite de l'ARNm d'histone en amont de l'extrémité polyA en agissant sur la stabilité de l'ARNm au cours du cycle cellulaire (Campbell *et al.*, 2002; Gunjan *et al.*, 2005).

Les plantes expriment également les histones de façon dépendante du cycle cellulaire par une régulation transcriptionnelle restreignant la production d'histones à la phase S. Ainsi, la transcription des gènes d'histones est activée au début de la phase S ; elle s'achève en fin de phase S puis les ARNm sont rapidement dégradés (Reichheld *et al.*, 1998).

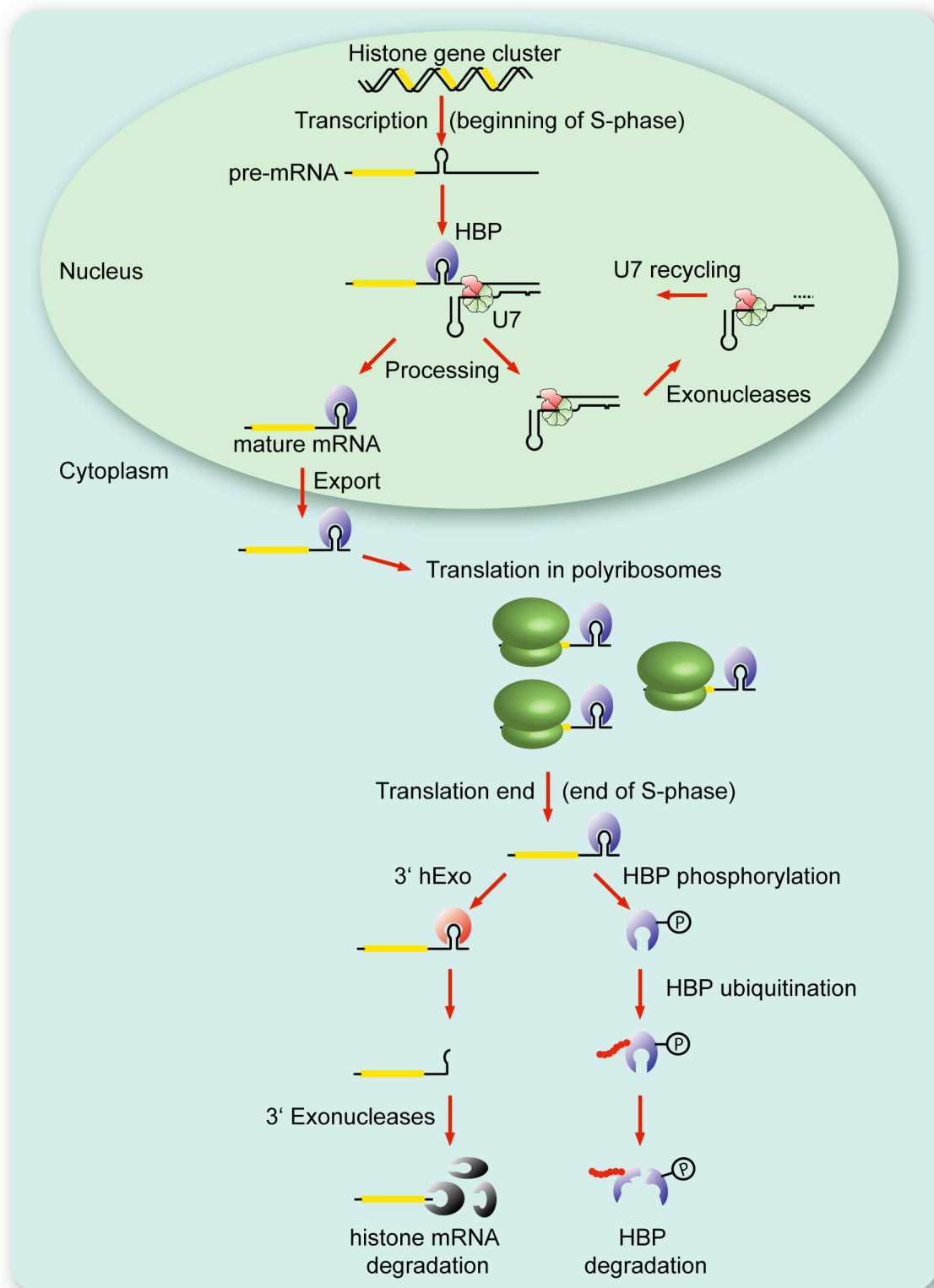


Figure 12 : Représentation schématique du métabolisme des ARNm d'histones de type réplication-dépendants chez les métazoaires. Au début de la phase S, les pré-ARNm sont transcrits à partir des « clusters » de gènes d'histones qui se localisent au niveau des « Cajal Bodies ». Après fixation de la protéine HBP et maturation de leur extrémité 3' dans le noyau, le complexe HBP:ARNm est exporté dans le cytoplasme. La protéine HBP est requise pour une traduction efficace et contribue à la stabilité de l'ARNm. En fin de phase S, HBP est phosphorylée puis ubiquitinylée, ce qui engendre sa rapide dégradation par la voie du protéasome. Les ARNm d'histones sont d'abord partiellement dégradés par une exonuclease spécifique appelée 3' hExo, puis par les exonucléases des voies classiques de dégradation. D'après Jaeger *et al.* (2005).

3.3 Expression et régulation des gènes réplication-dépendants chez les métazoaires

Cette partie de l'introduction qui décrit l'expression et la régulation des gènes d'histones dits réplication-dépendants est reprise dans ses grandes lignes dans l'article de revue placé à la fin de l'introduction, publié dans le journal *Biochimie* en 2005.

Les métazoaires ont développé des mécanismes beaucoup plus complexes afin de coordonner la production d'histones avec la synthèse d'ADN. Un mécanisme d'expression original coordonne en phase S l'expression massive des gènes d'histones de type réplication-dépendants. Ce mode d'expression original souligne l'importance de l'expression des histones dans la vie cellulaire. Il diffère de l'expression classique par différents points comme la maturation en 3', l'export du noyau vers le cytoplasme, la traduction et la dégradation des ARNm (**Figure 12**). L'expression des gènes d'histones réplication-dépendants fait également intervenir des facteurs qui lui sont spécifiques comme la protéine HBP (Hairpin Binding Protein) appelée aussi SLBP (Stem-Loop Binding Protein). La protéine HBP joue un rôle clé lors de plusieurs étapes du métabolisme des ARNm d'histones, en assurant le couplage de certains événements post-transcriptionnels avec le cycle cellulaire. Cette fonction régulatrice est obtenue grâce à la propre régulation de la protéine qui est dépendante du cycle cellulaire.

Une particularité supplémentaire des gènes d'histones de type réplication-dépendants vient de leur structure. En effet, ils ne présentent pas d'introns et ils ne sont pas polyadénylés en 3' contrairement à la quasi-totalité des ARNm. L'étape de maturation des ARN pré-messagers d'histones consiste en une coupure endonucléolytique au niveau de l'extrémité 3' non traduite, quelques nucléotides en aval d'une structure en tige-boucle. Le mécanisme orchestrant cette coupure est spécifique aux gènes d'histones et fait intervenir des facteurs bien particuliers.

Chez l'homme, on compte 65 gènes d'histones réplication-dépendants regroupés en 3 « clusters » localisés sur les chromosomes 1 et 6 dans des régions bien particulières du noyau appelées « Cajal bodies ». Les gènes d'histone H4 codent tous pour la même protéine et ceux de l'histone H3 ne comportent que deux variants (Marzluff *et al.*, 2002).

Ce sont toutes ces caractéristiques qui permettent une régulation coordonnée de la production massive et stœchiométrique des histones avec la réplication de l'ADN. Cette régulation est assurée à plusieurs niveaux du processus de biosynthèse des histones que nous allons détailler ci-dessous.

3.3.1 Transcription

La régulation de la transcription joue un rôle prépondérant dans l'expression des gènes d'histones réplication-dépendants (Zhao *et al.*, 2000). Au début de la phase S du cycle cellulaire, leur transcription par l'ARN polymérase II est augmentée de 3 à 5 fois (Marzluff & Duronio, 2002 et références contenues). Il a été proposé qu'une coordination de l'expression des gènes des histones permettrait de respecter leur stœchiométrie et faciliterait un couplage avec le cycle cellulaire. Cependant, malgré l'existence d'une transcription coordonnée, aucun élément régulateur commun n'a pu être identifié dans les promoteurs des gènes d'histones. Le seul élément commun intervenant dans la régulation de la transcription a été identifié dans les phases codantes des gènes d'histones. Cet élément appelé CRAS (Coding Region Activating Sequence) a été identifié dans les gènes des histones H2a, H2b H3 et H4 (Bowman *et al.*, 1996) et plus récemment pour l'histone H1 (Horvath *et al.*, 2003). Cette séquence contient notamment deux éléments de 7 nucléotides appelés alpha et oméga. Ils interagissent avec des facteurs nucléaires communs pour assurer une activation de la transcription spécifique à la phase S (Kaludov *et al.*, 1996). L'élément CRAS alpha fixe le facteur de transcription YY1 (YingYang1), un facteur ubiquitaire et très conservé, qui selon les cas active ou réprime l'expression des gènes. La fixation de ce facteur est nécessaire à l'activation de la transcription des gènes d'histones au cours de la phase S du cycle cellulaire (Eliassen *et al.*, 1998). À la transition de phase G1/S ce facteur devient majoritairement nucléaire, son activité de fixation de l'ADN est alors augmentée ce qui permet l'activation de la transcription des histones (Palko *et al.*, 2004). La cellule possède donc un moyen de réguler en fonction du cycle cellulaire et de manière simultanée la transcription des histones réplication-dépendantes de toutes les classes.

Notons ici, que le variant réplication-indépendant de l'histone H3.3 est naturellement muté au niveau des éléments CRAS alpha et oméga ce qui l'empêche d'interagir avec certains facteurs nucléaires et explique le fait que sa transcription reste basale (Bowman *et al.*, 1996; Kaludov *et al.*, 1996).

Une seconde voie de régulation permettant le couplage de la transcription des gènes d'histones avec le cycle cellulaire et la synthèse d'ADN a été mise en évidence. Elle est basée sur la régulation par un facteur central « NPAT » qui est un substrat du complexe cyclinE-Cdk2 kinase. Ce facteur établit un lien entre la machinerie de régulation du cycle cellulaire et la transcription des gènes d'histones (Zhao, 2004). Le complexe cyclinE-Cdk2 kinase est un régulateur clé de l'entrée en phase S des cellules animales, il est essentiel à la réplication de

l'ADN. Son activité est maximale à la transition de phase G1/S où il va notamment phosphoryler le facteur NPAT. À ce moment précis du cycle cellulaire, le complexe cyclinE-Cdk2 et le facteur NPAT vont se co-localiser avec les « clusters » de gènes d'histones au niveau des « Cajal bodies » du noyau. L'activation de NPAT par phosphorylation pourra alors s'effectuer. Une fois phosphorylé, NPAT devient capable d'activer la transcription des gènes d'histones mais également de promouvoir le passage en phase S du cycle cellulaire (Zhao *et al.*, 2000). Ainsi, la spécificité d'expression des histones en phase S est en partie la conséquence de la régulation du facteur NPAT dont l'expression est maximale à la transition G1/S où il est également activé par phosphorylation. C'est donc la régulation de NPAT qui permet d'activer la transcription des gènes d'histones et de la coordonner à l'activation de la réplication de l'ADN (Zhao *et al.*, 2000), ces deux fonctions étant indépendamment gouvernées par NPAT (Wei *et al.*, 2003).

L'activation de la transcription des gènes d'histones est également tributaire d'éléments du promoteur agissant *en cis* appelés SSCS (Subtype-Specific Consensus Sequence). Ces éléments sont reconnus par différents facteurs protéiques selon le type de gène d'histone qui est activé. Ainsi NPAT n'agirait pas en fixant directement l'ADN, mais via des co-facteurs spécifiques de chaque type de gène d'histone comme les facteurs liant les SSCS (Zhao *et al.*, 2000). Il a récemment été démontré que le régulateur de transcription spécifique HiNF-P recruté par les SSCS d'histones H4, interagit directement avec NPAT pour co-activer la transcription des gènes d'histones en phase S (Miele *et al.*, 2005).

Le facteur HIRA est un facteur qui réprime la transcription des gènes d'histones réplication-dépendants en dehors de la phase S du cycle cellulaire (Gunjan *et al.*, 2005) et dont l'expression ectopique bloque la synthèse d'ADN et arrête les cellules en phase S (Nelson *et al.*, 2002). Il est également phosphorylé par des kinases cyclin-dépendantes qui promeuvent la phase S (cyclinE-Cdk2 et cyclinA-Cdk2). Ces kinases permettraient d'activer NPAT et d'éteindre HIRA à l'abord de la phase S de manière à activer la transcription des gènes d'histones. Comme nous l'avons précédemment indiqué, le facteur HIRA a également une fonction dans l'incorporation du variant d'histone H3.3 dans le nucléosome. Ce double rôle reste à ce jour inexplicable (Gunjan *et al.*, 2005).

D'autres éléments impliqués dans la transcription des gènes d'histones ont été découverts. Les gènes codant pour les histones H3 possèdent en effet des promoteurs contenant une boîte TATA et deux boîtes CCAAT agencées en tandem dans des positions relativement fixes et cruciales pour leur aptitude à activer la transcription. À l'exception des histones H4, les promoteurs de gènes d'histones contiennent au moins une boîte CCAAT. Ces

boîtes interagissent avec le facteur CBF (CCAAT-binding factor)/NF-Y (Nuclear Factor-Y) ce qui suggère la formation d'un complexe d'initiation de la transcription particulier au niveau des promoteurs, comportant deux molécules de CBF/NF-Y ainsi que des facteurs de transcription généraux ou spécifiques. La manière dont tous ces facteurs agissent ensemble pour obtenir une coordination de la transcription est encore inconnue. Il semblerait que le couplage de l'expression avec le facteur cycle-cellulaire pourrait être médié par la localisation de CBF/NF-Y qui devient nucléaire en phase S (Koessler *et al.*, 2004).

Couplage transcription-maturation

La transcription des gènes d'histones produit des ARN messagers dotés d'une longue extrémité 3' qui doit être maturée par clivage endonucléolytique comme nous l'avons mentionné auparavant. Cette maturation est différente de la réaction de maturation en 3' des autres ARNm transcrits par l'ARN polymérase II. Ces derniers subissent en effet un clivage puis une polyadénylation de leur extrémité 3' non traduite, et ce de manière couplée à la transcription. Pour cela, les facteurs Cstf et CPSF interagissent avec le domaine C-terminal ou domaine CTD de l'ARN polymérase et sont immédiatement desservis au niveau du site de maturation en 3' au cours de la transcription (Marzluff, 2005).

Concernant les ARNm d'histones réplication-dépendants, la question du couplage entre leur réaction spécifique de maturation en 3' et leur transcription n'a pas encore été résolue. Des systèmes *in vitro* permettant à la fois la transcription et la maturation des transcrits ont été développés chez la souris (Gu & Marzluff, 1996) et *Drosophila melanogaster* (Adamson & Price, 2003). Certaines études menées avec ces systèmes de souris ont démontré la nécessité de la présence d'un site de maturation intact à l'extrémité 3' non traduite du transcrit naissant pour que ce dernier subisse une terminaison de transcription correcte, faute de quoi la transcription se poursuit en ne respectant pas le signal de terminaison du gène. La transcription s'arrêtera alors au niveau de l'élément de terminaison du gène d'histone suivant s'il est doté d'un site de maturation intact (Chodchoy *et al.*, 1991; Gu & Marzluff, 1996). Ces résultats suggèrent donc qu'il existe un couplage entre la transcription et la maturation de l'extrémité 3'. Cependant, la présence de transcrits non maturés dans ces systèmes *in vitro* démontre que tous les transcrits ne sont pas systématiquement maturés et que la réaction de coupure n'est pas un pré-requis pour la terminaison (Gu & Marzluff, 1996). À l'opposé, des expériences réalisées dans la drosophile ont montré que l'efficacité de maturation de transcrits contenus dans des complexes d'élongation et des transcrits libres était identique, ce qui suggère qu'au contraire, il n'y aurait

pas de couplage transcription-maturation (Adamson & Price, 2003; Gu & Marzluff, 1996). Une incertitude demeure donc à ce niveau.

Pour conclure, rappelons que les ARN messagers d'histones réplication-dépendants s'accumulent au cours de la phase S du cycle cellulaire pour permettre une synthèse massive d'histones. Cette augmentation du taux d'ARNm est en partie due à une régulation de leur transcription comme nous venons de le décrire, mais la plus grande contribution vient plutôt d'une régulation à un niveau post-transcriptionnel, c'est-à-dire par l'étape de maturation de l'extrémité 3' dans un premier temps (Gunjan *et al.*, 2005).

3.3.2 Maturation de l'extrémité 3'

Les gènes d'histones réplication-dépendants étant dépourvus d'introns, la seule étape de maturation requise pour l'obtention d'ARNm d'histones est la maturation de leur extrémité 3'. L'efficacité de cette étape augmente de 8 à 10 fois au cours de la phase S, permettant une accumulation rapide d'ARNm matures (Marzluff & Duronio, 2002).

La réaction de maturation de l'ARN pré-messager consiste en une coupure endonucléolytique qui s'effectue entre deux éléments agissant en *cis* dans la réaction (**Figure 13**). Un premier élément est une structure en tige-boucle très conservée qui se trouve en amont du site de coupure. Le second est une séquence riche en purines située en aval du site de coupure appelée HDE pour « Histone Downstream Element » (également appelé « Spacer Element »). Ces éléments *cis* permettent le recrutement de facteurs agissant en *trans* lors de la réaction de maturation en 3'.

La structure en tige-boucle interagit de manière très forte et très spécifique avec une protéine appelée HBP pour « Hairpin Binding Protein » ou SLBP pour « Stem-Loop Binding Protein ». (Martin *et al.*, 1997; Wang *et al.*, 1996). La séquence HDE est quant à elle en charge de recruter la snRNP U7. La particule U7 va s'y ancrer par hybridation de l'extrémité 5' de son snRNA (Dominski *et al.*, 1999). Le complexe ternaire formé par l'ARNm, la protéine HBP et la snRNP U7 est stabilisé par la protéine ZFP100 (Zinc Finger Protein 100), une protéine riche en motifs à doigts de zinc, qui interagit simultanément avec la protéine HBP fixée à l'ARN et l'extrémité amino-terminale de la protéine Lsm11, un des constituants de la particule U7. Cette protéine ZFP100 permet de ponter et de stabiliser la particule U7 et le complexe formé par la protéine HBP et la tige-boucle de l'ARN pré-messager d'histone (Azzouz *et al.*, 2005a; Dominski *et al.*, 2002; Pillai *et al.*, 2003).

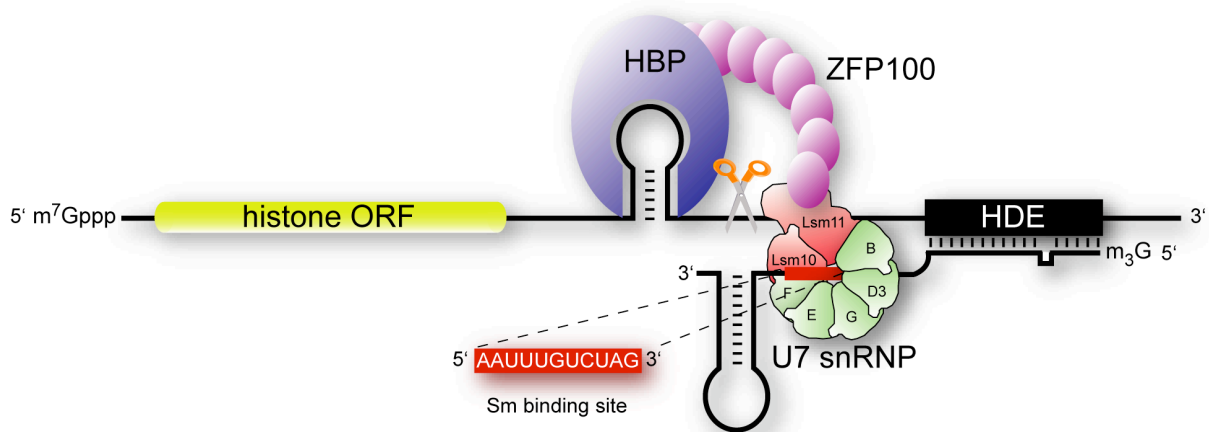


Figure 13 : Représentation schématique de la réaction de maturation en 3' spécifique aux pré-ARNm d'histones de type réplication-dépendants. Cette réaction consiste en un clivage endonucléolytique (indiqué par les ciseaux) qui s'effectue entre une structure en tige-boucle extrêmement conservée et une séquence riche en nucléotides puriques appelée HDE pour « Histone Downstream Element ». Les facteurs intervenant en *trans* dans la réaction sont la protéine HBP pour « Hairpin Binding Protein » (en violet), qui interagit avec la structure en tige-boucle, la snRNP U7 (en vert et rouge) qui s'ancre au niveau de la séquence HDE et la protéine ZFP100 pour « Zinc Finger Protein de 100 kDa » (en rose) qui pontre ces deux facteurs et permet ainsi de stabiliser l'ancrage de la particule U7 sur le pré-ARNm.

3.3.2.1 Les différents acteurs de la réaction de maturation

3.3.2.1.1 La snRNP U7

La snRNP U7 est une particule dite mineure car elle est très rare dans la cellule. Le noyau compte en effet environ 100 copies de snRNP U7 alors que 500.000 voir un million d'autres snRNPs qui composent le « spliceosome » ont été dénombrées (Marzluff, 2005). La particule U7 diffère également des snRNPs classiques dans sa composition. Elle contient 5 des 7 protéines Sm canoniques entrant dans la composition de toutes les autres snRNPs. Ce sont les protéines Sm B, D3, E, F, G, ainsi que deux protéines Sm-like spécifiques à U7, les protéines Lsm10 et Lsm11 qui remplacent les protéines Sm D1 et D2 (Pillai *et al.*, 2001; Pillai *et al.*, 2003) (**Figure 14**).

Des deux protéines Sm-like ou Lsm spécifiques à U7 que sont les protéines Lsm10 et Lsm11, la protéine Lsm11 (d'une masse moléculaire de 52 kDa) est la plus originale car elle possède une longue extension N-terminale de 170 acides aminés ainsi que deux motifs de type Sm (impliqués dans l'assemblage au snRNA) espacés de 138 résidus. Le snRNA U7 est

composé d'environ 57 nucléotides chez les vertébrés (**Figure 13**). Il présente à son extrémité 5' une coiffe tri-méthylée et une séquence d'hybridation aux ARN pré-messagers d'histones. Vient ensuite le site de fixation des protéines Sm ou « Sm binding site » dont la séquence est particulière à U7 et qui est directement impliqué dans l'assemblage de la particule. Enfin, cet ARN possède au niveau de son extrémité 3' une longue structure en tige-boucle probablement requise pour sa stabilité (Schümperli & Pillai, 2004).

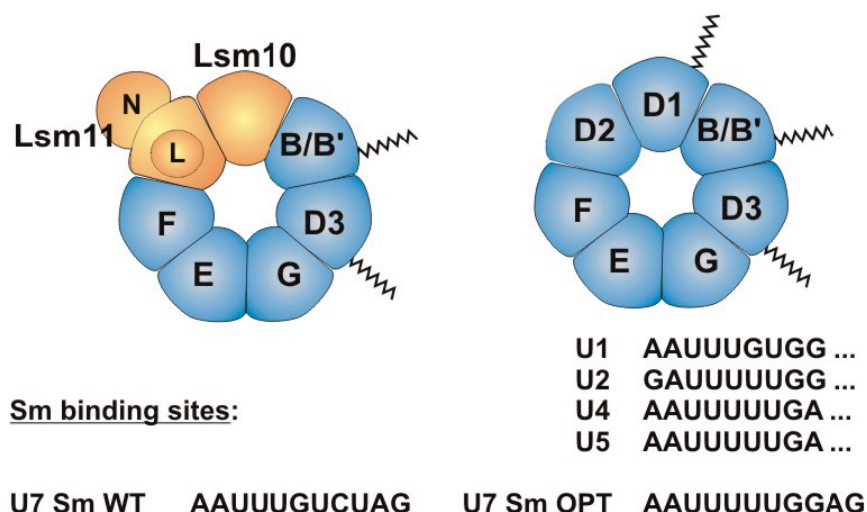


Figure 14 : La composition de la snRNP U7 diffère de celle des snRNPs classiques. Alors que ces dernières contiennent un noyau de sept protéines Sm classiques à savoir B/B', D1, D2, D3, E, F et G (à droite), la snRNP U7 (à gauche) possède deux protéines spécifiques de type Sm-like, Lsm10 et Lsm11 qui remplacent les protéines Sm D1 et D2. Lsm11 possède une extension N-terminale (N) et une longue séquence linker (L) entre ses deux motifs Sm impliqués dans l'assemblage au snRNA U7. L'extrémité N-terminale participe notamment à la maturation en interagissant directement avec la protéine ZFP100. La conversion du motif « Sm binding site » spécifique au snRNA U7 (U7 Sm WT) en motif consensus (U7 Sm OPT) dérivé des « Sm binding sites » de U1, U2, U4 et U5 entraîne l'assemblage en une particule classique contenant les protéines D1 et D2. L'assemblage de la particule U7 est donc déterminé par la séquence de son « Sm binding site » et est médié par des complexes SMN distincts contenant soit le dimère D1/D2, soit le dimère Lsm10/Lsm11. Les extensions représentées sur les protéines B/B', D1 et D3 correspondent à des répétitions du dipeptide Arginine-Glycine qui contiennent des modifications de type diméthyl-arginines symétriques introduites par le complexe méthyl-transférase PRMT5. D'après Schümperli & Pillai (2004).

Assemblage de la snRNP U7

L'assemblage des snRNP classiques s'effectue de manière spontanée *in vitro* en commençant par la fixation de dimères B/B'-D3, D1-D2 et du trimère E-F-G sur le « Sm binding site » du snRNA. Les protéines vont alors former une sorte d'anneau sur cette séquence particulière (**Figure 14**). *In vivo*, l'assemblage des snRNPs est conduit de façon ordonnée et dépendante de l'ATP par des facteurs protéiques. La protéine SMN est un facteur très important s'associant aux protéines Sm et Sm-like et régulant l'assemblage de la snRNP.

SMN forme un complexe avec d'autres protéines impliquées dans l'assemblage des particules: les gemins. Le « complexe SMN » interagit aussi avec le snRNA avec lequel il s'associe dans le cytoplasme. La mutation de la protéine SMN conduit à une grave pathologie neuromusculaire appelée SMA pour « Spinal Muscular Atrophy ».

L'assemblage de la snRNP U7 ressemble à celui des snRNP classiques mais la présence de son « Sm binding site » particulier (**Figure 14**) lui vaut d'être assemblé par un complexe SMN spécifique d'une composition différente du complexe SMN classique (Pillai *et al.*, 2003). C'est le type de « Sm binding site » qui définit la nature des protéines Sm se fixant sur le snRNA. Le remplacement du « Sm binding site » spécifique au snRNA U7, par le consensus des snRNPs classiques (Sm OPT, **Figure 14**), entraîne la fixation d'un heptamère classique de protéines Sm autour du snRNA U7 (Schümperli & Pillai, 2004).

Avant d'être incorporées au complexe SMN les protéines Sm B/B' D1 et D3 sont modifiées au niveau des motifs riches en résidus Arg et Gly (motifs RG) de leur extrémité carboxy-terminale de manière à produire des diméthyl-arginines symétriques (**Figure 14**). Ces modifications permettent d'augmenter l'affinité des protéines Sm pour le complexe SMN et favorisent ainsi leur prise en charge avant l'incorporation (Schümperli & Pillai, 2004). Le complexe protéique responsable de ces modifications contient la protéine PRMT5 qui est une méthyl-transférase, le facteur pICln qui fixe les protéines Sm, le facteur WD45, ainsi que d'autres sous-unités (Schümperli & Pillai, 2004).

Les protéines Lsm10 et Lsm11 interagissent quant à elles avec le complexe SMN en l'absence des modifications diméthyl-arginine symétriques. Il existerait au sein du complexe SMN, deux sites de fixation distincts pour les protéines Sm/Lsm. Le premier interagissant avec les motifs diméthyl-arginine, le second avec les motifs de type Sm (Azzouz *et al.*, 2005b). Ainsi, ces résultats confirment bien que l'assemblage de la particule U7 s'effectue selon un mécanisme atypique.

Implication dans la maturation

Des expériences ont montré que le déplacement du site d'ancrage de la particule U7 entraînait le même déplacement du site de coupure. La particule U7 semble donc définir la position du site de coupure ce qui lui a valu le qualificatif de « molecular ruler » (Scharl & Steitz, 1994).

Des expériences récentes ont montré que l'ancrage de la particule U7 à l'ARN pré-messager est stabilisé par l'interaction entre la protéine Lsm11 et la protéine ZFP100 (Dominski *et al.*, 2002). Deux sites d'interactions entre les protéines ZFP100 et Lsm11 ont été

identifiés (Azzouz *et al.*, 2005a). Le domaine N-terminal de la protéine Lsm11 est l'un de ces sites d'interaction. Il contient quatre motifs peptidiques très conservés. Des mutations dans l'un de ces motifs inhibent fortement la maturation tout en respectant l'intégrité structurale de la particule. Des mutations dans deux autres motifs entraînent une perte partielle de l'activité de maturation sans effet sur l'interaction avec la protéine ZFP100. Ces résultats suggèrent donc que ces motifs sont impliqués dans la réaction de maturation et non dans l'interaction avec la protéine ZFP100 (Azzouz *et al.*, 2005a).

3.3.2.1.2 La protéine ZFP100

Le facteur ZFP100 reste à ce jour peu connu. Il a été isolé par criblage d'une banque d'ADNc humain par une variante de la méthode du double hybride à l'aide d'un appât constitué du complexe HBP fixé sur la structure en tige-boucle d'un ARNm d'histone. Cette protéine de 100 kDa est très riche en motifs de liaison à l'ADN en doigts de zinc. Elle possède 18 motifs du type Cys₂His₂ dont certains interagissent avec le complexe HBP:tige-boucle (motifs 2 à 8) (Dominski *et al.*, 2002) et d'autres (non définis) avec la protéine Lsm11 spécifique de la particule U7 (Azzouz *et al.*, 2005a). La fonction de cette protéine serait de stabiliser l'ancrage de la particule U7 sur l'ARN pré-messager en la pontant à la protéine HBP fixée sur la structure en tige-boucle des ARNm d'histones (Dominski *et al.*, 2002) (**Figure 13**). ZFP100 possède en outre au niveau de son extrémité N-terminale un domaine KRAB (Krüppel-associated Box) dont la fonction reste inconnue à ce jour (Dominski *et al.*, 2002). Ce domaine possède habituellement, lorsqu'il est associé à des motifs en doigt de Zinc, une fonction de répresseur transcriptionnel (Urrutia, 2003).

3.3.2.1.3 Le complexe HBP:tige-boucle sur les ARN pré-messagers d'histones

3.3.2.1.3.1 Les partenaires

- La protéine HBP

Cette protéine a été clonée par criblage d'une banque d'ADN complémentaire de souris par la méthode du triple hybride dans la levure, en utilisant comme appât la structure en tige-boucle de l'ARNm de l'histone de souris H4-12 ou H2a-614 (Martin *et al.*, 1997; Wang *et al.*, 1996). Le caractère essentiel de la protéine HBP a été montré par des expériences d'ARN interférence dans des cellules HeLa, qui occasionnent une inhibition de l'expression des gènes d'histones, de la synthèse d'ADN, de la progression en phase S du cycle cellulaire

qui s'accompagne de l'inhibition de la prolifération cellulaire (Zhao *et al.*, 2004). D'autres expériences d'ARN interférence menées sur des cellules humaines d'ostéoblastes U2OS ont montré que l'accumulation des cellules en phase S pouvait être reversée par l'expression d'un mutant d'HBP résistant au traitement (Wagner *et al.*, 2005).

L'expression de la protéine HBP est régulée au niveau du cycle cellulaire. Il s'agirait même du seul facteur intervenant dans la maturation en 3' des ARN pré-messagers d'histones qui soit régulé de manière dépendante du cycle cellulaire. Ainsi, la protéine HBP serait l'unique facteur responsable de la synchronisation de la maturation en 3' avec le cycle cellulaire (Zheng *et al.*, 2003). La protéine HBP s'accumule juste avant la phase S par augmentation de sa traduction, puis elle est rapidement dégradée en fin de phase S (Whitfield *et al.*, 2000) après phosphorylation de deux résidus thréonine localisés dans son domaine N-terminal. La protéine contient également un site de reconnaissance pour une cycline requise pour la dégradation. En effet, les phosphorylations conduisant à la dégradation résulteraient de l'action de kinases dépendantes de cette cycline du type cyclin/cdk (Zheng *et al.*, 2003).

Les résidus de la protéine HBP impliqués dans la réaction de maturation ont été identifiés. Ils agissent indirectement en recrutant la protéine ZFP100 (Dominski *et al.*, 2002) et la snRNP U7 au niveau du site de maturation (Dominski *et al.*, 2001).

Des études par Résonance Magnétique Nucléaire (RMN) et dichroïsme circulaire portant sur le domaine central RBD (RNA Binding Domain) ont révélé la présence de conformations en hélices peu stables (Thapar *et al.*, 2004b). La présence de l'ARN en tige-boucle n'est pas suffisante pour stabiliser la conformation de la protéine HBP. Par contre, la phosphorylation de deux sérines situées à l'extrémité C-terminale (une modification post-traductionnelle normalement requise pour la maturation chez la drosophile) augmente significativement la stabilité du complexe ARN:protéine en créant des interactions électrostatiques favorables (Thapar *et al.*, 2004a).

- La tige-boucle d'ARN

Les études par RMN réalisées sur la tige-boucle des ARNm d'histones H4-12 murins ont confirmé la présence d'une structure en tige-boucle composée d'une tige de six paires de bases et d'une boucle à quatre ou six nucléotides (**Figure 15**) (DeJong *et al.*, 2002; Zanier *et al.*, 2002). Les nucléotides flanquant la tige sont désordonnés malgré le fait qu'ils soient impliqués dans la fixation de la protéine HBP (Battle & Doudna, 2001).

La mutagenèse des résidus conservés de la tige-boucle a montré leur implication dans la maturation. Celle-ci est considérablement affectée lorsque l'on modifie le premier et le

troisième résidu de la boucle ou si l'on utilise une tige présentant des paires de bases 00 inversées. Ces tiges-boucles mutantes ont également la particularité de ne plus interagir avec la protéine HBP (Pandey *et al.*, 1994) ce qui suggère que le défaut de maturation pourrait résulter de la perte d'interaction avec la protéine HBP et que par conséquent cette interaction est très importante pour la maturation en 3'.

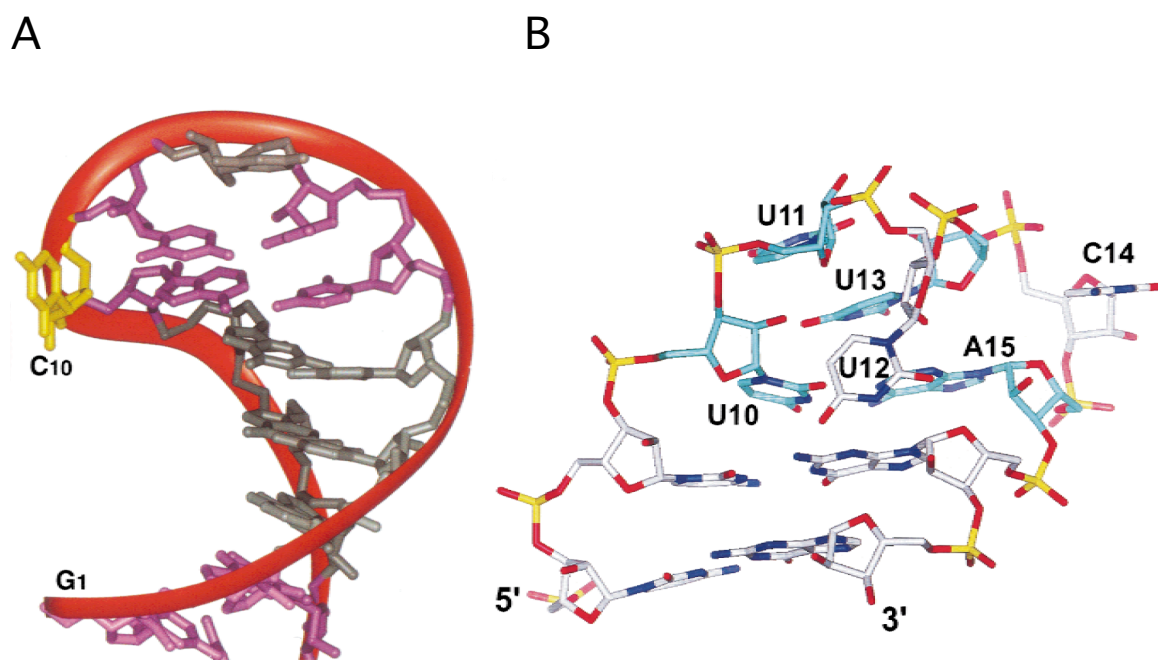


Figure 15 : Structures de la tige-boucle des ARNm d'histones résolues par Résonance Magnétique Nucléaire. (A) structure proposée par DeJong *et al.* (2002). (B) structure proposée par Zanier *et al.* (2002). Dans chacun des 2 modèles le résidu Cytosine de la boucle est rejeté en dehors de la structure. Par contre, seul le 1^{er} modèle propose un empilement des trois résidus Uridine de la boucle.

- Le complexe HBP:tige-boucle

Le complexe formé entre la protéine HBP et la structure en tige-boucle des ARNm d'histones a fait l'objet de nombreuses études. Une fois fixée sur la structure en tige-boucle, la protéine HBP va intervenir non seulement au niveau de la maturation de l'extrémité 3' non traduite (Martin *et al.*, 1997; Wang *et al.*, 1996) mais elle reste ensuite sur l'ARNm maturé et contribue à d'autres étapes du métabolisme des ARNm d'histones. L'étude de ce complexe comporte un autre intérêt. En effet, il représente un nouveau type d'interaction ARN:protéine qui n'a jusqu'à présent pas été caractérisé d'un point de vue structural.

Par filtration sur membrane de nitrocellulose, une très forte affinité pour l'ARN a pu être mesurée ; la constante de dissociation est de l'ordre de 1,5 nM. Des tests de fixation de tiges-boucles mutées ont permis d'identifier les nucléotides qui jouent un rôle essentiel pour l'affinité (Battle & Doudna, 2001).

En ce qui concerne HBP, la protéine peut se subdiviser en trois domaines : un domaine central minimal requis pour la fixation de l'ARN ou RBD pour « RNA Binding Domain » (défini par des expériences de délétion), flanqué de deux domaines N- et C-terminaux non essentiels pour la fixation à l'ARN (Wang *et al.*, 1996). Par mutagenèse dirigée de résidus conservés dans le RBD, un certain nombre de résidus impliqués dans l'interaction avec l'ARN ont été identifiés (Dominski *et al.*, 2001). Le complexe HBP:tige-boucle a également été étudié *in vivo* par la méthode du triple hybride dans la levure (Martin *et al.*, 2000). Des mutants d'ARN non fonctionnels sont ainsi venus confirmer les précédents résultats obtenus *in vitro*. Puis, grâce à une sélection positive effectuée sur la protéine mutée aléatoirement et ces tiges-boucles mutantes, des mutants d'HBP restaurant la fixation ont été isolés. Ces mutants d'HBP ont perdu leur spécificité de reconnaissance pour la tige-boucle sauvage et ils ont ainsi permis d'identifier les acides aminés participant au processus de reconnaissance de la tige-boucle. Étrangement, tous ces résidus sont localisés dans les domaines N- et C-terminaux de la protéine qui bordent le domaine RBD, suggérant que ces domaines jouent un rôle non négligeable pour la spécificité de reconnaissance des nucléotides de l'ARN. Par ailleurs, une sélection négative de mutants d'HBP a permis de caractériser des résidus qui abolissaient l'interaction de la protéine avec l'ARN. Des résidus différents de ceux caractérisés précédemment lors des études *in vitro* sont venus s'ajouter à la liste des résidus essentiels d'HBP. Ils sont tous localisés dans le domaine RBD et concernent des résidus très conservés.

3.3.2.1.3.2 Rôle du complexe HBP:tige-boucle dans la maturation en 3'

Le complexe formé par la structure en tige-boucle et la protéine HBP est un pré-requis indispensable pour la maturation de l'extrémité 3'. Toute abolition de l'interaction par mutagenèse ou par déplétion des extraits nucléaires (soit par des anticorps spécifiques d'HBP soit par des tiges-boucles compétitrices) entraîne une diminution de l'efficacité de maturation *in vitro* (Dominski *et al.*, 1995; Martin *et al.*, 1997; Pandey *et al.*, 1994; Vasserot *et al.*, 1989; Wang *et al.*, 1996). Le complexe est donc nécessaire pour obtenir une maturation efficace. Certaines études ont néanmoins démontré que dans le cas de l'ARNm H4-12 l'inhibition de l'interaction entre HBP et la tige-boucle par compétition n'avait qu'un effet mineur sur l'efficacité de maturation en 3' dans des extraits nucléaires murins de type K21 (Melin *et al.*, 1992; Streit *et al.*, 1993). Dans ces exemples, la protéine HBP n'est pas vraiment indispensable, la réaction se déroulant encore à environ 80%. Cette apparente contradiction avec les précédents travaux pourrait s'expliquer par l'origine des extraits nucléaires utilisés

pour tester la réaction de maturation ainsi que l'origine des gènes testés. Globalement, l'ensemble des études menées sur la contribution du complexe dans la réaction de maturation aboutit à une conclusion consensuelle attribuant à la fixation de la protéine HBP sur la tige-boucle une fonction de stimulation de la réaction de maturation. Elle n'est toutefois pas essentielle, la maturation s'effectuant toujours en l'absence de protéine HBP de manière plus ou moins efficace.

- Comment HBP stimule-t-elle la maturation ?

Très tôt, il avait été noté que les ARNm des histones réplication-dépendantes n'étaient pas maturés avec la même efficacité. Alors que les séquences des tiges-boucles sont très conservées, les séquences des HDE sont très variables en fonction des gènes (Liu *et al.*, 1989; Pandey *et al.*, 1994). Cette différence dans l'efficacité de maturation pourrait-elle s'expliquer dans les variations de séquence du HDE ? Dès 1989, Liu *et al.* avaient proposé que les ARNm efficacement maturés étaient ceux qui hybridaient mieux le snRNA U7. Ces observations suggéraient déjà une relation entre la complémentarité de séquence du HDE et du U7 snRNA et l'efficacité de maturation en 3' (Liu *et al.*, 1989).

Des travaux axés sur le HDE ont alors mis en évidence un lien entre l'efficacité de maturation d'un ARN pré-messager et sa capacité à former un complexe avec la particule U7. L'ARN pré-messager H4-53 est cinq fois moins bien maturé que celui d'H4-12. Cette différence d'efficacité est entièrement due à la nature du HDE. De plus, le HDE d'H4-53 n'est pas capable de former un complexe détectable avec U7 contrairement à l'ARN pré-messager H4-12 (Melin *et al.*, 1992). Ainsi, il semblerait que plus l'ARNm est capable de recruter la particule U7, mieux il sera maturé en 3'. Streit *et al.* ont par la suite testé des combinaisons chimériques associant les tiges-boucles et les HDE des ARNs H4-12 et H4-53 et d'autres ARNm d'histones. Il en résulte que l'efficacité de la maturation ne dépend pas de la protéine HBP mais plutôt de la séquence du HDE. Par ailleurs, il a été montré que l'inhibition de la maturation induite par la présence d'une tige-boucle mutante pouvait être compensée en améliorant l'hybridation du snRNA U7. Ainsi, une corrélation entre maturation et fixation d'HBP a été mise en évidence, pouvant s'expliquer par le fait que la protéine HBP stabiliserait l'ancrage de la particule U7 sur l'ARN pré-messager (Streit *et al.*, 1993).

Spycher *et al.* ont enfin démontré que la déstabilisation de l'interaction entre le snRNA U7 et l'ARN pré-messager d'histone entraînait une chute de l'efficacité de maturation accompagnée d'une dépendance accrue en HBP. Inversement, une mutation « stabilisante » qui améliore l'hybridation entre les deux ARNs rend la maturation moins dépendante de HBP

(Spycher *et al.*, 1994). En conclusion, plus le HDE est complémentaire à la snRNP U7, plus le complexe formé entre le pré-messager et la snRNP U7 sera stable et la maturation efficace. Dans ce cas, la protéine HBP ne sera que peu requise dans la maturation. Inversement, une faible complémentarité de séquence conduira à un ancrage de U7 peu stable et à une maturation peu efficace qui nécessitera davantage l'effet stabilisant de la protéine HBP.

Par la suite, d'autres résultats ont montré que la protéine HBP fixée à l'extrémité 3' non traduite des ARN pré-messagers d'histones stimulait la réaction de maturation par stabilisation de l'ancrage de la particule U7 (Dominski *et al.*, 1999). Dans ces travaux, la maturation *in vitro* des ARNm d'histones H1t et H2a-614 a été étudiée. Le gène H1t est connu pour avoir un HDE relativement différent de la séquence consensus et peu susceptible d'hybrider le snRNA U7 (**Figure 16**). Il en résulte que la maturation de l'ARN pré-messager d'H1t est strictement dépendante d'HBP. La caractéristique du gène H2a-614 est d'être maturé à hauteur de 5-10% en l'absence d'HBP ce qui le rend partiellement indépendant d'HBP, son HDE étant beaucoup plus complémentaire au snRNA U7 (**Figure 16**). Les auteurs ont démontré que le HDE d'H1t, en plus du fait d'être totalement dépendant d'HBP dans sa maturation, co-immuno précipitait moins de snRNA U7 en présence d'anticorps anti-HBP que le HDE d'H2a-614. Ces résultats confirment à nouveau que la séquence du HDE détermine son aptitude à fixer la particule U7 et la dépendance en HBP dans la maturation en 3'. De plus, l'insertion de 12 nucléotides entre le HDE et la structure en tige-boucle résulte en une diminution de l'efficacité de maturation accompagnée d'une incapacité à former un complexe stable avec la particule U7. Ainsi, si l'on éloigne la protéine HBP du HDE, la particule s'ancore encore plus difficilement sur l'ARN pré-messager, ce qui démontre plus directement le rôle joué par HBP dans la stabilisation de la snRNP U7 sur le HDE. HBP stimulerait donc la réaction de maturation en stabilisant l'ancrage de la snRNP U7 sur les ARN pré-messagers d'histones.

La découverte du facteur ZFP100 (Dominski *et al.*, 2002) (voir plus haut) a permis de comprendre comment la protéine HBP pouvait assurer cette fonction. HBP participerait à l'ancrage de la snRNP U7 grâce à ZFP100 qui assure le lien entre HBP fixée sur la tige-boucle et la protéine Lsm11 qui est l'un des constituants spécifiques à la snRNP U7. Ce « pont protéique » doublerait donc le duplexe formé entre les deux ARNs afin de le renforcer et de mieux stabiliser le complexe.

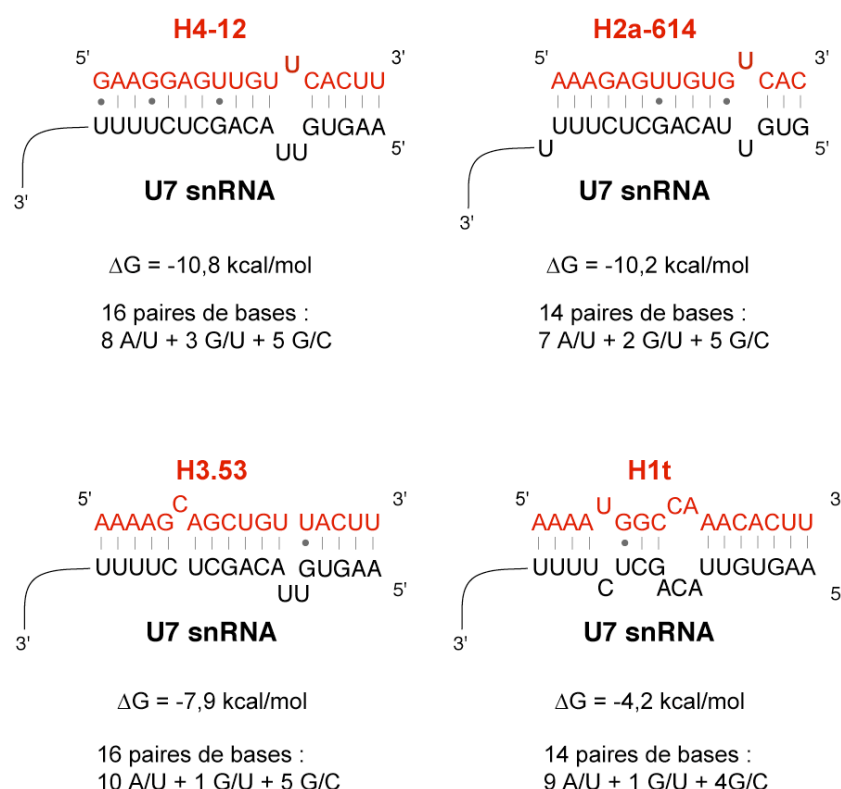


Figure 16 : Prédiction d'appariement du snRNA U7 avec la séquence HDE (en rouge) des ARN pré-messagers d'histones H4-12, H2a-614, H3.53 et H1t. La variation d'énergie libre (ΔG) reflétant la stabilité des appariements a été déterminée en utilisant le logiciel « Nucleic Acid 2 » (Zuker, 2003). Les complexes formés entre le snRNA U7 et les séquences HDE des gènes H4-12 et H2a-614 sont les plus stables. L'hybride formé avec la séquence HDE du gène H3.53 est presque tout aussi stable. Par contre, le duplexe correspondant à H1t est le plus faible.

-Dépendance d'HBP et maturation en 3' : de grandes variations sont parfois observées.

Les études de maturation et de dépendance d'HBP présentent parfois des résultats contradictoires. Alors que les études de Streit *et al.* (1993) montraient que la maturation de l'ARN H4-12 est très peu dépendante de la protéine HBP, d'autres travaux ont montré qu'au contraire sa contribution était essentielle (Martin *et al.*, 1997; Vasserot *et al.*, 1989).

Ces différences ont été expliquées par l'utilisation de substrats, ou d'extraits nucléaires différents. En effet, la protéine HBP était fortement requise pour la maturation *in vitro* de l'ARN pré-messager H4-12 dans des extraits nucléaires de cellules HeLa alors qu'elle ne l'est pas dans des extraits nucléaires murins de type K21 (Streit *et al.*, 1993). Une telle différence est également observée, mais de manière moins prononcée, avec le gène H2a-614 qui voit sa maturation passer de 10-20% dans des extraits nucléaires de cellules de myélome de souris (Wang *et al.*, 1996) à moins de 5% dans un extrait nucléaire de cellule HeLa (Pandey *et al.*, 1994) lorsque l'interaction entre HBP et la tige-boucle est abolie. Dans ces exemples, les variations ont été attribuées à des différences de concentration en snRNP U7 dans les extraits.

Cependant, d'autres ARNm comme H4-1 sont maturés de manière identique dans ces extraits ce qui tend à réfuter la précédente hypothèse (Streit *et al.*, 1993). Il persiste donc une absence d'explication satisfaisante dans ce cas.

Des différences de maturation ont également été observées pour un même mutant d'HBP avec des ARNs différents. Plusieurs mutants de la protéine HBP qui affectent la réaction de maturation mais pas la fixation à l'ARN ont été isolés (Dominski *et al.*, 2001). Dans le cas de l'ARNm H1t, ces mutants n'étaient pas ou peu capables de restaurer la maturation d'un extrait déplété en HBP, alors qu'ils restauraient celui de l'ARNm H2a-614 aussi efficacement que la protéine sauvage. Les auteurs ont montré que l'incapacité à rétablir une maturation normale était liée à une incapacité à recruter la snRNP U7. Ces expériences prouvent bien l'existence d'une spécificité de substrat mais au niveau fonctionnel dans ce cas.

Enfin, un dernier travail souligne encore davantage la confusion qui règne dans les tentatives d'attribuer une fonction dans la maturation à HBP (Streit *et al.*, 1993). Ce travail suggère que l'unique rôle d'HBP ne serait pas de stabiliser la snRNP U7 dans la réaction de maturation. En effet, le gène H3-53 présente un HDE très complémentaire au snRNA U7 et devrait être capable de s'apparier fortement au snRNA U7 (**Figure 16**). Or, les auteurs ont observé une grande dépendance d'HBP dans la réaction de maturation en 3' (Streit *et al.*, 1993).

Enfin, la confusion générale est aussi entretenue par la manière d'interpréter les résultats. Pour certains auteurs, l'interaction entre HBP et la structure en tige-boucle est jugée critique pour l'ARNm H2a-614 lorsque l'efficacité de maturation n'est plus que 5% du niveau initial (Pandey *et al.*, 1994). Pour d'autres auteurs au contraire, un résultat similaire indique que la maturation peut tout de même s'effectuer en l'absence d'HBP et que celle-ci ne joue qu'un rôle mineur (Dominski *et al.*, 1999).

3.3.2.1.4 Le site de coupure

À ce jour, nous disposons d'une bonne connaissance des facteurs impliqués dans la réaction de maturation de l'extrémité 3' des ARN pré-messagers d'histones de type réplication-dépendants. Ils sont tous nécessaires pour obtenir une réaction de coupure efficace, mais le mécanisme responsable de cette coupure reste inconnu. Diverses expériences de mutagenèse ont démontré que l'efficacité de clivage dépendait d'une liaison phosphodiester optimale qui est située idéalement en 3' d'une adénosine. La séquence nucléotidique du site de clivage contient des résidus adénosine et cytosine avec une

préférence pour une séquence « CA » au niveau du site de clivage (Furger *et al.*, 1998). Le remplacement de quatre résidus du site de clivage par des guanosines abolit totalement la réaction de maturation sans affecter l'ancrage de la particule U7, ce qui prouve l'importance de la séquence du site de coupure (Dominski *et al.*, 1999).

3.3.2.2 Lsm11 est-elle responsable de la coupure ?

À ce jour, le mécanisme de coupure endonucléolytique ainsi que le facteur responsable de cette coupure restent inconnus. Toutefois, il a été proposé que la protéine Lsm11 et plus particulièrement sa longue extrémité N-terminale serait responsable de la réaction de coupure (Azzouz *et al.*, 2005a). En effet, des mutants de délétion de cette extension sont inactifs pour la maturation. Il est aussi possible que l'extension N-terminale serve à recruter un autre facteur de type endonucléase qui serait responsable de la coupure. Dans une étude très récente, le facteur CPSF-73, un facteur impliqué dans la maturation des ARNm polyadénylés, a été ponté par les UV dans la région du site de coupure de l'ARN pré-messager de l'histone H2a-614. Bien que son rôle direct dans la maturation n'ait pas été démontré, il pourrait participer à la réaction de clivage (Dominski *et al.*, 2005).

Une fois la coupure réalisée, la protéine HBP reste fixée sur la structure en tige-boucle de l'ARNm mature et la snRNP U7 se retrouve appariée à l'ARN de la région 3'. Cette dernière sera alors recyclée par digestion exonucléolytique de type 5'-3' (Walther *et al.*, 1998) (Figure 12).

3.3.3 Export de l'ARNm mature du noyau vers le cytoplasme

Une fois maturés, les ARNm d'histones sont exportés du noyau vers le cytoplasme. Ils sont accompagnés de la protéine HBP fixée sur la structure en tige-boucle. Les ARNm d'histones ne présentant pas d'introns ni de queue polyA, ils se trouvent être dépourvus des signaux canoniques d'export comme les EJC (Exon Junction Complexes) ou la queue polyA. L'export des ARNm d'histones va donc devoir faire intervenir une voie d'export spécifique. Les ARNm d'histones ne sont efficacement exportés dans le cytoplasme que s'ils sont correctement maturés en 3'. En plus de la structure en tige-boucle, des séquences contenues dans la région codante sont également requises pour un export efficace (Eckner *et al.*, 1991). D'autres études ont montré que, dans le cas des gènes H2b-221 et H3.2-614, des ARNm délétés de leur structure en tige-boucle s'accumulaient dans le noyau, suggérant à nouveau l'importance fonctionnelle de l'extrémité 3' non traduite dans le processus d'export (Sun *et*

al., 1992). Les auteurs ont également montré que l'export est indépendant de la traduction cellulaire et que les ARNm, une fois dans le cytoplasme, se localisent par le biais de leur région 3' non traduite au niveau des polysomes (Sun *et al.*, 1992). D'autres travaux obtenus pour l'ARNm H2a-614 ont confirmé l'importance de la structure en tige-boucle et de l'interaction avec HBP dans le processus d'export (Williams *et al.*, 1994). Plus récemment, une étude a permis d'identifier un élément de 22 nucléotides localisé dans la région codante de l'ARNm H2a-614 qui était impliqué dans l'export de cet ARN. Sa transplantation dans d'autres ARNm dépourvus d'introns les rendait exportables (Huang & Carmichael, 1997). Cet élément recrute les protéines adaptatrices de l'export SRp20 et 9G8 (Huang & Steitz, 2001), protéines qui interagissent directement avec le récepteur d'export TAP (Huang *et al.*, 2003). Enfin, des expériences d'injection de transcrits dans les noyaux d'ovocytes de xénope ont confirmé que l'export des ARNm réplication-dépendants est conduit par le facteur TAP. Selon les auteurs, les signaux d'export de l'ARNm ne sont ni la région 3' non traduite, ni la structure en tige-boucle, mais plutôt la longueur de la séquence située en amont de la tige-boucle (Erkman *et al.*, 2005a). Les auteurs ont également pu exclure toute implication de la protéine HBP dans le processus d'export. Un modèle a été proposé dans lequel plusieurs adaptateurs lient plusieurs sites au niveau de l'ARNm en amont de la structure en tige-boucle. Ces adaptateurs permettent le recrutement du récepteur d'export TAP qui conduit l'export hors du noyau par le biais d'interactions avec les protéines du complexe protéique formant les pores nucléaires (Nuclear Pore Complex)(Erkman *et al.*, 2005a). Ces résultats, obtenus à l'aide de transcrits injectés dans les noyaux, contredisent les précédentes études qui avaient conclu au rôle de l'extrémité 3' non traduite dans le processus d'export. À la décharge de ces dernières, il convient de mentionner qu'elles étaient réalisées par une technique différente basée sur l'expression *in vivo* de gènes transfectés.

Après maturation, les ARNm sont conduits dans le cytoplasme. La protéine HBP reste fixée sur les ARNm exportés formant ainsi une particule ribo-nucléoprotéique messagère appelée aussi mRNP (Whitfield *et al.*, 2004). Elle sera retrouvée plus tard dans les polysomes au cours de la traduction (Dominski *et al.*, 1995; Sun *et al.*, 1992). Ceci pourrait expliquer le fait qu'HBP devienne majoritairement cytoplasmique au cours de la phase S du cycle cellulaire (Erkman *et al.*, 2005b). En effet, pour être exportée, la protéine doit impérativement fixer l'ARNm d'histone au niveau de sa structure en tige-boucle car elle ne contient pas de signaux d'export. Elle sera alors exportée de manière passive (Erkman *et al.*, 2005b). Son import dans le noyau s'effectue quant à lui par interaction directe avec les

récepteurs d'import Importin alpha/Importin beta et Transportin-SR2 comme il a été montré *in vitro*. Les expériences menées *in vivo* n'ont toutefois pas permis de confirmer l'implication de la Transportin-SR2. Une fois dans le noyau, HBP sera déplacée du récepteur d'import par une molécule de RanGTP, présente en forte concentration. En fin de phase S, la protéine sera dégradée par un mécanisme impliquant le protéasome.

3.3.4 Traduction

L'initiation de la traduction chez les eucaryotes nécessite un ensemble de facteurs d'initiation de la traduction qui permettent d'obtenir, par le biais d'un réseau d'interactions, la circularisation de l'ARNm, rapprochant ainsi les extrémités 5' et 3'. En ce qui concerne les ARNm polyadénylés, la liaison entre les deux extrémités est réalisée par le facteur eIF4G qui pontre les extrémités en interagissant avec la « Cap Binding Protein » (eIF4E) fixant la coiffe d'une part, et la « PolyA Binding Protein » (PABP) liée à la queue polyA d'autre part (Kapp & Lorsch, 2004) (**Figure 17**).

3.3.4.1 Rôle de la protéine HBP dans la traduction

Les ARNm d'histones n'étant pas polyadénylés au niveau de leur extrémité 3', il a été proposé que la protéine HBP pourrait remplacer la PABP manquante pour stimuler la traduction des ARNm. D'autres observations, comme le fait de trouver la protéine HBP associée aux polysomes tendrait à soutenir cette théorie. Par électroporation de transcrits rapporteurs codant pour la luciférase dans des cellules CHO, il a été montré que la structure en tige-boucle située à l'extrémité 3' des ARNm d'histones avait une fonction équivalente à la queue polyA des ARNm canoniques. De plus, la traduction des ARNm d'histones nécessite la présence de la coiffe au niveau de leur extrémité 5' (Gallie *et al.*, 1996). D'autres expériences ont montré que certaines mutations de la tige-boucle affectant la fixation de la protéine HBP diminuaient la stimulation de la traduction, ce qui suggère que cette protéine doit être présente au niveau de l'ARNm pour permettre une traduction efficace (Gallie *et al.*, 1996).

Une conclusion similaire a été obtenue par des expériences menées dans un contexte cellulaire hétérologue, par électroporation des ARNm rapporteurs dans des cellules de levure exprimant la protéine HBP humaine (Ling *et al.*, 2002). L'expression de formes tronquées de

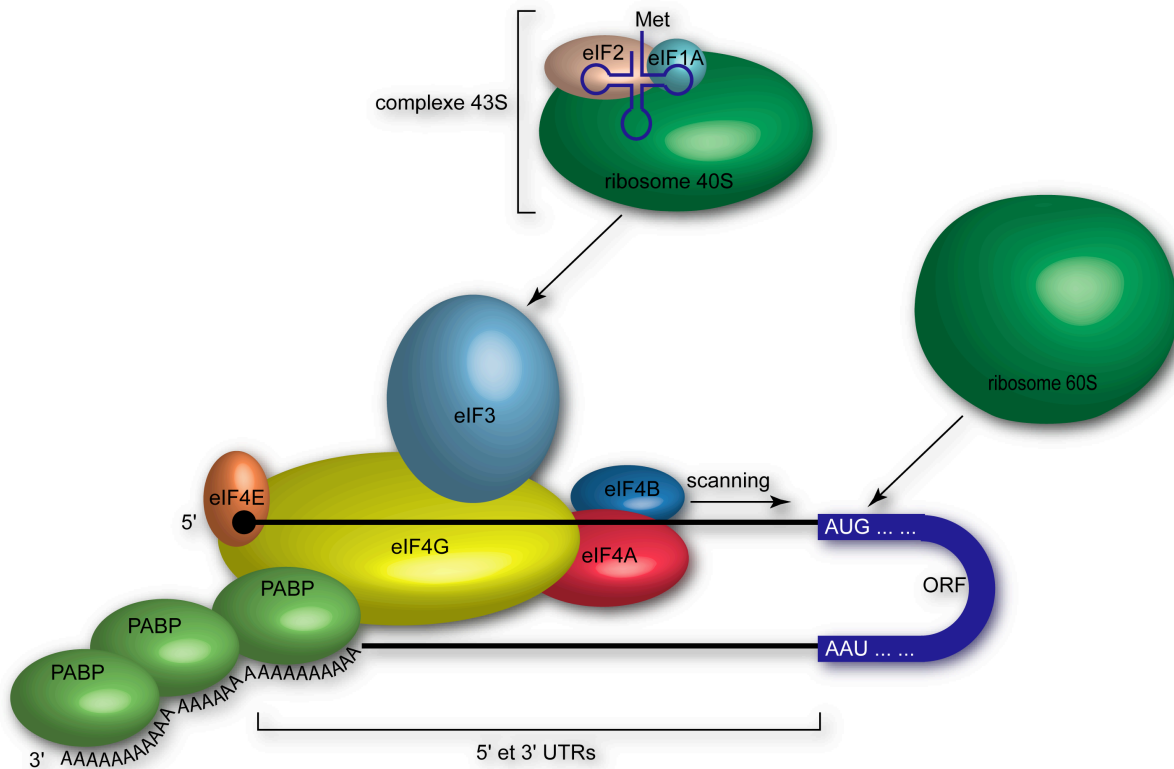


Figure 17 : Initiation de la traduction d'un ARNm eucaryotique polyadénylé. L'initiation de la traduction est un processus complexe qui nécessite l'intervention de nombreux facteurs protéiques. La fonction de ces facteurs est d'assister le ribosome en lui permettant de se positionner correctement sur les ARNm et d'identifier le codon initiateur. L'assemblage du ribosome nécessite de nombreux facteurs appelés eIF (eukaryotic Initiation Factor). Habituellement, les ARNm eucaryotiques sont coiffés à leur extrémité 5' et possèdent une queue polyA en 3'. La coiffe et la queue polyA sont reconnues respectivement par les facteurs eIF4E et la PABP (PolyA Binding Protein). Un autre facteur, eIF4G, sert de plate-forme d'assemblage sur laquelle viennent se fixer les facteurs eIF4E et PABP ce qui permet la circularisation des ARNm. D'autres facteurs viennent se greffer sur eIF4G, il s'agit d'eIF3, d'eIF4A et eIF4B. eIF4A est une ARN hélicase capable de déplier les structures secondaires; eIF4B est une protéine auxiliaire nécessaire à cette activité. Le facteur eIF3 permet le recrutement de la particule 43S qui contient la petite sous-unité du ribosome 40S et le Met-tRNA^{Met} initiateur ainsi que les facteurs eIF2 et eIF1A. Une fois assemblés au niveau de la coiffe, ces facteurs forment alors le complexe 48S. Ce complexe va ensuite glisser (scanning) sur l'ARNm en direction 5'-3' à la recherche du codon initiateur AUG. Lorsque celui-ci est atteint, la grande sous-unité 60S du ribosome s'assemble sur le complexe 48S et la traduction peut débuter.

la protéine HBP a permis de cartographier les régions importantes pour la stimulation de la traduction. Ainsi, la région centrale de la protéine (correspondant à un domaine RBD plus étendu) est requise pour la stimulation. Des tests de traduction semblables, conduits dans différentes souches de levures mutées (mutants nuls ou thermosensibles) pour différents facteurs impliqués dans l'initiation de la traduction, ont permis de mettre en évidence une interaction fonctionnelle de la protéine HBP avec certains de ces facteurs. Par la suite, ces interactions ont été clairement identifiées avec les facteurs eIF4G et eIF4E par co-purification avec HBP sur un support d'affinité couplé à une molécule de coiffe. L'ensemble de ces résultats a conduit à la conclusion que la protéine HBP stimule la traduction des ARNm rapporteurs en se fixant sur la structure en tige-boucle placée à l'extrémité 3' et en

interagissant avec les facteurs de traduction eIF4G, eIF4E et plus faiblement eIF3 de la levure (Ling *et al.*, 2002). Des expériences supplémentaires ont confirmé que la protéine HBP co-immuno précipite le facteur eIF4G issu de cellules de mammifères et le facteur eIF3 moins intensément. De plus, elle nécessite les facteurs eIF4E et eIF4G pour pouvoir interagir avec la coiffe. À l'issue de ces travaux, il a été proposé que la protéine HBP induirait une circularisation de l'ARNm par interaction avec le facteur eIF4G comme le ferait la PABP (Ling *et al.*, 2002).

Le rôle joué par HBP dans la traduction a été confirmé par injection d'ARNm similaires dans des ovocytes de xénope, et en utilisant des systèmes de traduction *in vitro* (Sanchez & Marzluff, 2002). La recherche des régions de la protéine requises pour la stimulation de la traduction a conduit à l'identification d'une séquence de 15 résidus située dans le domaine N-terminal. Le domaine C-terminal peut, quant à lui, être délété sans aucune incidence sur la stimulation de la traduction, contrairement à ce qui avait été montré précédemment (Ling *et al.*, 2002).

D'autres auteurs ont montré par double hybride que la protéine HBP interagit directement avec la sous-unité h du facteur d'initiation eIF3, et que l'interaction avec eIF4G ne s'effectue pas de manière directe contrairement à ce qui avait été démontré par Ling *et al.* (Gorgoni *et al.*, 2005). Ils ont également caractérisé un nouveau facteur interagissant directement avec la protéine HBP, le facteur Paip 1. Il s'agit d'une protéine homologue à la région C-terminale du facteur eIF4G qui participe à la traduction des ARNm polyadénylés. La protéine HBP interagit avec eIF3h par son RBD et avec le facteur Paip 1 par son domaine C-terminal. Par ailleurs, par traduction d'ARNm rapporteurs, les auteurs ont ré-examiné les régions importantes pour la traduction. Seul le domaine N-terminal de la protéine HBP induit une stimulation de la traduction comme précédemment observé (Sanchez & Marzluff, 2002). Les interactions de la protéine HBP avec les facteurs eIF3h et Paip 1 n'induisent donc pas une activation traductionnelle observable. Enfin, il a été démontré que la stimulation induite par HBP nécessite le facteur eIF4E mais sans interaction directe. Les auteurs proposent alors, un nouveau modèle plus complexe, impliquant des facteurs supplémentaires inconnus, mais où la stimulation de la traduction sera toujours une conséquence de la circularisation de l'ARNm induite par la protéine HBP.

L'ensemble des résultats obtenus converge vers un modèle de fonctionnement analogue à celui de la PABP, dans lequel la protéine HBP permettrait la circularisation des ARNm d'histones de type réplication-dépendants par interaction avec le complexe d'initiation. Cette circularisation pourrait promouvoir la formation d'un complexe de pré-

initiation de la traduction d'une part, mais aussi permettre le recyclage des ribosomes placés sur le site de terminaison. Il s'en suivrait une stimulation de l'efficacité de la traduction.

3.3.4.2 Des études très controversées

La plupart de ces travaux ont récemment été remis en cause dans une revue publiée par M. Kozak, dans laquelle certaines réserves ont été émises au sujet de la stimulation induite par la protéine HBP (Kozak, 2004). En premier lieu, les effets de stimulation sembleraient trop faibles pour être significatifs. Ils ne sont parfois pas reproductibles. Les stimulations *in vitro* ne peuvent être obtenues qu'à partir d'ARNm non coiffés, alors qu'il a déjà été démontré que la coiffe est requise pour un effet de HBP. Les interactions physiques ou fonctionnelles entre la protéine humaine et les facteurs de traduction de levures n'auraient, d'après Kozak, aucune véritable signification (Kozak, 2004). D'autre part, dans toutes ces études, les auteurs ont totalement écarté le fait que des signaux internes pourraient exister au sein même de la phase codante des ARNm d'histones. Tous ont utilisé des constructions contenant des gènes rapporteurs au lieu d'étudier les ARNm authentiques. Les dernières études de traduction réalisées par Gorgoni *et al.* utilisent des tests très éloignés du contexte cellulaire. Les auteurs négligent totalement l'éventuelle contribution de la phase ouverte de lecture, ils apportent la protéine HBP, voir même plusieurs facteurs, souvent d'origines différentes.

Par ailleurs, une comparaison des résultats obtenus dans tous ces travaux laisse apparaître quelques contradictions, au niveau des effets de stimulation, mais surtout en ce qui concerne les facteurs impliqués dans la réponse de HBP qui ne sont parfois pas les mêmes. De même, les régions de la protéine impliquées dans la stimulation de la traduction ne correspondent pas à celles interagissant avec les facteurs d'initiation. Ces dernières seront alors paradoxalement considérées comme non requises dans la traduction. Le cas du facteur eIF4G reste énigmatique. Selon Ling *et al.*, il recruterait par interaction directe la protéine HBP au niveau de la coiffe (Ling *et al.*, 2002). Selon Gorgoni *et al.*, il n'y aurait pas d'interaction directe entre ces deux facteurs (Gorgoni *et al.*, 2005).

Pour conclure, remarquons que les études portant sur le rôle de la protéine HBP dans la traduction ne conduisent pas à des résultats très concordants, hormis que la protéine a un effet stimulateur, quoique parfois très faible. Il nous semble indispensable que les études de traduction portant sur les ARNm d'histones puissent s'effectuer sur des ARNm les plus

proches de ceux trouvés dans la cellule tout en respectant une certaine homogénéité dans l'origine des facteurs ajoutés.

La traduction des ARNm d'histones a aussi un effet sur leur stabilité. Certaines études ont effectivement démontré que la dégradation des ARNm d'histones dépend de leur traduction. Ce n'est qu'une fois traduits que les ARNm pourront être dégradés (Graves *et al.*, 1987).

3.3.5 Stabilité des ARNm d'histones

Le taux des ARNm d'histones réplication-dépendants est régulé de manière coordonnée avec la synthèse d'ADN et le cycle cellulaire. Après la régulation de la transcription et de la maturation en 3', la stabilité est une autre étape de régulation dans le métabolisme des ARNm d'histones.

La durée de demi-vie propre aux ARNm d'histones de type réplication-dépendants passe de 45-60 minutes au cours de la phase S du cycle cellulaire à 10 minutes à la fin de cette phase, ce qui confirme que la stabilité des ARNm est également un moyen d'obtenir une régulation de la biogenèse des histones en fonction du cycle cellulaire (Marzluff & Duronio, 2002).

D'anciennes études portant sur la stabilité des ARNm d'histones avaient décrit une activité 3'-5' exonucléolytique capable de dégrader un ARNm d'histone H4 alors qu'il se trouvait lié aux ribosomes (Ross *et al.*, 1987). Cette activité exonucléolytique ne dépendait ni de la synthèse d'ADN ni de celle des protéines (Peltz *et al.*, 1989). Plus récemment, une nouvelle exonucléase baptisée 3' hExo a été caractérisée pour son aptitude à interagir avec la structure en tige-boucle de l'extrémité 3' non codante des ARNm d'histones (Dominski *et al.*, 2003). Étonnamment, cette enzyme est capable d'interagir avec la tige-boucle alors que la protéine HBP y est fixée, en formant un complexe ternaire. Lorsque la tige-boucle est libérée d'HBP, l'enzyme 3'hExo catalyse la dégradation de la tige-boucle dans le sens 3' vers 5' jusqu'à la moitié de la boucle (c'est-à-dire 13 nucléotides) (**Figure 12**). Le reste de l'ARNm serait ensuite dégradé par des enzymes de dégradation plus classiques (Dominski *et al.*, 2003; Marzluff, 2005).

La structure cristallographique du site catalytique de l'exonucléase 3' hExo a récemment été résolue (Cheng & Patel, 2004). La topologie de cette enzyme permet de

l'intégrer à la famille des exonucléases DDEDDh et son orthologue chez *C. elegans*, appelé ERI-1, est capable de dégrader des siARNs *in vitro* (Kennedy *et al.*, 2004).

L'ARNm d'histone et sa structure en tige-boucle sont donc protégés de la dégradation par la présence de la protéine HBP, ce qui suggère qu'elle est une composante clé de la dégradation des ARNm d'histones. Pourtant, si la protéine HBP est maintenue à un taux constant tout au long du cycle cellulaire, c'est-à-dire si elle perd sa propre régulation dépendante du cycle cellulaire, les ARNm d'histones seront tout de même dégradés en fin de phase S, ce qui semble indiquer que la seule présence d'HBP dans la cellule ne suffit pas à protéger l'ARNm (Zheng *et al.*, 2003). Ainsi, le re-largage de la protéine HBP de la tige-boucle serait une étape importante de la dégradation des ARNm puisqu'elle rendra la tige-boucle accessible à l'exonucléase 3' hExo.

D'autres travaux avaient montré que la présence de la tige-boucle à proximité du codon de terminaison était requise pour la dégradation des ARNm d'histones (Graves *et al.*, 1987). Plus récemment, il a été proposé que l'étape de terminaison de la traduction serait impliquée dans la dégradation par le biais du déplacement de la protéine HBP par un ribosome en cours de terminaison. L'ARNm libéré d'HBP ne serait alors plus protégé et subirait la dégradation par l'enzyme 3'hExo (Kaygun & Marzluff, 2005b).

Très récemment, une autre hypothèse a été avancée pour expliquer la dégradation des ARNm d'histones. La protéine HBP fixée sur la tige-boucle serait responsable du recrutement de la protéine Upf1 sur l'extrémité 3' non traduite. Une kinase de type PIKK appelée ATR pour « ATM and Rad 3-related » serait aussi impliquée dans la régulation de leur demi-vie (Kaygun & Marzluff, 2005a; Marzluff, 2005). La protéine Upf1 est impliquée dans le NMD (Nonsense Mediated Decay), la voie catabolique permettant l'élimination des ARNm comportant des codons de terminaison prématurés. Cette protéine est aussi impliquée dans le « Staufen Mediated Decay », une voie de dégradation d'ARNm où la protéine Staufen fixée sur l'extrémité 3' non traduite recrute Upf1 en aval du codon de terminaison et induit la destabilisation de l'ARN (Kim *et al.*, 2005). Dans ce cas, la protéine HBP doit rester fixée à l'ARN et ne doit pas en être relarguée pour que la dégradation s'effectue.

4 ARTICLE 1

Revue d'introduction

« Expression of metazoan replication-dependent histone genes »

Jaeger, S., Barends, S., Giegé, R., Eriani, G. and Martin, F.

Publiée dans Biochimie en 2005

Vol 87, 827-834.

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Expression of metazoan replication-dependent histone genes

Sophie Jaeger, Sharief Barends, Richard Giegé, Gilbert Eriani and Franck Martin

Biochimie, 2005, Vol 87, N° 9-10, Pages 827-834

Pages 827-834:

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur :

<http://dx.doi.org/10.1016/j.biochi.2005.03.012>

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP:

peb.sciences@scd-ulp.u-strasbg.fr

AVANT-PROPOS DE LA THÈSE

Les histones sont des protéines essentielles qui s'associent à l'ADN pour former la chromatine. La plupart des histones sont issues de gènes dont l'expression est synchronisée avec la réplication de l'ADN, leur valant le qualificatif de « réplication-dépendants ». Cette propriété les place en dehors des schémas classiques d'expression des gènes. En effet, certaines étapes de leur expression sont atypiques et leur permettent d'être exprimés de façon abondante et ciblée au cours de la phase S du cycle cellulaire.

Mon travail de thèse a consisté en l'étude de deux étapes atypiques de l'expression des gènes d'histones dits « réplication-dépendants ». Les résultats obtenus durant une période de 4 années sont présentés dans les 3 chapitres qui suivent ; les grandes lignes sont brièvement présentées ci-dessous.

Dans une première étude, nous nous sommes intéressés à la maturation de l'extrémité 3' non traduite des ARNm d'histones qui consiste en une coupure endonucléolytique se produisant après la fixation de la protéine HBP et de la snRNP mineure U7.

Nous avons tout d'abord étudié l'interaction entre la protéine HBP et l'ARNm de l'histone H4-12. Cette interaction avait déjà fait l'objet de nombreuses études car elle joue un rôle essentiel dans l'expression des gènes d'histones. Une fois fixée sur la structure en tige-boucle, la protéine HBP va intervenir non seulement au niveau de la maturation de l'extrémité 3' mais également dans d'autres étapes de l'expression des gènes d'histones. L'étude du complexe HBP:ARNm comporte un autre intérêt car il représente un type d'interaction ARN-protéine qui n'a, jusqu'à présent, pas été caractérisé d'un point de vue structural.

Pour étudier cette interaction, nous avons choisi une méthode de sélection de mutants dans la levure basée sur la technique du triple hybride. Rappelons qu'au commencement de cette étude nous ne disposions d'aucune donnée structurale pour nous orienter et d'un très faible nombre de séquences d'HBP. Le complexe HBP:tige-boucle avait déjà fait l'objet d'une étude *in vivo* par cette méthode et des mutants des deux partenaires avaient été étudiés (Martin *et al.*, 2000). C'est sur certains de ces mutants d'HBP incapables d'interagir avec l'ARN que nous avons effectué notre étude. Le problème que nous avons choisi de résoudre

consistait à rendre à ces mutants leur capacité à fixer la tige-boucle d'ARN par contre-sélection de nouvelles mutations compensant l'effet négatif initial. En effet, la recherche de supresseurs intra-géniques que nous avons effectuée est une approche couramment utilisée pour mettre en évidence les relations fonctionnelles qui peuvent exister entre les différents domaines d'une protéine. Nous verrons plus loin que les mutants sélectionnés ont montré l'importance des domaines N- et C-terminaux dans le processus de discrimination des tiges-boucles. Ces derniers participeraient à la mise en conformation du domaine central de liaison à l'ARN (RBD) de manière à accroître la spécificité de reconnaissance pour la structure en tige-boucle sauvage.

À la suite de ce travail, nous avons réalisé une seconde étude portant sur l'étape de maturation de l'extrémité 3' en nous intéressant plus particulièrement à l'effet de la fixation de la protéine HBP sur la structure de l'ARN. Des travaux antérieurs avaient montré que la fixation d'HBP pouvait stimuler la réaction de maturation-clivage de l'ARN pré-messager (voir introduction). Nous avons recherché si des changements structuraux pouvaient en être la cause. Les techniques de sondage en solution par les RNases et les agents chimiques ont été appliquées sur les extrémités 3' non traduites des histones H4-12, H1t et H2a-614. Nous verrons plus loin que dans les 3 ARNs la séquence d'hybridation à la snRNP U7 (la séquence HDE) n'est pas libre mais engagée dans des repliements secondaires qui pourraient limiter son accessibilité à la snRNP U7. Nous verrons encore que l'analyse de l'empreinte d'HBP sur les trois ARNs révèle des remaniements structuraux plus ou moins importants dans la région de la séquence HDE, remaniements qui se traduisent par une efficacité d'hybridation accrue pour un transcrit de snRNA U7. La protéine HBP pourrait donc augmenter l'accessibilité de l'élément HDE à la snRNP U7.

Le dernier volet de ce travail a porté sur l'étape de traduction de l'ARNm de l'histone H4. De vives controverses ont animé les études réalisées sur la traduction des ARNm d'histones (voir introduction § 3.3.4.2). Cela nous a conduit à repenser les études de traduction en utilisant des ARNm entiers que nous avons traduits *in vitro* dans des extraits acellulaires, et cela en opposition complète avec les études antérieures qui avaient été effectuées sur des ARNm rapporteurs. Nous verrons dans ce mémoire que l'ARNm H4-12 de souris peut être traduit aussi efficacement qu'un ARN viral et cela même en l'absence des

séquences 5' et 3' non traduites qui sont habituellement requises pour l'initiation de la traduction. Cette observation suggère que la séquence codante de l'ARNm contient un site interne d'entrée du ribosome (IRES).

Faisant suite à cette observation, nous avons effectué une étude structure-fonction sur l'ARNm en combinant sondage en solution et tests d'activité de traduction *in vitro*. La molécule native ainsi qu'un ensemble d'ARNm mutés ont été testés, conduisant à un modèle de structure secondaire, le premier à être réalisé sur un ARNm complet. Les tests de « toe print » ainsi que la microscopie électronique ont montré que 2 sites de fixation de ribosome existeraient sur l'ARNm. Un des sites serait situé au niveau de l'extrémité 3' de la séquence codante, l'autre serait situé au niveau du codon d'initiation. L'efficacité très élevée de traduction s'expliquerait par le fait que l'un des ribosomes pourrait parvenir directement sur l'AUG initiateur sans « scanning », en utilisant probablement un nombre réduit de facteurs d'initiation ; quant au second ribosome situé à la fin de la phase codante, il pourrait atteindre l'AUG initiateur par un phénomène de « ribosome shunting » qui assurerait un démarrage instantané de la traduction au lieu de se dissocier de l'ARNm comme le font habituellement les ribosomes ayant atteint le codon stop.

CHAPITRE 1

**Recherche des résidus de la protéine HBP
impliqués dans la reconnaissance de la structure
en tige-boucle des ARN messagers d'histones**

A conduit à la publication

Article 2

**« Critical residues for RNA discrimination of the histone hairpin binding protein (HBP)
investigated by the yeast three-hybrid system »**

Sophie Jaeger, Gilbert Eriani, Franck Martin

Publié dans FEBS Letters en 2004, vol. 556, pages 265-270.

1 INTRODUCTION

Avant que nous abordions cette première étude portant sur l'interaction entre la protéine HBP et l'ARNm de l'histone H4-12, de nombreuses études avaient montré que la formation du complexe HBP:tige-boucle des ARNm d'histones jouait un rôle prépondérant dans l'expression des gènes d'histones (voir introduction § 3.3).

Le clonage et le séquençage de la protéine HBP humaine dans deux laboratoires indépendants (Martin *et al.*, 1997; Wang *et al.*, 1996) a ouvert la voie à des études plus moléculaires sur la protéine et son complexe formé avec l'ARN.

Le gène codant pour la protéine HBP a été cloné par la technique du triple hybride à l'aide d'un appât constitué de la structure en tige-boucle de l'ARNm d'histone H4-12 de souris. La protéine HBP (270 résidus) se divise en 3 domaines fonctionnels comme l'on montré des expériences de délétion des extrémités N- et C-terminales (Martin *et al.*, 2000; Wang *et al.*, 1996). Le domaine central est appelé RBD pour « RNA Binding Domain ». Il correspond au domaine minimal requis pour la fixation de l'ARN. Il est flanqué de 2 domaines N- et C-terminaux qui sont non essentiels pour la fixation à l'ARN mais cruciaux pour la spécificité de reconnaissance de la tige-boucle (voir plus bas).

La méthode du triple hybride a été aussi utilisée pour décortiquer l'interaction HBP:tige-boucle. Par une sélection négative, des mutants d'HBP incapables d'interagir avec l'ARN ont été identifiés. Les mutations portées par HBP étaient toutes localisées dans le domaine central RBD, ce qui démontrait à nouveau son importance capitale pour l'interaction avec l'ARN (Martin *et al.*, 2000). Par ailleurs, la mutagenèse dirigée des résidus conservés du RBD, avait permis d'identifier un certain nombre de résidus impliqués dans l'interaction avec l'ARN, dont certains étaient redondants avec ceux identifiés par la méthode du triple hybride (Dominski *et al.*, 2001). Les domaines N- et C-terminaux ne sont cependant pas dépourvus d'importance. Des mutations y ont été isolées lors d'une sélection positive de mutants d'HBP capables de reverser l'effet de mutations dans la tige-boucle (Martin *et al.*, 2000). Ainsi, les domaines N- et C-terminaux de la protéine qui bordent le domaine RBD jouent un rôle non négligeable dans le processus de reconnaissance des nucléotides de l'ARN en modulant la spécificité de reconnaissance (Martin *et al.*, 2000).

Dans ce chapitre, je vais présenter l'étude qui a consisté à isoler, par la technique du triple hybride, des mutants de la protéine HBP capables de compenser les effets de premières mutations dans HBP qui abolissaient la fixation à la tige-boucle sauvage. En d'autres termes,

nous avons isolé des supprimeurs intragéniques qui permettaient de retrouver un phénotype de fixation à l'ARN. Expérimentalement, nous avons muté de manière aléatoire les gènes d'HBP codant pour des protéines inactives en fixation à l'ARN, puis nous avons sélectionné parmi les clones de la banque, les rares clones qui avaient reconstitué un triple hybride fonctionnel. Ces clones étaient porteurs de gènes d'HBP supprimant l'effet négatif de la mutation initiale. La recherche des supprimeurs intragéniques fournit des renseignements importants quant aux relations fonctionnelles ou structurales pouvant exister entre des domaines et des résidus distincts d'une protéine.

Dans les lignes qui suivent, j'exposerai le principe de l'approche du triple hybride utilisée puis je décrirai et analyserai succinctement les résultats obtenus.

2 LA MÉTHODE DU TRIPLE HYBRIDE

La méthode du triple hybride (SenGupta *et al.*, 1996) permet d'étudier l'interaction entre un ARN et une protéine dans la levure *Saccharomyces cerevisiae* (**Figure 18**). Elle consiste à co-exprimer trois hybrides dans la souche L40-coat. Deux des hybrides sont des chimères de protéines, le troisième est une chimère d'ARN. Le premier hybride protéique est constitué du domaine de liaison à l'ADN du répresseur LexA d'*E. coli* fusionné à la protéine de capsid du phage MS2. Cette dernière a la particularité d'interagir avec une séquence de 22 nucléotides de l'ARN génomique du phage MS2. Le second hybride est un ARN comportant ce site de liaison de 22 nucléotides de la protéine de capsid fusionné à la molécule X à tester. Enfin, le troisième hybride est une protéine de fusion entre la protéine analysée Y et le domaine d'activation de la transcription du trans-activateur Gal4 de levure. L'interaction entre l'ARN X et la protéine Y va permettre de reconstituer un trans-activateur fonctionnel qui va recruter la machinerie de transcription en amont des 2 gènes rapporteurs *HIS3* et *LacZ*.

La mesure du niveau d'activation des gènes rapporteurs permet d'évaluer la force d'interaction entre les partenaires X et Y et donc l'affinité de la protéine pour l'ARN.

Pour estimer le niveau d'expression du rapporteur *LacZ*, les cellules de levures sont cultivées sur un milieu sélectif supplémenté en X-Gal. Une coloration bleue indique l'activation du rapporteur *LacZ* et donc l'interaction ARN:protéine.

L'activité du gène *HIS3*, un des gènes de la voie métabolique de l'histidine, est mesurée par culture sur un milieu dépourvu d'histidine en présence de 3AT (3-aminotriazole), un inhibiteur compétitif du produit du gène *HIS3*. Le niveau d'expression d'*HIS3* (et son activité par conséquent) est évalué par des tests de viabilité des cellules sur milieux de culture

contenant des concentrations croissantes en 3AT. Plus l'expression du gène *HIS3* est élevée, plus la concentration de 3AT nécessaire pour inhiber la croissance de la souche de levure est élevée.

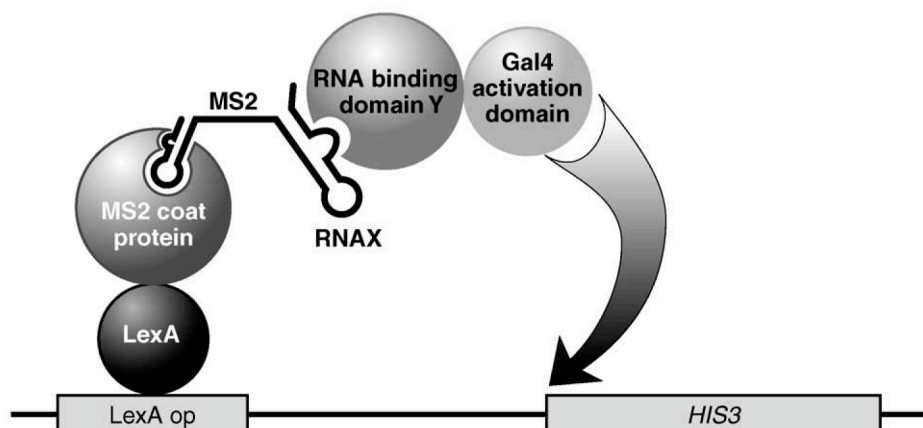


Figure 18 : Principe de la technique du triple hybride (SenGupta *et al.*, 1996). L'interaction entre trois molécules hybrides dans la levure permet de reconstituer un transactivateur fonctionnel et d'activer l'expression des gènes rapporteurs *HIS3* et *LacZ*. Les transformants obtenus sont auxotrophes pour l'histidine [His^+], et bleus sur milieu contenant de l'X-gal [LacZ^+].

Bien que simple dans son principe, la technique du triple hybride est relativement fastidieuse à mettre en œuvre et doit être accompagnée de très nombreux contrôles qui visent à éliminer les signaux faux-positifs qui sont trouvés en majorité. Les avantages du triple hybride sont cependant considérables. Nous n'en citerons que deux qui ont été essentiels pour notre étude. Il permet de cribler de très grandes bibliothèques de gènes mutés aléatoirement. D'autre part, l'expression du gène HBP dans la levure conduit à une protéine fonctionnelle en fixation, alors que la protéine HBP ne peut être surproduite et purifiée dans *E. coli*.

Pour mieux comprendre le système du triple hybride dans son principe ainsi que ses variants, nous invitons le lecteur à lire l'**article 3** situé à la fin de ce chapitre. Dans cette mini revue, nous y avons aussi exposé les avantages et inconvénients de la technique, et énuméré les études réalisées avec succès depuis 1996.

3 RÉSULTATS

Avant d'aborder cette étude, une sélection négative entreprise au laboratoire avait permis d'isoler des mutants de la protéine HBP dont la fixation à la structure en tige-boucle des ARNm d'histones était abolie. Le phénotype de ces mutants en triple hybride était donc [His^-] et [Lac^-] ce qui se traduit par une absence de croissance sur milieu sans histidine et une coloration blanche sur milieu supplémenté en X-gal et histidine. Parmi les différents mutants

disponibles au laboratoire, nous avons choisi 5 mutants dont les phénotypes étaient très sévères (Martin *et al.*, 2000). Nous les avons utilisés pour contre-sélectionner des mutations de réversion dans un criblage positif restaurant la fixation de la protéine sur la tige-boucle.

Ainsi, 2 µg des vecteurs pAct-HBP contenant les mutations initiales G₁₄₅R, P₁₇₂S, P₁₇₂L, R₁₈₁H et D₁₈₄N ont été soumis à un traitement mutagène par l'hydroxylamine en vue de l'obtention de banques de gènes mutés aléatoirement. Les banques obtenues ont servi à transformer la souche L40-coat exprimant les deux autres hybrides nécessaires au test d'interaction. Les cellules ont été sélectionnées sur milieu contenant un niveau basal de 10 mM 3AT. Les colonies [His⁺] ont ensuite été repiquées sur milieu contenant du X-gal afin de détecter la coloration bleue reflétant l'activation de la transcription de *LacZ*. Les clones « faux positifs » ont été éliminés grâce aux contrôles habituels (retransformation et reclonage). Les clones [His⁺, LacZ⁺] restants ont été séquencés.

Nous avons criblé 215.000 transformants dans les 5 banques. Nous avons identifié 7 mutations compensatoires capables de restaurer la fixation de la protéine sur la structure en tige-boucle sauvage.

Nous avons ensuite mesuré la force d'interaction des complexes ARN:HBP en mesurant le niveau d'expression du gène *HIS3* par des tests en gouttes sur des milieux contenant des concentrations croissantes de 3AT. Nous avons pu constater une amélioration de la fixation de l'ARN dans tous les cas. Puis, nous avons vérifié la spécificité de reconnaissance des mutants isolés par test en gouttes en présence de tige-boucles mutantes dont l'affinité pour HBP sauvage est très réduite (Martin *et al.*, 2000). Dans ce test de spécificité, aucune interaction forte n'a pu être mise en évidence avec les tiges-boucles mutantes. Les mutations compensatoires obtenues dans ce crible permettent donc d'augmenter l'affinité d'HBP pour la tige-boucle sauvage sans qu'il n'y ait d'élargissement ou de relaxation de la spécificité de reconnaissance.

Enfin, des expériences de gel-retardement ont confirmé que les mutations compensatoires isolées dans ce travail étaient bien capables de restaurer *in vitro* la fixation à la tige-boucle sauvage.

4 DISCUSSION

Les 5 mutations inactivantes précédemment identifiées lors de la première sélection négative (Martin *et al.*, 2000) se situaient toutes dans le domaine minimal de liaison à l'ARN appelé RBD. À partir de ces mutations, nous avons isolé 7 mutations compensatoires

restaurant la fixation de la protéine. Le séquençage des mutants a montré que 6 d'entre elles étaient situées dans les régions N- et C-terminales flanquant le domaine RBD (**Figure 1 de l'article 2**). Une seule mutation a été trouvée dans le RBD en suppression de la mutation P₁₇₂L. Il nous a été impossible d'isoler des révertants du mutant P₁₇₂S, ce qui suggère qu'une ou plusieurs mutations supplémentaires ne suffisent pas pour reverser l'effet de la mutation initiale. La distribution des mutations de compensation hors du RBD suggère l'existence de connexions fonctionnelles entre les domaines N-, C-terminaux et le RBD. Par ailleurs, le séquençage des mutations a révélé l'apparition de résidus chargés positivement en lieu et place de résidus chargés négativement ; ils pourraient permettre de nouvelles interactions avec l'ARN. Enfin, la meilleure fixation à la tige-boucle est obtenue pour une mutation qui enlève les 62 derniers résidus de la protéine (la quasi-totalité du domaine C-terminal). La perte du domaine C-terminal aurait donc un effet comparable à celui des mutations ponctuelles sélectionnées dans les domaines N- et C-terminaux. Ce mutant montre également que le domaine C-terminal joue un rôle négatif sur la fixation des tiges-boucles mal positionnées sur un RBD muté.

L'ensemble de ces résultats suggère que les domaines N- et C-terminaux d'HBP pourraient contrôler la fixation initiale de la tige-boucle sur le RBD et participer à la discrimination des structures d'ARN non spécifiques. Leur présence permettrait de guider la tige-boucle vers son site de fixation et d'améliorer la spécificité de reconnaissance. Cela expliquerait aussi qu'une mutation ponctuelle puisse leur permettre de reverser l'effet d'une mutation dans le site de fixation du RBD.

Pour conclure, rappelons que ces mutations mettent en évidence des effets de compensation auxquels il est tentant d'attribuer une responsabilité directe. Néanmoins, il convient de garder à l'esprit que ces effets peuvent n'être que l'aboutissement de changements structuraux à distance induits par les mutations secondaires. La résolution de la structure tridimensionnelle de la protéine devrait en principe permettre de trancher entre la participation directe ou l'effet indirect de ces mutations.

5 ARTICLE 2

« Critical residues for RNA discrimination of the histone hairpin binding protein (HBP) investigated by the yeast three-hybrid system »

Sophie Jaeger, Gilbert Eriani, Franck Martin

Publié dans FEBS Letters en 2004, vol. 556, pages 265-270.

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Critical residues for RNA discrimination of the histone hairpin binding protein (HBP) investigated by the yeast three-hybrid system

Sophie Jaeger, Gilbert Eriani and Franck Martin

FEBS Letters, 2004, Vol 556, N°1-3, Pages 265-270

Pages 265-270:

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur :

[http://dx.doi.org/10.1016/S0014-5793\(03\)01433-9](http://dx.doi.org/10.1016/S0014-5793(03)01433-9)

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP:

peb.sciences@scd-ulp.u-strasbg.fr

6 ARTICLE 3

Mini revue

« Results and prospects on the yeast three-hybrid system »

Sophie Jaeger, Gilbert Eriani, Franck Martin

Publié dans FEBS Letters en 2004, vol. 556, pages 7-12.

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Results and prospects of the yeast three-hybrid system

Sophie Jaeger, Gilbert Eriani and Franck Martin

FEBS Letters, 2004, Vol 556, N°1-3, Pages 7-12

Pages 7-12:

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur :

[http://dx.doi.org/10.1016/S0014-5793\(03\)01434-0](http://dx.doi.org/10.1016/S0014-5793(03)01434-0)

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP:

peb.sciences@scd-ulp.u-strasbg.fr

CHAPITRE 2

**Étude des changements structuraux
induits par la protéine HBP lors de sa fixation
sur les extrémités 3' des ARN pré-messagers d'histones**

A conduit à la publication

Article 4

en préparation :

**« Binding of human SLBP on the 3' UTR of histone precursor mRNA induces
structural rearrangements that enable U7 snRNP anchoring »**

Sophie Jaeger, Franck Martin, Joëlle Rudinger, Richard Giegé et Gilbert Eriani

¹ : co-premier sauteurs

1 INTRODUCTION

Les histones sont des protéines ancestrales apparues très tôt, avant même la divergence des différents eucaryotes. Leur conservation est très grande, tant au niveau protéique que nucléique ce qui suggère l'existence d'une pression de sélection sur la structure de l'ARNm lui-même (Huynen *et al.*, 1992). Les régions 5' et 3' non codantes sont moins conservées. Elles auraient été acquises au cours du temps pour permettre un contrôle de l'expression des gènes d'histones, celles-ci étant très toxiques pour la cellule en cas d'expression en dehors de la phase S du cycle cellulaire. La diversité des séquences retrouvées dans les régions non codantes suggère que différentes stratégies de contrôle ont été développées afin de confiner l'expression de ces gènes. Par exemple, les plantes et les levures présentent des extrémités 3' polyadénylées, alors que les métazoaires ont acquis une machinerie de maturation de l'extrémité 3' qui fait appel à plusieurs protéines et la snRNP U7. Ces différents facteurs pourraient contrôler l'expression des histones, au travers de la maturation, de la traduction, de la dégradation des ARNm. La participation de la protéine HBP a été observée dans ces différentes étapes. Son rôle dans la maturation et la traduction a été particulièrement étudié, mais il se dégage de ces différents travaux un certain nombre de controverses qui découlent du choix des substrats, ou des systèmes de maturation utilisés, parfois même de la manière d'interpréter les données (voir introduction). Néanmoins, un consensus se dégage de ces études pour attribuer à la fixation d'HBP un effet stimulateur, mais non essentiel, sur la maturation de l'extrémité 3'. *In vivo* la situation est plus claire ; la protéine HBP est bien essentielle dans les cellules humaines (Zhao *et al.*, 2004; Wagner *et al.*, 2005) et dans *C. elegans* (Kodama *et al.*, 2002; Pettitt *et al.*, 2002) et les lignées « knockout » de *D. melanogaster* sont viables mais stériles. Ainsi, ces résultats soulignent bien la difficulté à tester le rôle d'une protéine dans un processus aussi complexe que la maturation de l'extrémité 3', alors que tous les partenaires engagés ne sont pas encore identifiés. Le test cellulaire *in vivo* intègre quant à lui toutes les interactions avec les autres facteurs ainsi que les éventuelles autres fonctions. Il en ressort que la protéine HBP est bien essentielle chez les métazoaires.

Dans le cadre de la maturation de l'extrémité 3', nous nous sommes intéressés à une des premières étapes de la maturation qui correspond à la fixation de la protéine HBP sur la tige-boucle des ARNm. Des travaux ont suggéré qu'HBP stimule la maturation en stabilisant l'ancrage de la particule U7, plus précisément en renforçant la capacité de l'ARN à recruter

les particules U7 (Dominski *et al.*, 1999). Cet effet disparaît si le HDE est artificiellement éloigné du site de fixation de la protéine. Les mêmes auteurs ont aussi suggéré que la stabilité du duplexe formé entre le HDE et le snRNA U7 déterminerait les exigences en protéine HBP pour la réaction de maturation. Ainsi, plus le duplexe est stable, moins la protéine HBP est requise dans la réaction de maturation (Dominski *et al.*, 1999). Par la suite, alors que la présente étude était engagée, les mêmes auteurs ont identifié la protéine ZFP100 qui établit un pont protéique entre la snRNP U7 et le complexe HBP:tige-boucle (**Figure 1 de l'article 4**). Cette interaction permettrait à HBP de faciliter l'ancrage de la particule au niveau de l'ARN pré-messager (Dominski *et al.*, 2002).

Ces études ont montré que la protéine HBP agirait plus ou moins directement sur la formation de l'hétéroduplexe ARN pré-messager:snRNA U7. Ce rôle pourrait être plus ou moins essentiel en fonction de la stabilité de l'hétéroduplexe formé. Au laboratoire, nous avons recherché une approche qui nous permettrait de visualiser l'impact structural que pourrait avoir la fixation de la protéine HBP sur l'ARN pré-messager. Nous avons retenu l'approche du sondage en solution qui permet d'établir la structure secondaire de molécules d'ARN et d'examiner l'empreinte résultant de la fixation d'une protéine sur ces molécules. Nous avons sondé plusieurs ARN pré-messagers d'histones en présence ou non de protéine HBP. L'impact de la fixation d'HBP sur l'hybridation du snRNA U7 a été également examiné. Ces travaux sont décrits plus en détail ci-dessous et dans l'article 4.

2 RÉSULTATS

Dans un premier temps, nous avons choisi des gènes d'histones pour effectuer notre étude. Le choix s'est orienté vers 3 gènes d'histones différentes, choix justifié par l'existence de travaux antérieurs ayant porté sur la maturation et par la connaissance précise des extrémités 3' non maturées (rappelons que ces ARNs ne sont pas polyadénylés et ne figurent donc pas dans les banques de cDNAs). Nous avons retenu le gène H4-12 pour lequel HBP est essentielle pour la maturation de l'extrémité 3' (Martin *et al.*, 1997). Nous avons également choisi les gènes des histones H1t et H2a-614 dont la maturation avait été étudiée (Dominski *et al.*, 1999). H1t avait montré une dépendance en HBP pour la maturation en 3' alors qu'H2a-614 était indépendant d'HBP.

Les études en solution, qui font l'objet de ce chapitre, ont été réalisées sur les régions 3' non traduites, c'est-à-dire débutant au nucléotide qui suit le codon stop, jusqu'à la fin identifiée de l'ARN pré-messager (voir **figure 1 de l'article 4** pour les séquences

nucléotidiques). Ces fragments de gènes ont été construits par assemblage d'oligonucléotides. Ils ont ensuite servi de matrices pour la transcription par l'ARN polymérase du phage T7. Les transcrits purifiés ont été marqués à la radioactivité pour réaliser les sondages « directs ». Seul le transcrit de l'ARN pré-messager **entier** de H4-12 a été sondé par la méthode « indirecte », c'est-à-dire que les modifications réalisées sur l'ARN froid ont été révélées par l'extension d'une amorce marquée. Les réactivités observées dans ce cas étaient quasiment identiques à celle des régions 3' isolées.

Des gels représentatifs des sondages réalisés sur les 3 ARNs sont présentés dans les **figures 2, 3 et 4 de l'article 4**. On y retrouve les résultats des coupures et des modifications aux RNases V1, T1, T2 et aux ions Pb^{2+} , ainsi que les modèles des structures secondaires obtenus par M-fold (Zuker, 2003), rectifiés manuellement.

L'examen de ces structures révèle des éléments communs et des grandes dissemblances, qui étaient attendues compte tenu de l'absence d'homologies de séquences entre les régions 3' non codantes.

Un élément structural commun retrouvé dans ces 3 structures est la structure en tige-boucle sur laquelle la protéine HBP se fixe. Sa structure est clairement définie dans les trois cas, bien que quelques différences mineures apparaissent, en particulier lors de la fixation de HBP (voir discussion de chaque cas plus bas). Dans les 3 modèles, la tige-boucle est suivie d'une région simple brin dans laquelle les résidus du site de maturation sont souvent accessibles aux sondes spécifiques des résidus non appariés. Par ailleurs, cette région présente des répétitions C/A qui sont caractéristiques des régions fragiles. Dans certains cas, nous y avons observé des coupures spontanées en l'absence de réactif. Enfin, le dernier point commun aux 3 modèles est l'appariement du motif HDE (en jaune sur les figures) avec la région située en amont de la tige-boucle fixant HBP. La nature et la force de l'appariement sont néanmoins très différentes d'un ARN à l'autre.

ARNm H4-12 : une extrémité 3' non traduite très structurée et sensible à la fixation d'HBP

Le modèle de structure secondaire établi pour l'ARN H4-12 est présenté dans la **figure 19**. Sa forme évoque très sommairement celle d'une feuille de trèfle. Nous y avons défini 4 domaines en numérotant les 3 tiges-boucles I à III ainsi qu'un bras IV renfermant 6 nucléotides non appariés. L'élément HDE qui doit s'hybrider au snRNA U7 est presque entièrement apparié au niveau des bras des domaines III et IV.

L'empreinte résultant de la fixation de la protéine HBP sur le domaine I est visible à plusieurs niveaux. Les domaines I et II présentent tous deux des diminutions de réactivité qui

suggèrent que la protéine s'y fixe en masquant ces résidus. Une face du bras du domaine I est cependant non réactive aux sondes, ce qui ne permet pas de conclure à la fixation d'HBP sur ces résidus. Des effets remarquables sont obtenus sur les domaines III et IV porteurs du HDE. Le bras III, qui contient la première moitié du HDE devient plus accessible aux coupures V1 ce qui suggère que la fixation de HBP sur le domaine I libère ce bras des contraintes structurales dans lesquelles il se trouvait dans l'ARN libre. Des réactivités « simple brin » font aussi leur apparition dans la boucle III. Parallèlement à cette libération du bras III on note une augmentation de la réactivité aux sondes « simple brin » dans la seconde moitié du HDE qui est localisée dans le bras IV. Ainsi, cette partie du HDE semble se déshybrider de son vis-à-vis, pour éventuellement devenir accessible au snRNA U7.

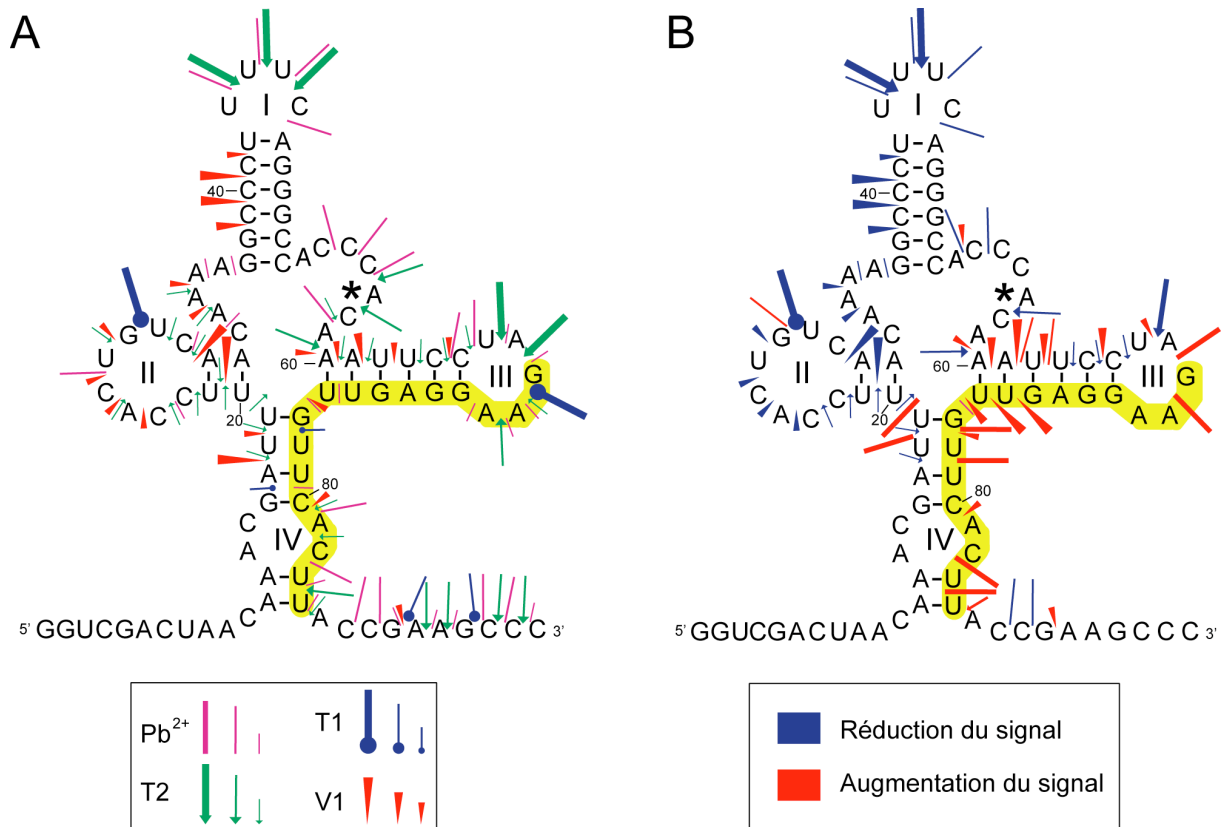


Figure 19 : Modèle de structure secondaire de l'extrémité 3' non traduite du pré-ARNm H4-12 libre et effet structuraux induits par la fixation de la protéine HBP. (A) Repliement de l'extrémité 3' non traduite de l'histone H4-12. Trois intensités de coupure sont décrites pour chaque sonde. L'élément HDE est surligné en jaune et le site de maturation est marqué d'un astérisque. (B) Modèle issu du sondage de l'ARN en présence de protéine HBP. Les changements induits par la fixation de HBP ont été reportés sur le modèle. Les signaux dont l'intensité est diminuée lors de la fixation de la protéine sont représentés en bleu. En rouge sont représentées les réactivités accrues ou nouvelles.

Ces résultats suggèrent que la fixation de la protéine HBP sur le domaine I et II entraîne un relâchement de la structure tertiaire de l'ARN qui se traduit dans les domaines III et IV par une déstabilisation avec ouverture du bras IV et une exposition nouvelle du bras III

aux différentes sondes. L'effet de la fixation d'HBP est donc très important sur la structure secondaire et probablement tertiaire de l'ARN.

ARNm H1t : une structure secondaire plus relâchée qui est sensible à la fixation d'HBP

Le modèle de repliement de l'extrémité 3' non traduite est présenté dans la **figure 20**. Nous y avons à nouveau désigné 4 domaines par analogie avec l'ARN H4-12. Des différences notables existent entre les deux molécules. Premièrement, le bras du domaine I présente des réactivités sur les 2 faces. On peut en déduire que cette face dans H4-12 est masquée par un autre domaine de son ARN. Il pourrait s'agir du domaine III qui présente la possibilité d'établir une interaction à distance avec la boucle I. Le domaine III du gène H1t forme une large boucle alors que la structure du domaine II rappelle celle qui est trouvée dans l'ARN H4-12. La quasi totalité du HDE se trouve apparié dans le domaine IV.

La fixation de la protéine HBP induit plusieurs effets remarquables. La protéine se fixe, comme dans le cas d'H4-12, sur les domaines I et II, mais il s'en suit une augmentation de la réactivité de la face 3' du bras I. Ceci démontre qu'HBP n'interagit qu'avec une face du bras, la seconde restant libre et accessible. Ce résultat est en accord avec des travaux qui ont montré que la protéine 3' hExo pouvait accéder à cette partie de la molécule en présence d'HBP fixée sur le domaine I (Dominski *et al.*, 2003). Dans la boucle III, on note une augmentation des réactivités aux environs du site de maturation, qui suggère une nouvelle accessibilité de ce site pour la réaction de clivage. Le bras du domaine II s'ouvre, comme le montre l'apparition de nouvelles coupures aux ions Pb^{2+} . Enfin, une grande partie des bras du domaine IV devient bien plus accessible aux coupures « double-brin », alors que le haut du bras s'ouvre comme le montrent les réactivités nouvelles aux ions Pb^{2+} (réactivité des résidus de la face opposée).

Ces résultats montrent que la fixation de la protéine HBP sur le gène H1t induit des changements de conformation de l'ARN. Mais dans ce cas, nous observons une déshybridation du côté 5' de la séquence du HDE, prélude éventuel d'une hybridation à l'ARN de la snRNP U7. Contrairement à l'ARN H4-12, l'ARN H1t apparaît moins fortement structuré au niveau de sa structure secondaire, il est néanmoins capable de transmettre un signal structural à l'autre extrémité de sa molécule consécutivement à la fixation de la protéine HBP.

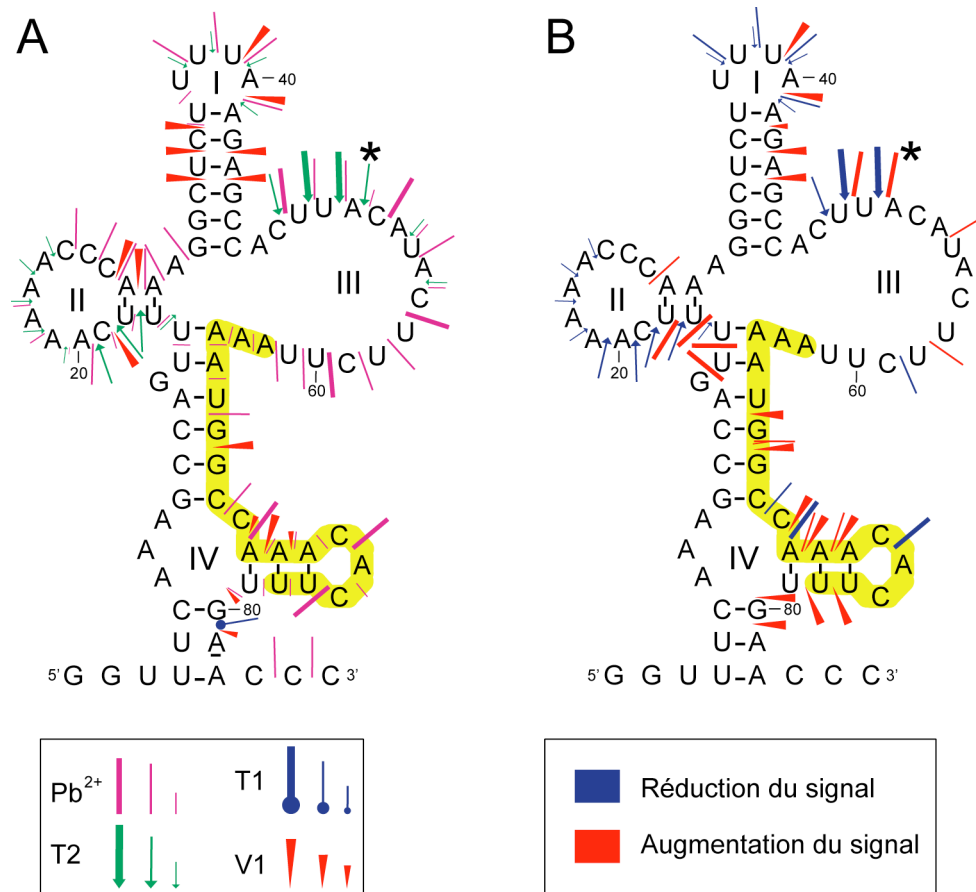


Figure 20 : Modèle de structure secondaire de l'extrémité 3' non traduite de l'ARN pré-messager H1t. Le code couleur de présentation est identique à celui utilisé dans la figure précédente.

ARNm H2a-614 : une structure secondaire relâchée insensible à la fixation d'HBP

Le modèle de structure secondaire établi pour l'ARN H2a-614 diffère sensiblement des deux modèles précédents (**Figure 21**). Seule la tige-boucle du domaine I est clairement identifiée. Les domaines II et III semblent ne former qu'une grande boucle. Le bras IV contient les résidus du HDE dont seulement 8 nucléotides sont appariés. Par ailleurs, l'extrémité 5' du HDE est non appariée et son extrémité 3' présente une bulle refermée par le dernier résidu du HDE.

La fixation de la protéine HBP est bien visible sur le domaine I et II. Dans le domaine I, HBP semble protéger les deux faces du bras. Enfin, aucun changement majeur n'est observé au niveau du HDE suggérant qu'HBP n'induit pas de changement structural à ce niveau.

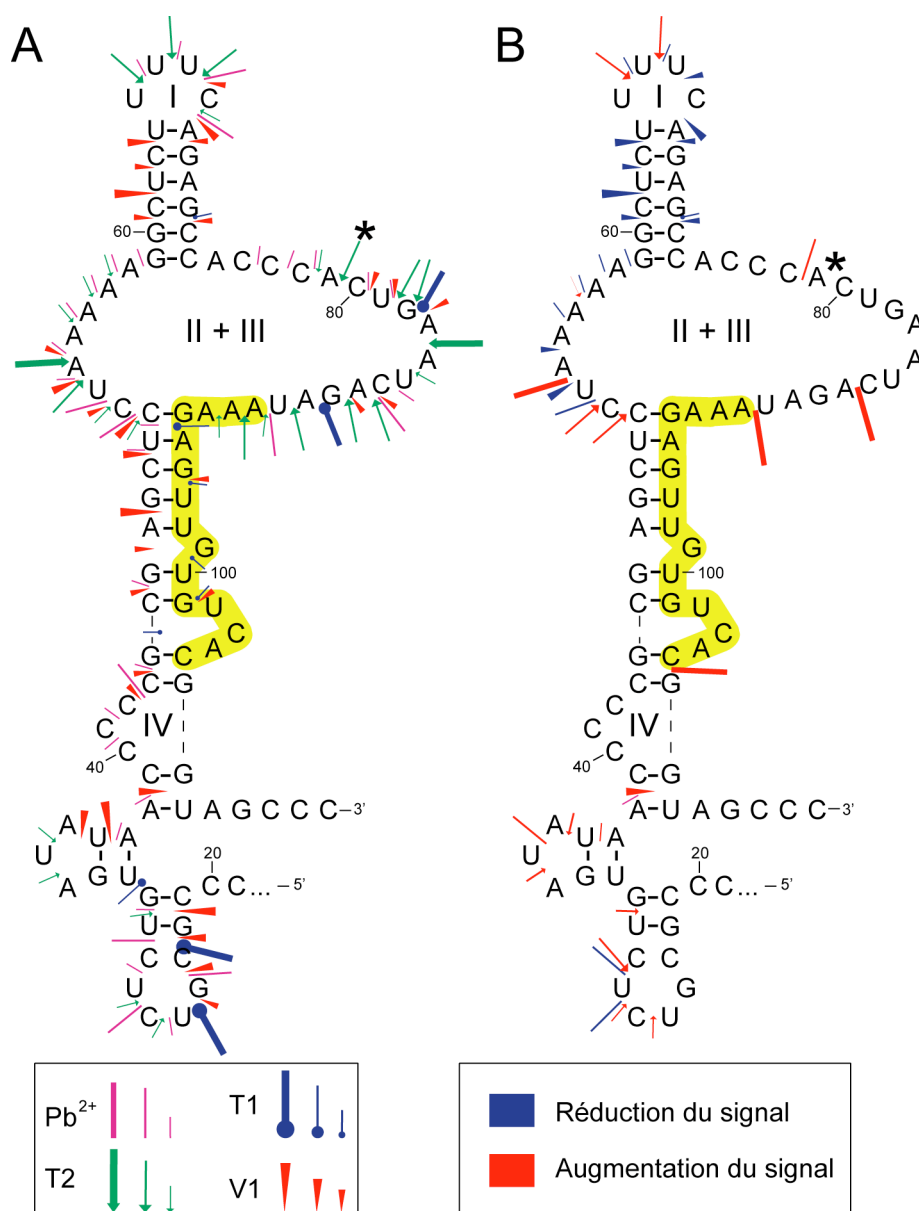


Figure 21 : Modèle de structure secondaire de l'extrémité 3' non traduite de l'ARN pré-messager H2a-614. Le code couleur de présentation est identique à celui utilisé dans la figure précédente.

Autres sondages chimiques sur l'ARN H4-12

Nous avons sondé au DMS et à la CMCT l'extrémité 3' non traduite d'H4-12. Les résultats obtenus étaient en accord avec les résultats précédents. Un remarquable résultat mérite cependant d'être mentionné. Nous avons observé à l'aide du DMS, que la fixation d'HBP sur sa tige-boucle (domaine I) induisait l'apparition d'une modification de la cytosine de la boucle (**Figure 22**). Cette réactivité observée en présence d'HBP suggère que la protéine n'interagit pas avec ce résidu. Elle signifie aussi qu'en l'absence d'HBP, ce résidu établit une interaction tertiaire avec un autre résidu ce qui le protège de l'action du DMS dans l'ARN libre. Cet autre résidu pourrait être le résidu G68 localisé dans la boucle du bras III comme le

montre l'augmentation de réactivité aux ions Pb^{2+} détectée lors de la fixation de HBP. Ainsi, les boucles des domaines I et III pourraient interagir en l'absence de HBP. La protéine HBP pourrait, lors de sa fixation, rompre cette interaction et séparer les deux partenaires.

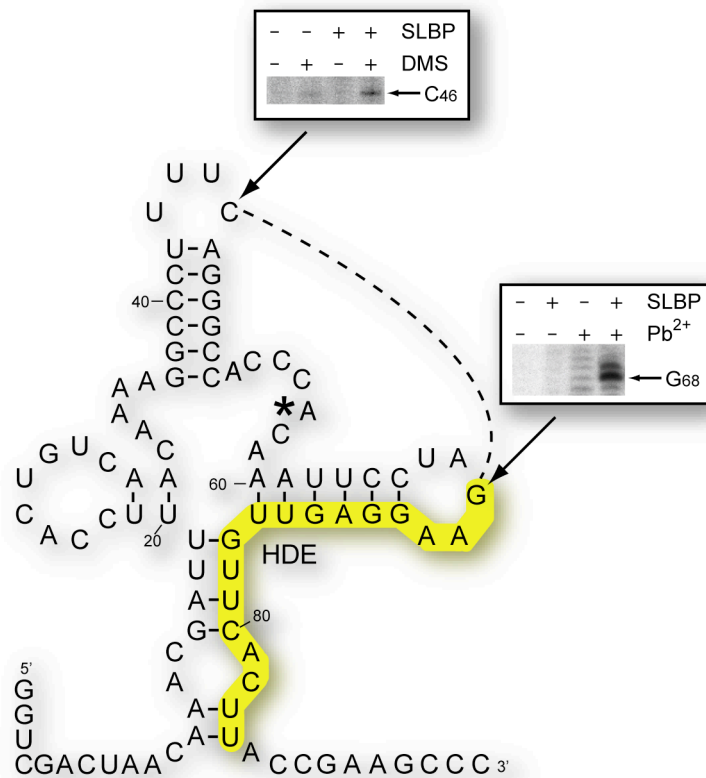


Figure 22 : Effet remarquable de la fixation de la protéine HBP sur la tige-boucle de l'extrémité 3' non traduite du gène H4-12. Le résidu C46 de la boucle devient accessible au DMS en présence de la protéine. Parallèlement, la fixation d'HBP induit l'apparition de coupures aux ions Pb^{2+} dans la région du résidu G68. Ces résultats suggèrent l'existence d'une interaction tertiaire entre les deux positions, ce qui expliquerait leur absence de réactivité dans l'ARN libre.

Cet exemple illustre également l'avantage que présentent les sondes chimiques qui peuvent atteindre, grâce à leur petite taille, des résidus enfouis mais aussi détecter des interactions tertiaires. Nous avons restreint le sondage chimique à l'extrémité H4-12 car il nous a été techniquement impossible de sonder la région du HDE, trop proche de l'extrémité 3' de la molécule.

La fixation d'HBP peut-elle induire un meilleur recrutement du snRNA U7 ?

Nous avons observé que la fixation d’HBP sur les ARNs H4-12 et H1t induisait des changements de réactivité de la séquence HDE de ces ARNs. Dans les 2 cas, une meilleure accessibilité semblait se dessiner, au niveau 5’ dans un cas, et 3’ dans l’autre, ce qui permettait de supposer que l’accessibilité à l’ARN de la snRNP U7 s’en trouverait être

améliorée. Nous avons alors tenté de visualiser cet effet à l'aide d'expériences de « binding » révélées par des gels de retardement. Nous avons également testé l'ARN H2a-614 qui ne subissait pas de changement de conformation lors de la fixation d'HBP.

Ne disposant pas de particules de snRNP U7, nous avons utilisé un transcrit du gène U7 marqué par la radioactivité. Des gammes de concentrations croissantes de transcrits froids des 3 gènes d'histones (extrémités 3' non codantes) ont été dénaturées-renaturées, puis mises en présence de protéine HBP et de transcrit de U7 marqué (préalablement dénaturé-renaturé) (voir Matériel et méthodes § 6.1.2). Les complexes formés ont ensuite été visualisés par migration en conditions natives.

Nous avons ainsi pu montrer que la présence de la protéine HBP améliorait la fixation du transcrit du gène U7 sur le gène H4-12 (**Figure 5 de l'article 4**). En effet, un « super shift » est observé en présence des 3 partenaires HBP+(ARN de H4-12)+(snRNA de U7), que l'on distingue du « simple shift » correspondant au complexe (ARN de H4-12)+(snRNA de U7). Par ailleurs, la présence de la protéine HBP améliore la fixation du snRNA U7, puisqu'à une concentration égale d'ARN de H4-12, davantage de snRNA de U7 libre est entraîné en présence d'HBP.

Dans le cas de l'ARN de H1t nous n'avons pu observer de complexe ternaire. Les changements structuraux observés sur l'ARN de H1t lors de la fixation de HBP, ne sont donc pas suffisants à eux seuls pour permettre la fixation de U7. Il est aussi possible que nous n'ayons pas trouvé les conditions expérimentales adéquates, ou encore que l'utilisation du transcrit de U7 soit insuffisante pour obtenir la fixation. Ce dernier point suggère que les protéines de U7 ne sont pas dépourvues de fonction, ou que le facteur ZFP100 est indispensable dès le stade initial de la fixation.

Enfin, nous n'avons pu obtenir la formation de complexe ternaire avec l'ARN du gène H2a-614. Ce résultat est moins surprenant, compte tenu de l'absence d'effet de la fixation d'HBP sur la structure secondaire de l'ARN. Dans ce cas aussi, la présence d'une snRNP U7 complète ou du facteur ZFP100 pourrait être requise pour obtenir la formation du complexe.

3 DISCUSSION

La maturation de l'extrémité 3' des ARNs pré-messagers d'histone est complexe à étudier du fait du grand nombre de partenaires requis pour la réaction. La coupure endonucléolytique finale est le résultat d'interactions ARN:protéine, protéine-protéine et ARN:ARN entre au moins 4 partenaires différents : l'ARN pré-messenger, HBP, ZFP100, la

snRNP U7 ; le facteur responsable du clivage de l'ARN étant encore inconnu. Il faut rajouter à cela qu'il existe plusieurs ARN pré-messagers différents, qui vont fixer les partenaires avec des affinités différentes, car leurs sites de fixation varient d'un ARN à l'autre.

Par exemple, les tiges-boucles qui fixent HBP présentent des séquences relativement conservées. Elles ne fixent cependant pas HBP de la même manière. Nous avons observé par gel-retardement qu'à quantité égale d'HBP, 96% des molécules du gène H4-12 étaient retardées, contre 55% dans le cas d'H1t et 94% dans le cas d'H2a-614.

Il en va de même lorsque l'on compare les complémentarités de séquence entre les différents HDE et la séquence complémentaire du snRNA U7 (**Figure 16**). Par gel-retardement, nous avons pu constater qu'en l'absence de protéine, il existe une corrélation entre la complémentarité de séquence HDE/snRNA U7 et l'efficacité d'hybridation. Ainsi, l'ARN d'H4-12, qui est le plus complémentaire, forme aisément un duplexe avec le snRNA U7 (**Figure 5 de l'article 4**). L'ARN d'H1t, qui est peu complémentaire, forme peu de duplexes en présence du snRNA de U7. Mais lorsque l'on reproduit ces expériences en remplaçant le transcrit du snRNA U7 par des extraits de fractions enrichies en particules U7 (don des Prof. R. Lührmann et D. Schümperli), on constate une hybridation relativement efficace ce qui suggère que l'intégrité de la snRNP est indispensable et qu'elle peut contrebalancer un défaut d'hybridation par l'apport de contacts supplémentaires.

Un autre facteur non négligeable est la structure de l'ARN lui-même, qui peut être plus ou moins compacté par les repliements secondaires et tertiaires. Les sites de fixation des partenaires peuvent alors ne plus être accessibles. Nous avons pu constater dans les 3 ARNs testés que la séquence d'hybridation au snRNA U7 (HDE) n'était pas accessible dans l'ARN libre, et qu'elle s'appariait spontanément avec une séquence située en amont de la tige-boucle.

Tous ces éléments permettent d'entrevoir la complexité de la réaction de maturation. Les sites de fixation peuvent varier en affinité, mais ils peuvent être compensés par une meilleure affinité d'un autre site pour un autre partenaire. Notre étude a mis en évidence un degré supplémentaire de complexité avec les changements de conformation de l'ARN qui sont induits par la fixation de HBP. Ils ne sont cependant qu'un des maillons qui conduisent à la réaction de maturation. Une fois l'ancrage du snRNA U7 réalisé, le duplexe doit être stabilisé et c'est là qu'interviendrait le facteur ZFP100. Celui-ci stabilise la particule U7 en interagissant avec Lsm11 d'une part et avec la protéine HBP d'autre part. Ainsi, HBP qui est intervenue au stade initial en dépliant la structure intervient à nouveau en fin de réaction pour stabiliser le complexe quaternaire, et peut-être recruter l'endonucléase qui clivera l'ARN.

4 ARTICLE 4

- en préparation -

« Binding of human SLBP on the 3' UTR of histone precursor mRNA induces structural rearrangements that enable U7 snRNP anchoring »

Sophie Jaeger¹, Franck Martin¹, Joëlle Rudinger-Thirion, Richard Giegé and Gilbert Eriani*

Département “Architecture et Réactivité des ARN” UPR 9002 du CNRS,
Université Louis Pasteur
Institut de Biologie Moléculaire et Cellulaire,
15, rue René Descartes, F-67084 Strasbourg cedex, France

¹These authors contributed equally to this work.

*Corresponding author

Fax: (33)388602218

Tel: (33)388417042

E-mail address: G.Eriani@ibmc.u-strasbg.fr

Abbreviations:

SLBP, Stem-Loop Binding Protein; HBP, Hairpin Binding Protein; RBD, RNA Binding Domain; HDE, Histone Downstream Element; EMSA, Electro Mobility Shift Assay

ABSTRACT

In metazoans, cell-cycle-dependent histones are produced from poly(A)-lacking mRNAs. The 3' end of histone mRNAs is formed by an endonucleolytic cleavage of longer precursors between a conserved stem-loop structure and a purine-rich Histone Downstream Element (HDE). The cleavage requires at least two *trans*-acting factors: the Stem Loop Binding Protein (SLBP), which binds to the stem loop; and the U7 snRNP, which anchors to histone pre-mRNA by annealing to HDE. Using chemical (Pb^{2+} , DMS) and enzymatic probes (V1, T1, T2), we determined the secondary structure of the 3' UTR of the mouse histone pre-mRNAs H4-12, H1t and H2a-614. Surprisingly, HDE is embedded in hairpin structures and is therefore not easily accessible for U7 snRNP anchoring. Probing of the 3' UTR in complex with SLBP revealed structural rearrangements leading to an overall opening of the structure especially at the level of HDE. Electro Mobility Shift Assays demonstrated that the SLBP-induced opening of HDE actually facilitates U7 snRNA anchoring on the histone H4-12 pre-mRNA 3' end. These results suggest that prior SLBP binding functions in making the HDE accessible for U7 snRNA anchoring.

INTRODUCTION

Histones are among the most conserved proteins in the eukaryotic kingdom. Their amino-acid sequence is highly conserved as well as the coding sequence suggesting a strong selection pressure on the secondary structure of the corresponding mRNAs (1,2). In contrast, sequences of the 5' and 3' UTRs are more divergent. They are believed to be modern acquisition in order to restrict histone biosynthesis to the S phase of the cell cycle. Different strategies have been developed by eukaryotes for that purpose. Indeed, histones are produced from classical polyadenylated mRNAs in yeast and plants and are produced in metazoans by a sophisticated machinery involving many factors and ribonucleoprotein particles (RNPs) (3,4). In the latter case, histone mRNAs are not polyadenylated but instead end in a highly conserved hairpin structure. The histone mRNAs are synthesized as precursor mRNAs with a 3' extension. The mature histone mRNAs are generated by a single endonucleolytic cleavage occurring between two *cis*-acting elements: a highly conserved hairpin structure upstream of the cleavage site and a so-called Histone Downstream Element (HDE also called spacer element; see figure 1). These two sequences are the scaffold for the assembly of the whole processing machinery. The highly conserved hairpin structure is bound by the Stem-Loop Binding Protein (SLBP, also called HBP for hairpin binding protein) (5,6). Downstream of the cleavage site, HDE anneals to the 5' end of the minor U7 snRNP. Then, the zinc-finger rich protein (ZFP100) bridges the two *cis*-elements by interacting on one hand with the U7-specific protein Lsm11 and on the other hand with SLBP bound to its target hairpin (7-9). Recently, UV-crosslinking studies suggest that CPSF-73, a known component of the cleavage/polyadenylation machinery, is the cleavage factor, which is recruited in a U7-dependent manner (10). The mature histone mRNAs are exported to the cytoplasm for translation. Therefore, the 3' end processing of histone pre-mRNA is the corner stone of histone biosynthesis. Early *in vitro* processing experiments have suggested that the hairpin for SLBP binding is necessary for efficient 3' end processing. However, mutations in the hairpin inhibiting binding of SLBP can be compensated by optimization of the base pairing between the HDE and the U7 snRNP (11). Furthermore, the distance between the hairpin and HDE is critical for efficient processing (12-14). *In vitro* experiments showed that efficient processing of mouse H2a-614 mRNA can be observed in absence of SLBP whereas processing of mouse H1t is fully dependent on the presence of SLBP on its target hairpin because of weak base-pairing with U7 (15). Early *in vivo* experiments on mouse histone mRNA H2a-614 showed that an intact hairpin target for SLBP is required for efficient 3' end processing (16). Recently, this has been validated using the nowadays-available knockout techniques and indeed SLBP is an

essential protein in *C. elegans* (17,18) and in HeLa cells (19,20). Drosophila histone genes are special in that they contain polyadenylation sites downstream to the normal 3' end site. Therefore, SLBP depletion by RNAi in drosophila is not lethal but only causes infertility because the histone mRNAs become polyadenylated. This allows *in vivo* examination of 3' processing reaction, which shows that processing is strongly affected in the absence of SLBP (21). Altogether, these data suggest that SLBP is absolutely required for efficient histone 3' end processing. In spite of all these studies, the reason why SLBP is required for efficient processing remains elusive.

In this work, we investigated the secondary structure of the 3' UTR of histone H4-12 pre-mRNA, which is processed *in vitro* in a SLBP-dependent manner (6,11). Our results show that SLBP induces conformational rearrangements of HDE that enables U7 snRNP anchoring. This is also observed for the histone gene H1t, which displays SLBP-dependent but not for H2a-614 (15).

MATERIALS & METHODS

Construction of templates

Templates for T7 *in vitro* transcription of the 3' UTRs and U7 snRNA have been cloned into the *Sma*I site of pUC19. Sequences of the 3' UTRs of mouse H4-12 (accession number X13235), H2a-614 (accession number AY158924), H1t gene (accession number AY158908) and mouse U7 snRNA (accession number X54165) have been reconstructed by shot-gun oligonucleotides ligation. Correct sequences have been verified by sequencing.

In vitro transcription

In vitro transcription of the 3' UTRs was performed as previously described (22). Transcripts were separated by denaturing 10% PAGE and electro-eluted from gel slices using a Biotrap apparatus (Schleicher and Schuell). The pure transcripts were dephosphorylated using bovine alkaline phosphatase (Fermentas) and 5'-radiolabelled with T4 polynucleotide kinase (New England Biolabs). Radioactive transcripts were further purified by denaturing 10% PAGE and eluted overnight at 4°C in buffer A (0,3 M NaCl, 0,5 mM EDTA, 10 mM Tris-HCl pH 7,5). Before use, transcripts are renatured in water by incubation at 80°C for 2 min, slow cooling to 35°C and kept on ice.

Enzymatic and chemical probing experiments

Enzymatic and chemical structural probing of RNA were conducted by established procedures (23). Probing mixtures (15- μ L) containing radioactive transcripts (50,000 cpm) were incubated for 20 min on ice in buffer B (10 mM Tris-HCl [pH 7,5], 50 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 10 % glycerol) supplemented with 2 mM vanadyl ribonucleoside RNase inhibitor (New England Biolabs), 30 U RNasin (Promega) and 10 pmoles of yeast tRNA. Statistical RNA cleavage was achieved after 10 min incubation at 20°C with $5,9 \cdot 10^{-5}$ U of RNase V1, 1.27 U RNase T1 or 5.6 U of RNase T2. Nucleases V1, T2 and T1 were from KemoTex, Invitrogen and BRL respectively and lead (II) acetate from Merck. Reactions were stopped by rapid cooling on ice and addition of 15 μ L of stop mix A (0,6 M sodium acetate [pH 6], 3 mM EDTA and 0,1 μ g/ μ L total tRNA). Lead probing required a 5-min-incubation at 20°C with 10.7 mM final concentration of a freshly prepared lead-acetate solution. Rapid cooling on ice and addition of EDTA (17 mM final concentration) was used for quenching the reactions. After phenol-chloroform extraction, RNA was ethanol precipitated in the presence of 1 μ g glycogen and washed twice with 80% ethanol. Pellets were dried and dissolved in Formamid Dye. Samples were separated by denaturing 10% PAGE in parallel with RNase T1 and alkaline ladders of the corresponding end-labelled RNA for band assignments. RNase T1

ladders were made as previously described (24). Alkaline ladders were obtained by incubation of the labelled transcript with 1 µg of yeast total tRNA for 10 min at 80°C in a buffer containing 80 mM Na₂CO₃/NaHCO₃ pH 9. As Control, we compared probed RNAs with RNAs treated similarly but without probe.

Enzymatic and chemical foot print experiments

Footprinting reactions were done as described above in the presence of SLBP. Recombinant human His-tagged SLBP was obtained as previously described (25). The amount of SLBP used for each footprint was determined by EMSA (see below) in order to obtain at least 80% of the transcript (50,000 cpm) shifted. Before use, appropriate amounts of SLBP were dialysed against buffer B and then mixed with the radioactive transcripts (50,000 cpm).

Electro Mobility Shift Assay with U7 snRNA transcripts

For EMSAs, 50,000 cpm of 5'-labelled U7 snRNA transcript were incubated in buffer B for 30 min on ice, with increasing concentrations of H4-12 transcript (0.2, 0.5, 1 and 10 µM) in presence or in absence of SLBP. Complexes were separated by native 5% PAGE and visualized by phosphorimaging.

RESULTS

Methodological considerations

A 95-nt transcript encompassing the 3' UTR of histone pre-mRNA H4-12 has been synthesized by T7 run-off transcription. The transcript has been purified by denaturing 10% PAGE after transcription and after 5' labelling. The integrity and the homogeneity of the transcripts have been verified before probing. The structure of this RNA has been probed by RNases T1, T2, V1 and by lead. The sequence of the precursor 3' UTR used here starts downstream of the stop codon as previously identified (26) (Figure 1). For better T7 transcription yield, two G residues have been inserted at the 5' end of the sequence and two C residues are generated at the 3' end after *Sma*I digestion of the template for run-off transcription. RNase T2 and lead cut after unpaired residues and RNase T1 specifically cuts after unpaired G residues. RNase V1 cuts in double-stranded sequences. Spontaneous cuts (background hydrolysis by water and/or traces of contaminating metal cations) were discriminated from probe induced-cuts by control experiments with and without probes.

The H4-12 3' UTR contains 4 structured domains

Based on the probing results, we propose a 2D-structure model of the 3' UTR of H4-12 precursor mRNA (Figure 2). It contains a cloverleaf structure comprising four domains named I, II, III and IV. The presence of the highly conserved hairpin structure (domain I), which is bound by SLBP, is clearly visible. This hairpin is highly stable since no T1 cuts after the five G residues of stem I (G37, G38, G48, G49 and G50) are detected even under denaturing conditions (Figure 2A, lane T1). More strikingly, HDE is embedded in two hairpin structures named domains III and IV. The 5' part of HDE is involved in domain III, a 6-bp-stem and 4-nt-loop. G residues of HDE located in domain III (G71, G72 and G74) are not cleaved by T1 under denaturing conditions indicating high stability of domain III. In contrast G77 which lies in the 3' part of HDE in domain IV is very well cut by T1 under denaturing conditions suggesting that stem IV can be unfolded more easily than stem III. The processing site clearly lies in a single-stranded sequence indicated by presence of lead, T1 and T2 and absence of V1 cuts.

SLBP footprint experiments reveal structural rearrangements in HDE

Using recombinant SLBP produced in baculovirus, we performed footprinting of the SLBP-3' UTR complex (Figure 2C, shown in blue). As expected, SLBP binding prevents V1 cleavages occurring in the 5' part of the stem I and single-strand-specific cleavages in loop I (lead and T2). More surprisingly, SLBP also protects residue 19 to 36 in domain II upstream of its target domain I indicating that SLBP also interacts with domain II. Finally, a weak protection is detected in domain III at position U66. SLBP binding to the 3' UTR not only leads to protection but also induces the appearance of new cleavages or enhancement of existing cuts. This shows that binding of SLBP induces structural rearrangements of the precursor 3' UTR (Figure 2C, shown in red). Strikingly, all these SLBP-enhanced cuts are exclusively located in domains III and IV containing the HDE. Stem III becomes accessible to V1 cuts confirming that this hairpin is still present after SLBP binding, which is in good agreement with the high stability previously observed by non-accessible G residues under denaturing conditions. In addition, loop III is also more accessible to lead. These results demonstrate that domain III becomes more accessible to the probes upon SLBP binding. On the contrary, the conformation of domain IV is significantly changed by the binding of SLBP as shown by strong lead cleavages appearing on both sides of stem IV. Altogether, these data reveal that SLBP binding induces an opening domain III and stem IV, which leads to a more accessible 3' end of HDE. The 3' UTR of the full-length pre-mRNA has also been probed by reverse

transcription using a primer annealing to the 15 last residues at the 3' end of the transcript. Importantly, the structure of the 3' UTR in the context of the full-length molecule is identical to the 3' UTR transcript confirming that the folding of this part of the molecule is not influenced by the 5' UTR and the coding region of the mRNA (data not shown).

The HDE of H1t and H2a-614 3' UTR are also embedded in hairpins

Next, we performed similar experiments on two other histone precursor 3' UTR, H1t and H2a-614. These histone genes have been selected for our study because they show different behaviour with respect to the influence of SLBP as shown by previous *in vitro* processing experiments (15). 3' end processing of histone H1t pre-mRNA fully requires SLBP whereas H2a-614 does not need the presence of SLBP. The probing of H1t 3' UTR pre-mRNA has been performed using the same probes (Figure 3) and the 2D model also contains four domains that have been named I to IV in analogy to H4-12. Domain I is the stem-loop structure bound by SLBP, it is highly stable as shown by absence of reactivity to G residues of the stem under denaturing conditions and by strong stacking at positions 44 to 48 in the AH ladder (Figure 3A). Stem I is accessible to V1 on both sides in contrast to the semi-accessibility observed in the case of H4-12. This suggests that the 3' part of stem I in H4-12 might be masked by another RNA domain, probably domain III that can form a putative kissing loop interaction between loop I and loop III. Domains II of H4-12 and H1t are very similar whereas domain III is completely unpaired in H1t. HDE is almost completely embedded in a structured domain IV. Like for H4-12, SLBP foot-print experiment on the H1t 3' UTR showed protection in domain I and II but only the 5' half of stem I is protected (Figure 3C). This is in good agreement with previous experiments showing that SLBP interacts only with the 5' part of domain I allowing binding of another protein, 3' hExo, on the second part leading to ternary complex formation (27). Binding of SLBP led to appearance of lead cleavages in stem II and the upper part of stem IV and enhancement of accessibility to the probes in the lower part of domain IV. Like for H4-12, these results suggest that SLBP induces conformational rearrangements resulting in opening the HDE-containing domain IV. However, in the case of H1t, it is the 5' part of HDE that opens upon SLBP whereas the 3' part of HDE is opened in the case of H4-12. To conclude this part, one can say that the overall 2D-structure of H1t 3' UTR is more opened than H4-12 but SLBP also induces structural rearrangement in both domains IV. The probing of H2a-614 revealed a rather different organization (Figure 4). Only domain I displays clear V1 cuts on both sides of stem I and single strand-specific cuts in loop I. Domains II and III are completely unpaired

while domain IV, which contains HDE, is composed of eight base pairs interrupted by two bulges. The 5'- and 3'-ends of the HDE are unpaired. Similar to H4-12 and H1t, binding of SLBP induces protections exclusively on domains I and II; however no significant changes are observed in domain III and IV, except four lead cleavages at positions 53, 86, 89 and 105 flanking stem IV (Figure 4C). In contrast to the previous histone 3' UTRs no changes in the HDE are detected upon SLBP binding to the H2a-614 3' UTR. Altogether, these data suggests that the 3' UTR of H2a-614 is less structured than H4-12 and H1t and that binding of SLBP has a minor effect on the structure of HDE.

SLBP enables U7 snRNA anchoring on H4-12 3' UTR

To confirm that SLBP binding enables U7 snRNP anchoring, we performed EMSAs with the 3' UTR of H4-12 pre-mRNA transcript and [³²P]-labelled U7 snRNA transcript. H4-12 3' UTR and U7 snRNA were mixed in the presence or absence of SLBP and complexes separated by native PAGE (Figure 5). Incubation of H4-12 3' UTR and U7 snRNA with SLBP improves the anchoring of U7 snRNA as indicated by the presence of the supershift SLBP/H4-12/U7 which is well separated from the complex H4-12/U7 upon addition of increasing concentration of H4-12. Furthermore, the presence of SLBP stabilizes the SLBP/H4-12/U7 complex since equivalent concentrations of H4-12 can shift more U7 snRNA in the presence of SLBP. The same experiment has been done with H1t and H2a-614 transcripts. Quantification of the free U7 snRNA with increasing concentrations of H4-12 3' UTR with and without SLBP confirmed that U7 snRNA anchoring is facilitated by prior SLBP binding (Figure 5) and allows to propose the following hierarchy of SLBP-dependency for U7 snRNA anchoring: SLBP strongly favours the anchoring of U7 to H4-12, weakly to H1t and not at all to H2a-614. This is well correlated with the overall structure of the respective 3' UTR, since the more structured H4-12 is and the less structured H2a-614 is. Our results are also in good agreement with previous experiments showing that H4-12 and H1t are SLBP-dependent for efficient processing whereas H2a-614 does not require SLBP for *in vitro* processing (6,15).

DISCUSSION

The 3' end processing of histone pre-mRNAs is a complicated process because it requires at least two proteins (SLBP, ZFP100) and the minor U7 snRNP. Numerous RNA-RNA and RNA-protein interactions enable assembly of the processing machinery and the cleavage factor has not been yet identified. In addition, the unique U7 snRNP triggers 3' end processing of many different histone pre-mRNA by annealing to distinct HDEs. In order to decipher the function of each component, we investigated the influence of SLBP on the anchoring of U7 snRNA by chemical and enzymatic probing.

We decided to probe the secondary structure of the full-length 3' UTR of H4-12, H1t and H2a-614. We also checked that the folding of the 3' UTR of full-length H4-12 pre-mRNA is identical to the minimal 3' UTR to be sure that the structure of the 3' UTR is not influenced by sequences upstream of the stop codon. Our results show that in each case the HDE is embedded in hairpin structures. These structures are formed by interaction of HDE with flanking sequences and also with sequences upstream of the SLBP hairpin-binding site. Based on footprint experiments using recombinant SLBP, we propose the following model (Figure 6). The secondary structure of H4-12 pre-mRNA consists in a cloverleaf structure that is closed by a kissing loop interaction between the SLBP binding-site and the 5' end of HDE. NMR structures of the hairpin revealed that the C residue, which is not recognized by SLBP, is protruding outside of the loop and therefore should be largely accessible to probes (28,29). However, chemical probing experiments (DMS, lead) confirmed that the C residue of the SLBP binding site loop and the first residue G of HDE are only accessible upon SLBP binding (data not shown). This suggests that C residue of the loop is involved in a long-distance interaction that is destroyed by SLBP binding. The footprint of SLBP revealed that it interacts mainly with the 5' part of the molecule and with 5' half of the hairpin. This is in good agreement with previous experiments showing that 3' hExo can bind to the already formed SLBP-hairpin complex and form a ternary complex (27). This implies that 3' hExo can start trimming before or even without the release of SLBP. Furthermore, ZFP100 has been selected by three-hybrid screen using SLBP-hairpin complex and biochemical experiments showed that ZFP100 does not bind to SLBP on its own (27). The C-residue of the loop is a good candidate for a specific RNA-protein interaction with ZFP100 since it is specifically accessible only in the SLBP-hairpin complex. Other structural rearrangements are induced by SLBP binding; they lead to opening of the 3' part of HDE. We showed that these rearrangements actually facilitate the anchoring of U7 snRNA. Processing experiments in mammalian and drosophila extracts revealed that SLBP are not interchangeable and strongly

inhibits processing in the heterologous extracts indicating a co-evolution of the hairpin and HDE (30). Co-evolution of the 3' UTR can also be explained by the fact that the secondary structure of the 3' UTR is important for efficient processing. It has also been suggested that the strength of annealing between U7 and HDE is directly correlated with the dependency of SLBP for processing (11,31,32). However, an exception for that is the mouse H3-53, which has an HDE that can make strong base pairing with U7 snRNA. Despite strong complementarities between H3-53 and U7 snRNA, *in vitro* processing of this pre-mRNA is strictly dependent on SLBP (11). These experiments have been done with *in vitro* transcripts encompassing a part of the 3' UTR of the pre-mRNA, usually starting from a few nucleotides upstream the histone hairpin till the end of the HDE. In our case, the folding of the whole 3' UTR has been studied and interestingly, we demonstrate that, in the 3 histone mRNAs studied in this work, HDE anneals to residues upstream of the histone hairpin. Therefore, the structure of the whole 3' UTR cannot be compared with a minimal processing substrate encompassing the histone hairpin and the HDE. This might explain some discrepancy with previous histone 3' processing experiments. Another implication of our model is that 3' end processing of histone pre-mRNA is an ordered process. SLBP binds first, enabling anchoring of U7 snRNP and finally the whole complex is locked by ZFP100, which bridges SLBP-hairpin complex to U7 snRNP. This model is in good agreement with previous reports showing that, among the distinct processing factors, SLBP is the sole cell-cycle regulated factor (33,34). The whole processing machinery is present constitutively in the cell but cannot promote processing because of lack of SLBP. The fact that HDEs of histone precursor mRNAs are “hidden” in secondary structures from U7 snRNA anchoring might be looked at as a safety mechanism to prevent processing and therefore histone biosynthesis in non-S phase especially at the G1-S transition phase when histone transcription is increasing and SLBP is not yet available (33). SLBP is the cornerstone of histone expression, its binding on histone pre-mRNA “gives a go” to the whole process. To conclude, SLBP is necessary for 3' end processing of histone pre-mRNA. Therefore, the strict cell-cycle regulation of histone expression can only be explained by cell cycle regulation of the sole SLBP.

ACKNOWLEDGMENTS

We are grateful to Agnès Clénet for skilful technical assistance, to Sharief Barends for critical reading of the manuscript, Pascale Romby, Hervé Moine and Jean-Christophe Paillart for helpful discussions. This work was supported by grants from xxx. S.J. was recipient of a “Ministère de l'Éducation Nationale, de la Recherche et de la Technologie” fellowship.

REFERENCES

1. DeBry, R.W. and Marzluff, W.F. (1994) Selection on silent sites in the rodent H3 histone gene family. *Genetics*, **138**, 191-202.
2. Huynen, M., Konings, D. and Hogeweg, P. (1992) Equal G and C contents in histone genes indicate selection pressures on mRNA secondary structure. *J Mol Evol*, **34**, 280-291.
3. Marzluff, W.F. (2005) Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Curr Opin Cell Biol*, **17**, 274-280.
4. Jaeger, S., Barends, S., Giegé, R., Eriani, G. and Martin, F. (2005) Expression of metazoan replication-dependent histone genes. *Biochimie*, **87**, 827-834.
5. Wang, Z.F., Whitfield, M.L., Ingledue, T.C., 3rd, Dominski, Z. and Marzluff, W.F. (1996) The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes Dev*, **10**, 3028-3040.
6. Martin, F., Schaller, A., Eglite, S., Schümperli, D. and Müller, B. (1997) The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. *EMBO J*, **16**, 769-778.
7. Pillai, R.S., Will, C.L., Lührmann, R., Schümperli, D. and Müller, B. (2001) Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein. *EMBO J*, **20**, 5470-5479.
8. Dominski, Z., Erkmann, J.A., Yang, X., Sanchez, R. and Marzluff, W.F. (2002) A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3'-end processing. *Genes Dev*, **16**, 58-71.
9. Azzouz, T., Gruber, A. and Schümperli, D. (2005) U7 snRNP-specific Lsm11 protein: dual binding contacts with the 100 kDa zinc finger processing factor (ZFP100) and a ZFP100-independent function in histone RNA 3' end processing. *Nucleic Acids Res.*, **33**, 2106-2117.
10. Dominski, Z., Yang, X.C. and Marzluff, W.F. (2005) The Polyadenylation Factor CPSF-73 Is Involved in Histone-Pre-mRNA Processing. *Cell*, **123**, 37-48.
11. Streit, A., Koning, T.W., Soldati, D., Melin, L. and Schümperli, D. (1993) Variable effects of the conserved RNA hairpin element upon 3' end processing of histone pre-mRNA in vitro. *Nucleic Acids Res*, **21**, 1569-1575.
12. Cho, D.C., Scharl, E.C. and Steitz, J.A. (1995) Decreasing the distance between the two conserved sequence elements of histone pre-messenger RNA interferes with 3' processing in vitro. *RNA*, **1**, 905-914.

13. Scharl, E.C. and Steitz, J.A. (1994) The site of 3' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. *EMBO J*, **13**, 2432-2440.
14. Scharl, E.C. and Steitz, J.A. (1996) Length suppression in histone messenger RNA 3'-end maturation: processing defects of insertion mutant premessenger RNAs can be compensated by insertions into the U7 small nuclear RNA. *Proc Natl Acad Sci U S A*, **93**, 14659-14664.
15. Dominski, Z., Zheng, L.X., Sanchez, R. and Marzluff, W.F. (1999) Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. *Mol Cell Biol*, **19**, 3561-3570.
16. Pandey, N.B., Williams, A.S., Sun, J.H., Brown, V.D., Bond, U. and Marzluff, W.F. (1994) Point mutations in the stem-loop at the 3' end of mouse histone mRNA reduce expression by reducing the efficiency of 3' end formation. *Mol Cell Biol*, **14**, 1709-1720.
17. Pettitt, J., Crombie, C., Schümperli, D. and Müller, B. (2002) The *Caenorhabditis elegans* histone hairpin-binding protein is required for core histone gene expression and is essential for embryonic and postembryonic cell division. *J Cell Sci* 857-866.
18. Kodama, Y., Rothman, J.H., Sugimoto, A. and Yamamoto, M. (2002) The stem-loop binding protein CDL-1 is required for chromosome condensation, progression of cell death and morphogenesis in *Caenorhabditis elegans*. *Development*, **129**, 187-196.
19. Zhao, X., McKillop-Smith, S. and Müller, B. (2004) The human histone gene expression regulator HBP/SLBP is required for histone and DNA synthesis, cell cycle progression and cell proliferation in mitotic cells. *J Cell Sci.*, **1**, 6043-6051.
20. Wagner, E.J., Berkow, A. and Marzluff, W.F. (2005) Expression of an RNAi-resistant SLBP restores proper S-phase progression. *Biochem Soc Trans*, **33**, 471-473.
21. Marzluff, W.F., Godfrey, A., Dominski, Z., Kupsco, J.M. and Duronio, R.J. (2005) Mutants in the histone mRNA 3' processing system in *Drosophila*. *Communication at the RNA 2005 meeting*, **Abstract N° 158**.
22. Perret, V., Garcia, A., Grosjean, H., Ebel, J.P., Florentz, C. and Giegé, R. (1990) Relaxation of a transfer RNA specificity by removal of modified nucleotides. *Nature*, **344**, 787-789.
23. Giegé, R., Helm, M. and Florentz, C. (2001) Classical and novel chemical tools for RNA structure probing. In *RNA* (Söll, D., Nishimura, S. and Moore, P.B., eds), **Pergamon/Elsevier, Amsterdam**, 79-89.
24. Peattie, D. and Gilbert, W. (1980) Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci U S A.*, **77**, 4679-4682.

25. Martin, F., Michel, F., Zenklusen, D., Müller, B. and Schümperli, D. (2000) Positive and negative mutant selection in the human histone hairpin-binding protein using the yeast three-hybrid system. *Nucleic Acids Res*, **28**, 1594-1603.
26. Meier, V.S., Bohni, R. and Schümperli, D. (1989) Nucleotide sequence of two mouse histone H4 genes. *Nucleic Acids Res*, **17**, 795.
27. Dominski, Z., Yang, X.C., Kaygun, H., Dadlez, M. and Marzluff, W.F. (2003) A 3' exonuclease that specifically interacts with the 3' end of histone mRNA. *Mol Cell*, **12**, 295-305.
28. Zanier, K., Luyten, I., Crombie, C., Müller, B., Schümperli, D., Linge, J., Nilges, M. and Sattler, M. (2002) Structure of the histone mRNA hairpin required for cell cycle regulation of histone gene expression. *RNA*, **8**, 29-46.
29. DeJong, E.S., Marzluff, W.F. and Nikonowicz, E.P. (2002) NMR structure and dynamics of the RNA-binding site for the histone mRNA stem-loop binding protein. *RNA*, **8**, 83-96.
30. Dominski, Z., Yang, X.C., Raska, C.S., Santiago, C., Borchers, C.H., Duronio, R.J. and Marzluff, W.F. (2002) 3' end processing of *Drosophila melanogaster* histone pre-mRNAs: requirement for phosphorylated *Drosophila* stem-loop binding protein and coevolution of the histone pre-mRNA processing system. *Mol Cell Biol*, **22**, 6648-6660.
31. Spycher, C., Streit, A., Stefanovic, B., Albrecht, D., Koning, T.H. and Schümperli, D. (1994) 3' end processing of mouse histone pre-mRNA: evidence for additional base-pairing between U7 snRNA and pre-mRNA. *Nucleic Acids Res*, **22**, 4023-4030.
32. Dominski, Z., Erkmann, J.A., Greenland, J.A. and Marzluff, W.F. (2001) Mutations in the RNA binding domain of stem-loop binding protein define separable requirements for RNA binding and for histone pre-mRNA processing. *Mol Cell Biol*, **21**, 2008-2017.
33. Whitfield, M.L., Zheng, L.X., Baldwin, A., Ohta, T., Hurt, M.M. and Marzluff, W.F. (2000) Stem-loop binding protein, the protein that binds the 3' end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms [In Process Citation]. *Mol Cell Biol*, **20**, 4188-4198.
34. Zheng, L., Dominski, Z., Yang, X.C., Elms, P., Raska, C.S., Borchers, C.H. and Marzluff, W.F. (2003) Phosphorylation of stem-loop binding protein (SLBP) on two threonines triggers degradation of SLBP, the sole cell cycle-regulated factor required for regulation of histone mRNA processing, at the end of S phase. *Mol Cell Biol*, **23**, 1590-1601.

Figure 1

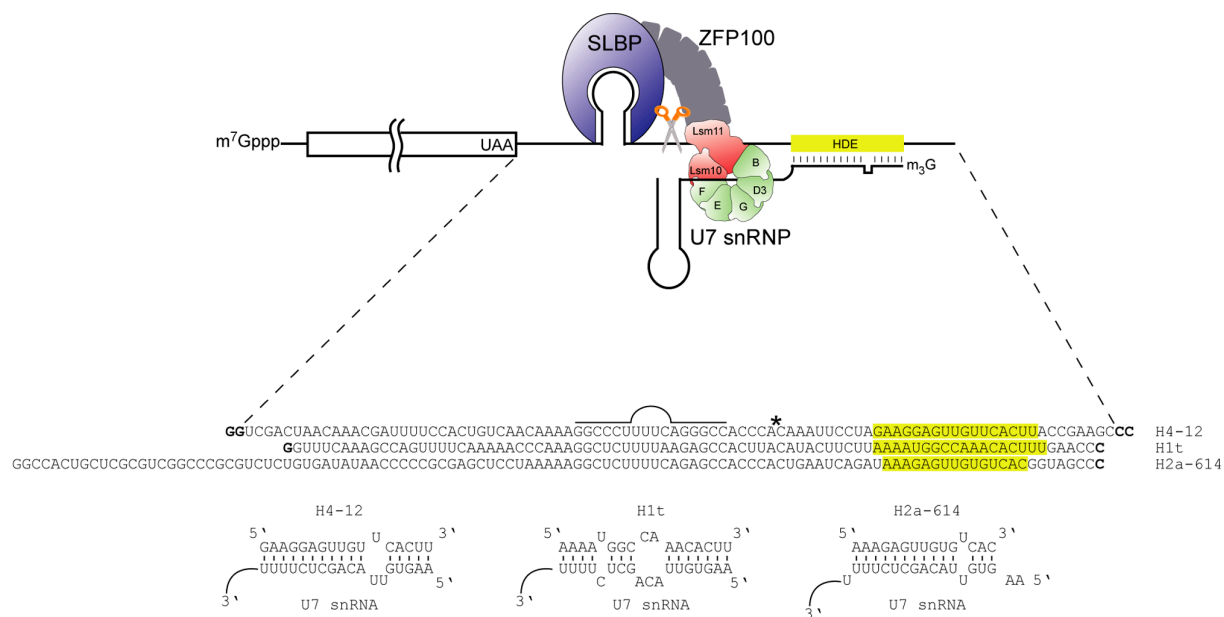


Figure 2

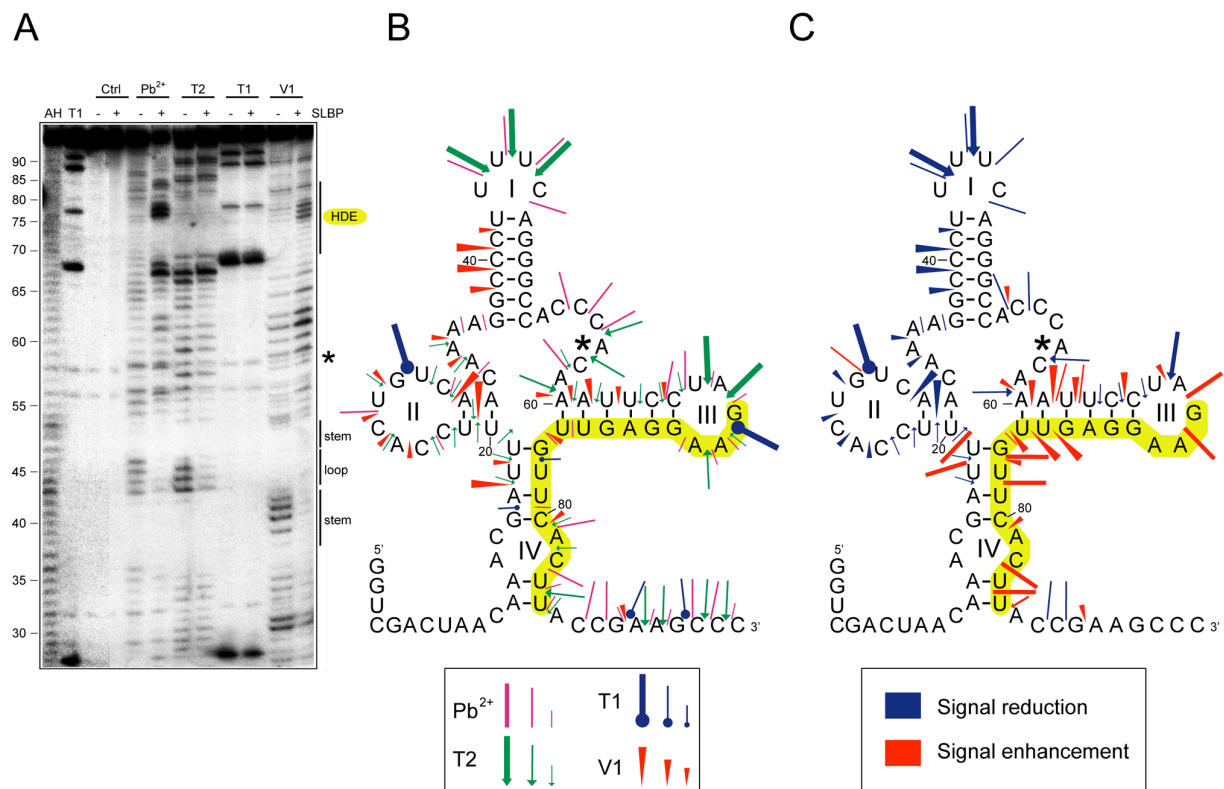


Figure 3

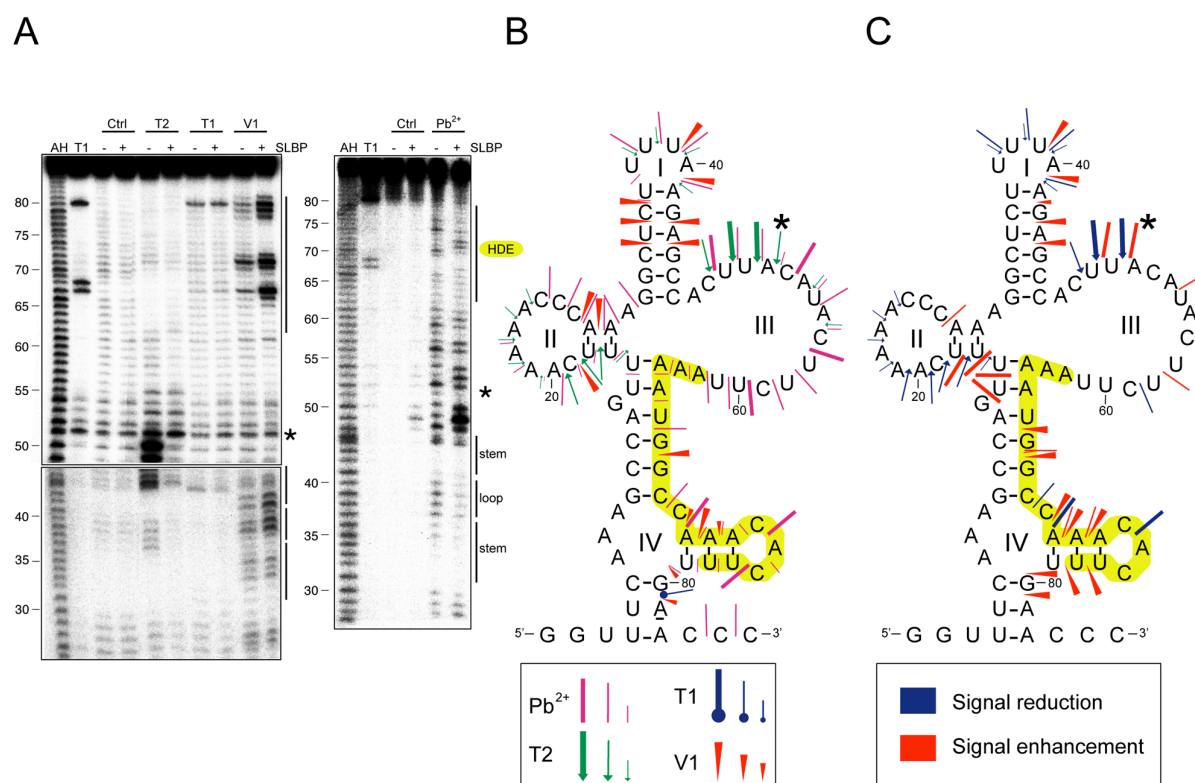


Figure 4

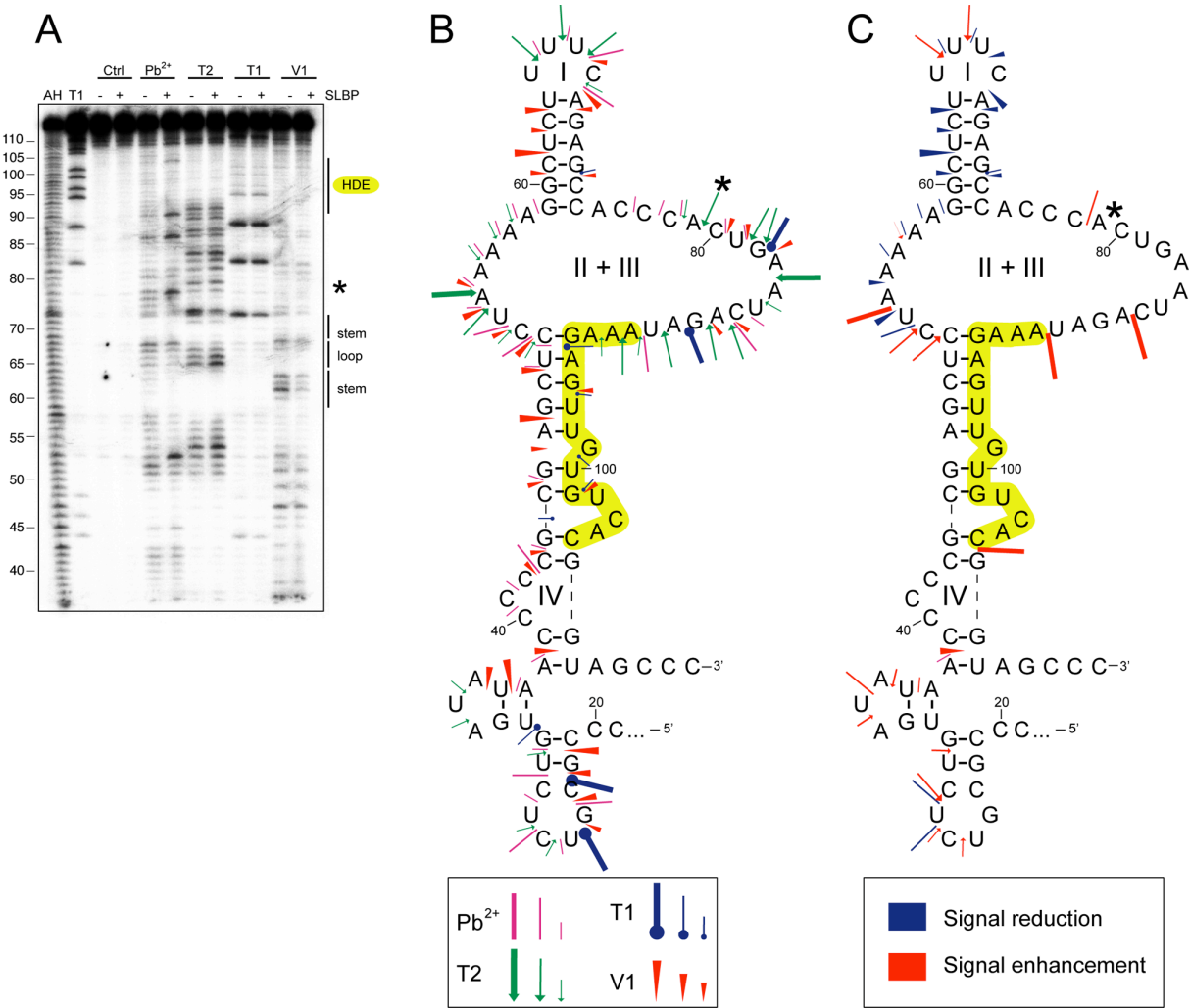


Figure 5

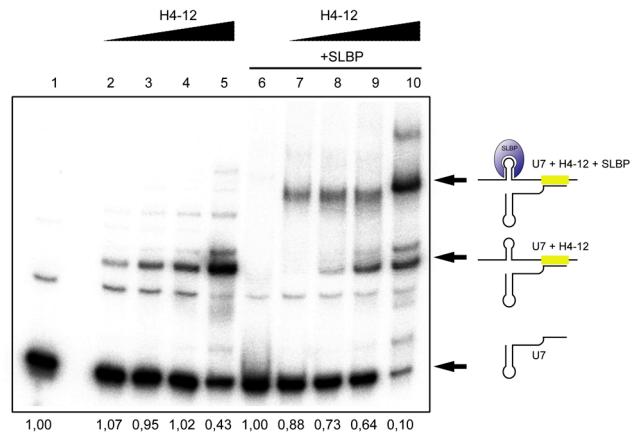


Figure 6

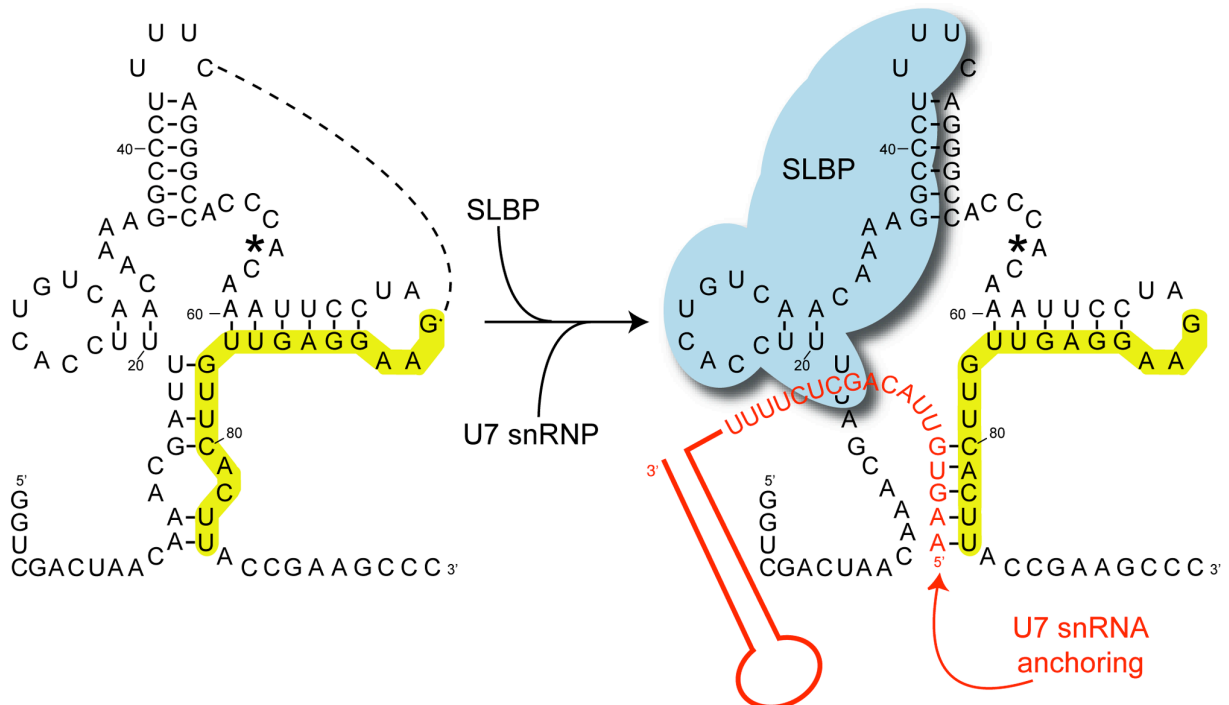


FIGURE LEGENDS

Figure 1

Schematic representation of the 3'-end processing of replication-dependent histone pre-mRNAs. Cleavage of the RNA occurs between a highly conserved stem-loop structure and the HDE (indicated by scissors). *Trans*-acting factors are SLBP, which binds to the hairpin structure and U7 snRNP, which anchors the pre-mRNA by annealing to the spacer element. U7 snRNP contains two U7-specific proteins named Lsm10 and Lsm11. ZFP100 bridges U7 and SLBP by interacting with SLBP and Lsm11. The sequences of the 3' UTRs used in this work are indicated under the cartoon. Nucleotides shown in bold are added for optimized transcription yields. Position of the stem loop, the cleavage site (*) and the HDE (in yellow) are indicated. Base-pairing of H4-12, H1t and H2a-614 with mouse U7 snRNA is shown at the bottom of the figure.

Figure 2

Establishment of an experiment-based folding model of the 3' UTR of **histone H4-12 pre-mRNA** in free state and effect of SLBP binding. (A) Chemical and enzymatic structural probing. Probing was done in the presence (+) or absence (-) of recombinant SLBP. AH and T1 are ladders of the RNA under denaturing condition; Ctrl is the control without any probe. Probes were lead (Pb^{2+}), RNase T2, RNase T1 and RNase V1. Domains of the stem-loop structure, the processing site (*) and the spacer element are indicated on the right of the gel. (B) Folding of the 3' UTR of histone H4-12 pre-mRNA with probing data. The colour code for the probes is indicated in the figure. Three intensities of cuts/modifications for each probe are shown. The spacer element is highlighted in yellow. (C) Changes in the probing profile upon SLBP binding shown on the RNA folding. Protection by SLBP is shown in blue (signal reduction) and cut/modification appearance and enhancement (signal enhancement) are shown in red; symbols for each probe are the same as in (B). The HDE (in yellow) and the processing site (*) are highlighted in each panel.

Figure 3

Establishment of an experiment-based folding model of the 3' UTR of histone **H1t pre-mRNA** in free state and effect of SLBP binding. (A) Chemical and enzymatic probing data. (B) Folding of the RNA. (C) Changes in the probing profile upon SLBP binding shown on the RNA folding. Legend is the same as for Figure 2.

Figure 4

Establishment of an experiment-based folding model of the 3' UTR of histone **H2a-614 pre-mRNA** in free state and effect of SLBP binding. (A) Chemical and enzymatic probing data. (B) Folding of the RNA. (C) Changes in the probing profile upon SLBP binding shown on the RNA folding. Legend is the same as for Figure 2.

Figure 5

Electro Mobility Shift Assay using U7 snRNA and H4-12 pre-mRNA transcripts. The 5'-radiolabelled U7 snRNA was incubated with increasing concentrations of H4-12 pre-mRNA (0.2, 0.5, 1 and 10 μ M) in the absence (lane 2 to 5) or presence (lane 6 to 10) of recombinant SLBP. Cartoons on the right depict the distinct complexes and the free U7 snRNA. Relative quantification of the free U7 snRNA is reported under the gel.

Figure 6

Model of the assembly of SLBP and U7 snRNA on histone H4-12 3' UTR pre-mRNA. The model shows an ordered process where SLBP binds prior to U7 snRNP and induces structural rearrangements leading in the opening of the 3' part of the spacer element (in yellow). The secondary structure of histone H4-12 pre-mRNA 3' UTR is shown before and after SLBP binding. A putative long-range interaction is depicted with a dashed line. The U7 snRNA 5' end (in red) can anneal and gradually open the hairpin encompassing the 5' part of the HDE.

CHAPITRE 3

Étude structurale et fonctionnelle de l'ARNm H4-12 **Mise en évidence d'un mécanisme de traduction original**

A conduit à la publication

Article 5

en préparation

« Two IRES elements in the coding region of histone H4 mRNA
enable highly efficient translation »

Sharief Barends¹, Sophie Jaeger¹, Roman I. Koning, Henk K. Koerten,
Gilbert Eriani et Franck Martin

¹ co-premiers auteurs

1 INTRODUCTION

L'étude décrite dans ce chapitre a trouvé son inspiration dans les travaux que nous avons précédemment effectués sur la protéine HBP. Après avoir étudié la réaction de maturation de l'extrémité 3', nous avons réfléchi au rôle que pourrait jouer HBP dans la traduction des ARNm d'histones. Un certain nombre de travaux avaient déjà conclu, peut-être hâtivement, à un rôle d'HBP dans la traduction. Il est vrai que par sa localisation sur sa tige boucle à l'extrémité 3' de l'ARNm, HBP est un candidat idéal pour mimer la PABP fixée sur la queue de polyA des ARNm cellulaires ordinaires. Cependant, ces études avaient été sérieusement critiquées (voir § 3.3.4.2 de l'introduction) en raison des faibles effets mesurés, et de certaines contradictions. Le principal reproche que l'on pouvait adresser à ces études était qu'elles avaient été menées avec des gènes rapporteurs fusionnés avec l'extrémité du gène d'histone, et qu'elles ne tenaient nullement compte d'un éventuel rôle de la phase codante. Or, une observation plus poussée de l'ARNm murin d'H4-12 révèle dans sa phase codante certaines caractéristiques inhabituelles qui suggèrent qu'une certaine pression de sélection s'est exercée au cours de l'évolution sur la structure de cette région de l'ARNm d'histone.

Particularités des ARNm d'histones

Les ARNm d'histones font partie des plus petits ARNm cellulaires, non seulement parce qu'ils codent pour des petites protéines mais surtout parce qu'ils possèdent de très courtes extrémités 5' et 3' non traduites. Ainsi, la séquence 5' non traduite de l'ARNm d'H4-12 ne comporte que 9 nucléotides (Meier *et al.*, 1989). Cette taille extrêmement limitée semble *a priori* incompatible avec une initiation de la traduction efficace, pourtant ces ARNm sont traduits de façon très efficace lors de la phase S du cycle cellulaire.

Les histones sont aussi les protéines eucaryotiques les plus conservées. Les histones H3 et H4 sont les plus conservées, et la protéine H4 est totalement invariante. L'histone H4-v.1, qui est le seul variant connu de l'histone H4, présente quelques rares variations de la séquence nucléotidique mais aucune variation au niveau de la séquence peptidique. Cependant, la conservation de séquence dans les ARNm d'histones ne concerne que la région codante. On note une grande divergence de séquence dans les régions 5' et 3' non traduites (Piontkivska *et al.*, 2002).

Caractéristiques des séquences suggérant une conservation structurale

Des analyses bioinformatiques portant sur les séquences nucléotidiques des gènes d'histones ont conduit à des observations singulières. En effet, en plus d'être très conservées, les régions codantes des ARNm d'histones sont anormalement riches en résidus Guanosine (G) et Cytidine (C) qui y sont retrouvés à des niveaux équivalents (suggérant un grand nombre d'appariements G:C). Les séquences flanquantes ne possèdent pas de telles caractéristiques. De plus, quelles que soient les espèces, les variations des rapports G/C trouvés au niveau des 2 premières positions des codons sont toujours contrebalancées par les rapports G/C à la troisième position des codons de façon à ce que le contenu global en G et C demeure constant (Huynen *et al.*, 1992). Ceci est particulièrement vrai pour les histones H4. Enfin, il a été décrit que les phases codantes des histones H3 de rongeurs avaient un fort biais pour les nucléotides G et C en troisième position du codon (DeBry & Marzluff, 1994). Or, si la séquence peptidique doit être conservée, cette position reste celle qui est la plus susceptible d'être modifiée.

L'ensemble de ces éléments constitue des arguments forts qui suggèrent que la conservation de séquence nucléotidique est le reflet d'une pression de sélection sur la structure secondaire de l'ARNm au cours de l'évolution. Cependant, avant nos travaux, cette hypothèse n'était étayée par aucune donnée expérimentale, tant au niveau structural que fonctionnel. Au laboratoire, nous avons noté que les ARNm d'histones présentent des régions 5' non traduites de tailles très réduites, qui semblent incompatibles avec une initiation de la traduction classique. Il est apparu dans nos esprits qu'un lien pouvait exister entre l'énigme de l'initiation de la traduction et l'existence d'une structure secondaire particulière. Nous avons alors entrepris d'étudier la traduction de l'ARNm de l'histone H4-12 de souris et d'établir sa structure secondaire par sondage en solution. Bien entendu, nous avons pris soin d'étudier les ARNm complets tels qu'ils sont retrouvés dans la cellule.

Ces deux axes d'étude, structure et fonction, ont été développés en parallèle, les résultats de l'un permettant d'expliquer le second et inversement. Ce travail a impliqué plusieurs personnes au sein du groupe. Je me suis plus personnellement impliquée dans l'établissement des structures secondaires des ARNm natifs ou mutés. Un premier modèle de structure secondaire a été obtenu, il a été progressivement affiné au fil des expériences de sondage et de mutagenèse pour arriver au modèle que nous proposons dans l'article 5. Pour des raisons de place dans ce mémoire, je ne détaillerai pas l'intégralité des expériences, ni les différents modèles intermédiaires de structures secondaires précédant l'établissement du modèle final (à ce jour).

2 RÉSULTATS

2.1 L'ARNm de l'histone H4 présente un mécanisme de traduction original

Nous avons entrepris de travailler sur l'ARNm codant pour l'histone H4-12. En effet, c'est l'un des ARNm d'histones dont la séquence nucléotidique est la plus conservée, et dont l'extrémité 3' non traduite est la plus courte (seulement 9 nucléotides).

Lors d'un travail préliminaire, nous avons amplifié le gène H4-12 par PCR à partir d'ADN génomique. Le produit de PCR (375 pb) a été cloné dans le plasmide pUC19 pour donner le plasmide pUC19-H4-12. Par la suite, ce gène H4-12 a été systématiquement utilisé comme matrice pour toutes les amplifications PCR destinées à la préparation des matrices de transcription. Les réactions de transcription par la T7 ARN polymérase ont été effectuées sur ces produits de PCR. La coiffe ($m^7G_{ppp}G$) a été incorporée lors de la réaction de transcription. Les transcrits ont été purifiés par électrophorèse sur gel préparatif en conditions dénaturantes et extraits du gel par électroélution. Nous les avons utilisés à une concentration de 250 nM dans des tests de traduction *in vitro* composés d'extraits de germes de blé ou de lysats de réticulocytes de lapin. La détection des protéines traduites est rendue possible par l'incorporation de [^{35}S]-méthionine. La phase codante du gène H4-12 contient deux codons méthionine. Nous avons déterminé expérimentalement que la première Met était éliminée à hauteur de 80% de manière post-traductionnelle. Les protéines synthétisées comportent donc en grande majorité une seule méthionine radioactive. Connaissant cette valeur, nous avons pu quantifier la synthèse d'histone dans les tests de traduction. Les premières expériences de traduction réalisées sur les molécules d'ARNm d'histone H4-12 natives ont montré que la traduction s'effectuait très efficacement, avec une efficacité comparable, voir-même supérieure, à celle d'une protéine virale témoin.

Pour déterminer si l'initiation de la traduction était dépendante de la présence de la coiffe en 5', nous avons effectué des expériences de compétition en présence de concentrations croissantes de coiffe libre (**Figure 2 de l'article 5**). L'inhibition totale de la traduction a été observée à partir d'une concentration de 60 μM en analogue de la coiffe dans le milieu de traduction. Cette faible valeur démontre que la traduction est dépendante de la présence de la coiffe au niveau de l'extrémité 5' de l'ARNm.

Nous avons ensuite vérifié si l'extrémité 3' non traduite pouvait présenter un élément déterminant pour la traduction. Certains travaux antérieurs semblaient montrer que la protéine HBP était un stimulant de la traduction qui pouvait jouer un rôle semblable à celui de la PABP (voir § 3.3.4.1 de l'introduction). Nous avons alors examiné l'influence de la protéine HBP dans nos tests de traduction en utilisant un ARNm complet, ce qui diffère des études antérieures entreprises sur des fusions avec des gènes rapporteurs. Nos résultats ont montré que la protéine HBP n'a aucune influence sur l'efficacité de traduction lorsqu'elle est ajoutée en *trans* dans le milieu réactionnel. Mais ce sont surtout les délétions des extrémités de l'ARNm, et en particulier de la structure en tige-boucle fixant HBP, qui ont confirmé l'absence de rôle essentiel de la protéine HBP. En effet, pour déterminer si les extrémités 5' et 3' non codantes étaient vraiment nécessaires, elles ont été progressivement délétées. Les délétions ont été obtenues par transcription de produits de PCR raccourcis, obtenus à partir d'oligonucléotides appropriés. Ces produits de transcription coiffés ont été testés en traduction. Les résultats ont montré que les extrémités 3' et 5' non codantes ne sont pas requises pour la traduction (**Figure 1 de l'article 5**).

Ces expériences de délétion montrent que la région codante de l'ARNm est à elle seule suffisante pour être traduite et permettre au ribosome d'initier la traduction sur l'AUG initiateur. Ces résultats suggèrent fortement que le ribosome pourrait se placer directement sur le codon d'initiation sans qu'il n'y ait l'étape normale de « scanning » par la petite sous-unité du ribosome.

Ces premiers résultats démontrent que le codon d'initiation est reconnu sans qu'un « scanning » de la région 5' non traduite ne soit nécessaire. Afin de nous assurer qu'il s'agissait bien d'un phénomène propre au gène de l'histone H4-12, nous avons délété de manière identique l'extrémité 5' non traduite de l'ARNm de la protéine GFP. Dans ce cas, la traduction de l'ARNm tronqué a conduit à la synthèse d'un peptide « p7 » résultant de l'initiation non pas sur le bon AUG initiateur, mais sur l'AUG suivant situé sur une deuxième phase codante (ce dernier étant atteint par « scanning » conventionnel) (**figure S1 de l'article 5**). Cette expérience contrôle montre bien que la phase codante de l'histone H4 présente la capacité unique de recruter le ribosome au niveau du codon initiateur.

Pour compléter ces résultats nous avons construit des mutants de l'ARNm où le codon initiateur était muté en codon Thr (mutant M₁T) et recréé plusieurs AUG initiateurs en amont, ou en aval afin d'étudier la plasticité de ce système d'initiation de la traduction (**Figure 3 de l'article 5**). Le mutant M₁T abolit complètement la traduction, par contre l'introduction de codons AUG aux positions 5 et 7 restaurent la traduction alors qu'au-delà de la position 7,

ainsi qu'en amont de la position 1 (codon -4), l'efficacité de traduction diminue considérablement. Ceci renforce l'hypothèse qu'il n'y a pas de « scanning » du ribosome. Il existe cependant une certaine flexibilité dans le processus d'initiation de la traduction puisque la traduction est encore très efficace lorsque l'AUG initiateur est déplacé de 5 codons, soit 15 nucléotides.

Tous ces résultats réunis suggèrent que la phase codante de l'ARNm de l'histone H4 se comporterait comme un site d'entrée interne du ribosome (IRES) qui positionnerait le ribosome au niveau du codon d'initiation. Ces résultats soulèvent un certain nombre de questions relatives au mécanisme d'initiation que le ribosome utilise. Pour y répondre, nous avons entrepris le sondage de la structure en solution de la molécule d'ARN en vue d'en déterminer la structure secondaire, puis la recherche des régions fonctionnelles de la molécule impliquées dans l'initiation de la traduction.

2.2 Étude de la structure secondaire de l'ARNm de l'histone H4

Nous avons sondé l'ARNm de l'histone H4-12 à l'aide de sondes spécifiques de l'ARN simple brin comme les sondes chimiques DMS, CMCT et ions Pb^{2+} ou enzymatiques (RNases T1, T2) et de l'ARN double brin comme la RNase V1. La taille relativement grande de l'ARNm (375 nucléotides) nous a conduit à utiliser une méthode indirecte de détection, c'est-à-dire que les coupures ou les modifications par les sondes ont été appliquées au transcrit froid, la révélation s'effectuant de manière indirecte par transcription inverse en utilisant 7 amorces distribuées le long de la séquence. Ces expériences constituent, à notre connaissance, le premier exemple de sondage en solution d'un ARNm cellulaire dans sa totalité. À l'aide des données obtenues et du logiciel M-fold (Zuker, 2003) nous avons établi un modèle de structure secondaire de l'ARNm qui a ensuite été rectifié manuellement.

Le modèle présenté dans la **figure 23** apparaît très stable et compact avec 62 paires de base G:C sur les 77 que compte le modèle. Les extrémités 5' et 3' non traduites de l'ARNm sont libres, ce qui est en accord avec les expériences de délétion qui ont souligné leur caractère non essentiel.

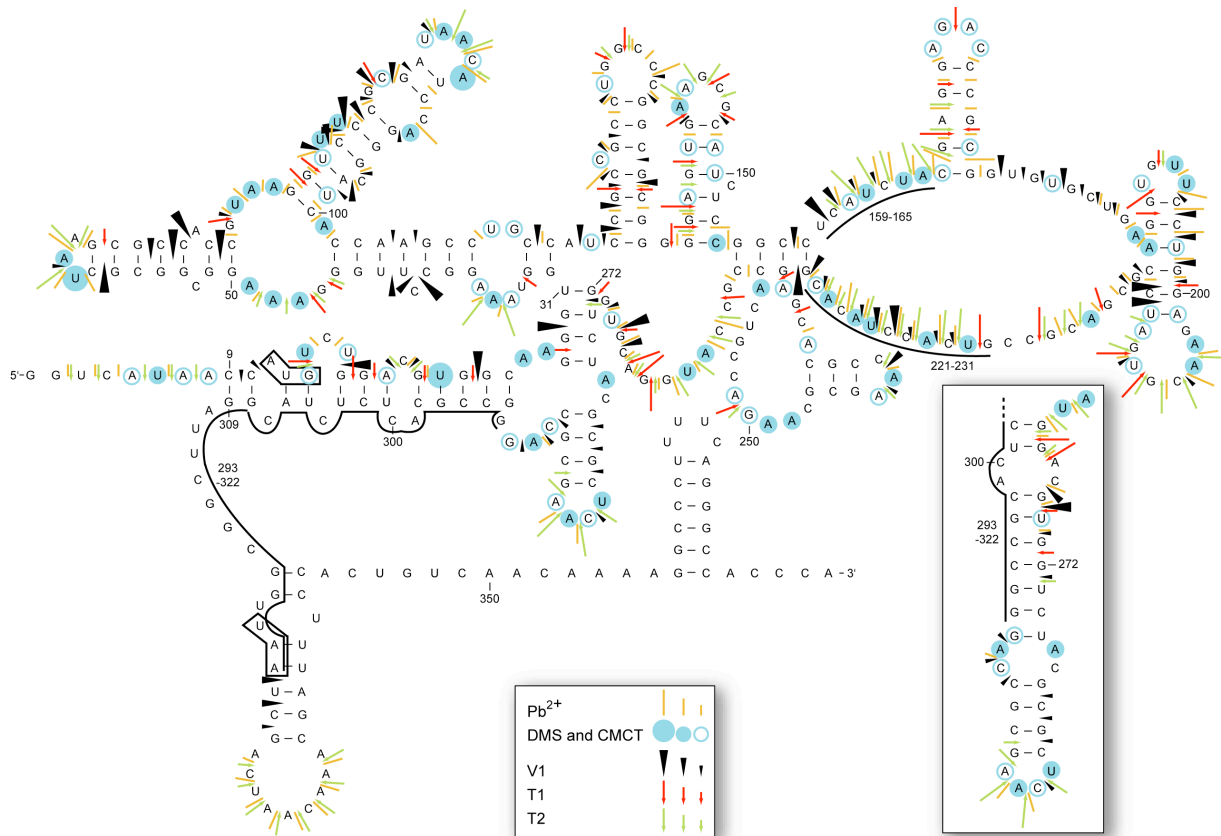


Figure 23 : Modèle de structure secondaire proposé d'après le sondage en solution de l'ARNm H4-12 par les sondes chimiques et enzymatiques. Les séquences 159-165 et 221-231 sont soulignées car elles contiennent des signaux contradictoires indiquant à la fois la présence d'appariements et une propriété simple-brin. La séquence 293-322 est soulignée car il est impossible d'y détecter des modifications ou coupures avec aucune sonde. La séquence 263-302 peut également se replier différemment et former une structure alternative (encart de droite) en accord avec les réactivités et qui libère la séquence 9-31 de son appariement.

Le modèle comporte plusieurs zones structurées très bien définies par les sondes, c'est notamment le cas de la structure en tige-boucle formée par les nucléotides 50 à 68 où la boucle est de type « tri-loop ». Cette structure est stable comme le montre le phénomène de compression observé sur gel après séparation électrophorétique des produits de synthèse de la transcriptase inverse. Une autre région présumée fortement structurée se situe entre les nucléotides 293 et 322 (soulignée dans la **figure 23**) ; il est impossible d'y détecter la moindre modification. Le modèle comporte également d'autres régions où la structure semble plus instable ou vraisemblablement équilibrée entre différentes conformations. On y détecte des signaux contradictoires, c'est-à-dire des réactivités simultanées pour les sondes simple brin et double brin. Ces régions comprennent les nucléotides 159-165 et 221-231 (soulignées dans la **figure 23**) ; elles sont probablement impliquées dans un équilibre dynamique entre différentes conformations. Une autre région comportant les nucléotides 9-31 est quant à elle réactive aux sondes simples et doubles brins bien qu'elle soit en mesure de former un appariement avec les

nucléotides 272-275 et 294-309 de l'extrémité 3' de la phase codante (dans l'encart de la **figure 23**). L'existence de cet appariement a été confirmée plus tard lorsque nous avons étudié la traduction d'ARNm mutés. L'effet de la mutation de la position C25 en G, qui réduit de dix fois l'efficacité de traduction de l'ARNm H4-12, peut être compensé par une mutation du résidu G294 en C qui restaurera l'interaction entre les positions 25 et 294 (**Figure 4 de l'article 5**). Ces résultats confirment bien qu'il existe une interaction entre les résidus des deux positions et donc un appariement de la région 9-31. Ainsi, cette région serait soit libre comme l'indiquent certaines données en solution ou appariée comme le confirme la traduction d'ARNs mutés. La région 9-31 pourrait donc adopter deux conformations alternatives.

Afin de vérifier l'existence de ces conformations alternatives, nous avons déposé sur gel de polyacrylamide en conditions non dénaturantes (sans urée), des transcrits de l'ARNm marqués au niveau de leur extrémité 5'. La migration en conditions natives à 4°C a permis de séparer deux espèces moléculaires (**Figure 24**). L'extraction de chacune de ces espèces suivie d'une nouvelle migration en conditions natives a montré que leur conformation restait stable en conditions natives. Cependant, la dénaturation-renaturation de chacune des espèces purifiées entraîne la réapparition des deux conformations en équilibre. Ces deux conformations alternatives pourraient correspondre aux repliements alternatifs suggérés plus haut pour la séquence 9-31 ou les régions 159-165 et 221-230.

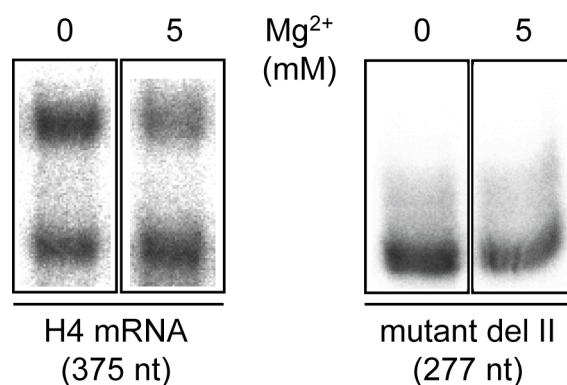


Figure 24 : Gels natifs présentant les profils de migration des ARNm des gènes H4-12 sauvage ou délétant « del II ». Les 2 ARNs ont été renaturés puis incubés en absence ou présence d'ions Mg²⁺ (pistes 0 et 5 mM). Deux conformations sont visibles dans le cas de l'ARNm sauvage, alors qu'une seule conformation est visible dans le cas du mutant « del II ».

Nous avons alors sondé la structure en solution de ces 2 conformations de l'ARNm sauvage en respectant les conditions natives. Les résultats obtenus avec les RNases T1, T2, V1 et les ions Pb²⁺ ont conduit à des résultats quasi-similaires, y compris dans les extrémités 5' et 3' non traduites. La structure secondaire reste donc identique, à moins qu'il n'y ait eu un

retour à l'équilibre entre les 2 conformations lors du sondage. Les seules différences observées l'ont été dans les profils de dégradation des 2 espèces moléculaires marquées au ^{32}P .

Les expériences de traduction ont montré l'importance de la coiffe pour la traduction. Compte tenu du coût élevé de la coiffe, les expériences de sondage en solution l'ont été sur des ARNm non coiffés. Il nous a paru essentiel de sonder la structure en solution de l'ARNm coiffé afin de déterminer si la présence de la coiffe sur la molécule traduite allait modifier sa structure secondaire. Les résultats obtenus n'ont pas mis en évidence des changements structuraux dans les ARNm coiffés. Notre modèle correspond donc bien à la structure d'une molécule efficacement traduite par le ribosome.

2.3 Identification des éléments structuraux de l'ARNm impliqués dans la traduction

Afin de localiser les régions de l'ARNm impliquées dans l'initiation de la traduction, nous avons réalisé des tests de traduction en présence de petits ARNs complémentaires. Le principe de cette expérience repose sur le fait que le ribosome en phase d'élongation est capable de déplacer un duplexe ARN:ARN alors qu'il ne pourra pas démarrer sur un tel duplexe situé sur le site d'initiation de la traduction (Takayar *et al.*, 2005). La phase codante de l'ARNm a été parcourue par 15 petits ARNs complémentaires de 20 nucléotides chacun. L'expérience baptisée « primer walk » a consisté à hybrider individuellement un excès molaire de 10 fois des 15 petits ARNs (**Figure 4 A de l'article 5**). Les duplexes ont été purifiés puis testés en traduction dans des extraits de germe de blé et de réticulocytes de lapin. Les résultats nous ont permis d'identifier trois ARNs agissant sur la traduction dans les deux types d'extraits. L'ARN 1 est le contrôle positif car il s'hybride au niveau du codon initiateur. Il a l'effet le plus drastique sur la traduction. Les deux autres ARNs, à savoir l'ARN 7 et l'ARN 9, réduisent l'efficacité de traduction de l'ARN sauvage de plus de 50%. Ces résultats indiquent que les régions complémentaires à ces ARNs sont impliquées dans le recrutement du ribosome. Trois autres ARNs (ARN 4, 11 et 13) inhibent la traduction mais uniquement dans les extraits de germe de blé (**Figure 4 A**). Ces différences pourraient s'expliquer par les dissemblances pouvant exister entre les systèmes de traduction des plantes et des métazoaires. Des contrôles ont été réalisés afin de vérifier que chaque ARN hybridait bien la séquence qui lui était complémentaire. Nous avons dans un premier temps réalisé des expériences de

digestion à la RNase H, mais en utilisant des petits oligonucléotides ADN de séquence identique à celle de nos petits ARNs (§ 7.2.6 de Matériel et méthodes). Les résultats obtenus ont révélé dans l'ensemble une très bonne spécificité d'hybridation ; seule une hybridation croisée mineure a été observée avec l'oligonucléotide 13 qui hybride dans la région complémentaire à l'ARN 1. Nous avons aussi séquencé l'ARNm par transcription inverse avec des oligonucléotides radioactifs correspondant aux petits ARN hybridés (§ 7.3.2.3 de Matériel et méthodes). Les séquences ont confirmé les bons résultats des tests à la RNase H. Seule, la séquence obtenue avec l'oligonucléotide 1 se double d'une séquence correspondant à la région d'hybridation de l'ARN 13. Enfin, une preuve supplémentaire de la bonne hybridation des petits ARNs a été apportée lors du sondage en solution des duplexes inactifs (voir plus loin). La présence du petit ARN hybridé à l'ARNm a pu être mise en évidence par une pause de la transcriptase inverse utilisée lors de la détection indirecte des coupures, pause engendrée par une collision de l'enzyme avec le duplexe ARNm:ARN antisens.

Ces expériences de « primer walk » sur l'ARNm nous ont permis d'identifier 2 régions impliquées dans l'initiation de la traduction. À ce stade de l'étude, les causes de l'inhibition de la traduction demeuraient au rang d'hypothèse. L'inhibition pouvait résulter d'un défaut de recrutement du ribosome par masquage de l'ARN, il pouvait aussi agir par effet indirect en détruisant un élément de structure de l'ARNm.

2.4 Sondage de l'ARNm en présence de petits ARNs anti-sens

Pour déterminer si l'hybridation des petits ARNs induisait des changements structuraux dans l'ARNm, nous avons sondé aux ions Pb^{2+} la structure en solution des duplexes ARN:ARN dont la traduction est inhibée. Nous avons ainsi évalué l'effet de la fixation des ARNs 4, 7, 9 et 13 qui diminuaient l'efficacité de traduction, ainsi que l'effet des ARN 1 et ARN 5 utilisés comme contrôles positifs et négatifs respectivement (**Figure 4A de l'article 5**). Les seuls changements que nous ayons observés l'ont été avec l'ARN 9 ; ils se limitaient à l'apparition de nouvelles pauses de la transcriptase inverse à proximité de l'AUG initiateur au niveau des nucléotides 24, 25, 28, 29, 30 et 31 (**Figure 4C de l'article 5**). Ces pauses suggèrent l'existence d'un lien entre ces nucléotides et la séquence d'hybridation de l'ARN 9, qu'il soit direct ou indirect. Elles traduisent l'apparition d'une difficulté pour l'enzyme à déplier la structure locale, difficulté dont la provenance reste à découvrir.

Ces nouvelles expériences de sondage en présence des ARNs inhibiteurs n'ont pas permis de révéler des changements structuraux majeurs pouvant expliquer de manière simple

leur effet sur la traduction. Elles ont tout de même permis d'identifier des signaux de pause qui sont à mettre en relation avec la fixation d'un ARN inhibiteur sur l'ARNm. L'étude des nucléotides correspondant à ces signaux est décrite dans le paragraphe suivant.

2.5 Identification de nucléotides impliqués dans la traduction par mutagenèse dirigée

Le sondage en solution des molécules d'ARNm bloquées en traduction a mis en évidence des pauses de transcriptase inverse au niveau des nucléotides 24, 25, 28, 29, 30 et 31. Ces nucléotides se situent précisément dans une des régions de l'ARNm pour laquelle nous avons proposé un repliement alternatif (**Figure 4B de l'article 5**). Nous avons effectué 9 mutations dans cette région ou dans la région en vis-à-vis afin d'étudier l'effet de mutations de compensation. Les mutations ont été introduites dans le produit de PCR destiné à la transcription en utilisant des amorces mutées.

L'effet le plus remarquable est obtenu pour la substitution C25G qui est silencieuse au niveau protéique (Gly₅Gly) ; l'efficacité de la traduction de l'ARNm est diminuée de 10 fois pour ce mutant (**Figure 4C de l'article 5**). Notre modèle de structure secondaire prédit un appariement de C25 avec G294. En conséquence, nous avons additionné la mutation G294C à C25G afin de restaurer l'interaction Watson-Crick. Le double mutant présente une nette restauration de l'efficacité de traduction, en accord avec notre modèle qui prédisait une interaction entre ces deux résidus. Les autres mutations agissant fortement sur l'efficacité de la traduction touchent le nucléotide 31 (mutation silencieuse Gly₇Gly) et les nucléotides 29 à 31 (mutation Gly₇Pro).

Cependant toutes les mutations affectant l'appariement de la séquence 9-31 avec la région 272-309 ne présentent pas un effet drastique sur la traduction. Ceci pourrait s'expliquer par la possibilité de former des paires de bases alternatives par le jeu du glissement de certains appariements. Cette malléabilité n'est cependant pas envisageable dans le cas du mutant C25G (Gly₅Gly)(**Figure 4C** encart inférieur).

Ces expériences de mutagenèse confirment l'existence de l'appariement entre les séquences 9-31 et 272-309. Une certaine tolérance dans l'organisation des appariements semble possible exception faite du nucléotide 25 qui doit rester apparié pour une traduction efficace.

2.6 Sondage de la structure des ARNm mutés

Nous avons identifié par mutagenèse des résidus impliqués dans la traduction. En conséquence, nous avons sondé la structure en solution des mutants présentant les plus forts effets sur la traduction (C25G, U31G et GGU29-31CCG). Aucun changement de la structure secondaire n'a pu être détecté malgré l'utilisation des RNases habituelles et des ions Pb^{2+} . Les seules différences avec la structure native résidaient en de nouvelles pauses de la transcriptase inverse, avec notamment une pause remarquable en position 142 pour les trois mutants.

Nous avons également procédé à un sondage aux ions Pb^{2+} du mutant Met₁Thr où la traduction est abolie. Aucune différence dans les signaux n'est apparue lors de la comparaison effectuée avec l'ARNm sauvage. La structure secondaire de la molécule ne semble donc pas être affectée par cette mutation.

2.7 Recrutement du ribosome

Nous avons tenté de déterminer l'endroit où se positionne le ribosome lorsqu'il est recruté par la phase codante, en procédant à des expériences de « toe print ». Le test « toe print » permet d'identifier les régions de l'ARNm où se localise le ribosome par un arrêt de l'élongation de la transcriptase inverse engendré par la collision avec le ribosome.

Les molécules d'ARNm sont dans un premier temps incubées dans un extrait de traduction de réticulocyte de lapin supplémenté en un inhibiteur dont la présence empêchera le démarrage de la traduction, de manière à figer le ribosome en phase d'initiation sur l'ARNm. Les complexes sont ensuite soumis à une étape de transcription inverse en utilisant deux amorces radioactives, l'une s'hybridant à la fin de la molécule, l'autre au milieu de manière à couvrir l'ensemble de l'ARNm, excepté les 45 nucléotides de la fin. Les régions où se localise le ribosome sont identifiées par un arrêt de l'élongation de la transcriptase inverse.

Les résultats obtenus avec l'ARNm H4 sauvage sont inhabituels par la présence de multiples signaux d'arrêt de l'enzyme dispersés dans la phase codante. Quatre signaux forts se localisent au niveau des nucléotides 85-87, 248-250, 294 et 307-310, le nucléotide 1 étant le résidu A de l'AUG initiateur (**Figure 5 de l'article 5**). La plupart des signaux forts, à l'exception des positions 85-87, se situent à proximité de l'extrémité 3' de la séquence codante dans la région de l'ARNm où existe la possibilité d'appariement avec les nucléotides 9-31. Des signaux plus faibles ont été détectés au niveau des nucléotides 19, 70-72, 76, 78 et

96. La présence de toutes ces pauses dans la phase codante démontre bien que le ribosome est recruté par celle-ci.

La présence de ces multiples signaux d'arrêt pourrait s'expliquer par le fait que le ribosome emprunte plusieurs éléments structuraux distincts de l'ARNm pour se fixer, à moins que plusieurs ribosomes ne se fixent à l'ARNm. Dans les deux cas, la fixation du ou des ribosomes s'effectue sur l'ARNm actif ou inactif, car le mutant M₁T présente essentiellement les mêmes signaux. Les arrêts observés vers la fin de la phase codante pourraient signifier que cette région est particulièrement importante pour la fixation des ribosomes.

2.8 Recrutement de deux ribosomes, cartographie du deuxième site de recrutement

Afin de déterminer si l'ARN recrutait effectivement plus d'un ribosome, nous avons procédé à une observation du complexe formé entre l'ARNm H4-12 et le ribosome par microscopie électronique en coloration négative. La purification des complexes a été effectuée par ligation de 600 pmoles d'ARNm coiffées en 5' à un oligonucléotide complémentaire biotinylé en 3' (**Figure S2 de l'article 5**). Cette ligation s'est réalisée par l'intermédiaire d'un oligonucléotide stabilisateur permettant, par appariement aux deux protagonistes, de maintenir à proximité l'extrémité 3' OH de l'ARNm et l'extrémité 5' phosphate de l'oligonucléotide biotinylé. Après ligation, l'ARNm a été immobilisé par interaction de la biotine avec la streptavidine greffée sur des billes magnétiques. Les oligonucléotides libres ainsi que l'oligonucléotide utilisé pour faciliter la ligation ont alors été éliminés par lavage. À l'issue de cette étape, environ 60 à 80 pmoles d'ARN ont été immobilisées sur les billes. Les complexes de traduction ont été formés par incubation des billes dans des extraits de réticulocytes de lapin contenant environ 8 pmoles de ribosomes. Après deux lavages, les complexes de traduction ont été élués par digestion de l'oligonucléotide biotinylé à la Dnase et utilisés dans une analyse par microscopie électronique.

Les résultats obtenus indiquent qu'une majorité d'ARNm recrutent un ribosome en formant des particules appelées monosomes. Une proportion non négligeable de molécules recrute deux ribosomes pour former des disomes (**Figure 6A de l'article 5**). Des résultats identiques sont obtenus avec l'ARNm mutant M₁T, indiquant que l'ARNm H4 est capable de recruter deux ribosomes avant que la traduction ne débute. Ainsi, le deuxième ribosome n'est

pas un ribosome en phase d'élongation, mais bien un ribosome recruté en phase d'initiation de la traduction.

Nous avons alors voulu localiser ces deux ribosomes sur la molécule. Rappelons que les expériences de « toe print » ont révélé de fortes pauses dans la seconde moitié de la phase codante. Nous avons alors transcrit des ARNm tronqués à partir de matrices de PCR délétées progressivement de l'extrémité 3' terminale. Les ARN transcrits ont ensuite été testés en traduction. Des résultats étonnants ont été obtenus pour le délétant II qui s'achève au niveau du résidu 277 et qui ne contient plus les séquences 294-309 capables d'interagir avec l'extrémité 5' de la molécule (**Figure 4B de l'article 5**). L'efficacité de traduction de ce mutant est réduite de moitié par rapport à l'ARNm sauvage (**Figure 6C de l'article 5**) à la fois dans les extraits de réticulocytes de lapin et de germes de blé. Ce résultat confirme à nouveau l'importance de la région 3' terminale de la phase codante.

La délétion de séquence pouvant être très préjudiciable au repliement structural de l'ARNm, nous avons sondé par les RNases habituelles et par les ions Pb^{2+} , la structure en solution du délétant II. Le profil des coupures obtenues est semblable à celui de l'ARNm entier. L'effet de la délétion sur la traduction n'est donc pas le résultat de remaniements structuraux majeurs de l'ARNm restant, mais bien le résultat de la perte d'une structure ou d'une séquence dont le rôle serait de recruter le ribosome. De plus, la migration de ce délétant en gel natif démontre la présence d'une seule conformation (**Figure 24**).

Les résultats obtenus avec le délétant II suggèrent que la délétion de l'extrémité 3' de la phase codante de l'ARNm H4-12 empêche le recrutement d'un des deux ribosomes, l'efficacité de traduction se trouvant diminuée de moitié. Ceci suggère aussi que le recrutement du 2^{ème} ribosome est pour moitié responsable de l'efficacité de traduction exceptionnelle de l'ARNm H4. Par microscopie électronique des complexes d'initiation réalisés sur le délétant II, nous avons pu montrer que la proportion de disomes est fortement diminuée alors que celle des monosomes augmente. La présence de quelques disomes résiduels, malgré le fait qu'un des sites de fixation des ribosomes ait disparu, pourrait s'expliquer par le démarrage de la traduction en dépit de la présence de l'inhibiteur (cycloheximide). En d'autres termes, en s'engageant dans la traduction de l'ARNm, le premier ribosome aurait libéré le site de fixation de l'AUG qui serait devenu disponible pour un second ribosome (le tout étant finalement bloqué par la cycloheximide).

3 DISCUSSION

Les histones réplication-dépendantes sont exprimées abondamment et précisément au cours de la phase S du cycle cellulaire. Ce double contrôle peut s'expliquer par certaines caractéristiques propres aux gènes d'histones : maturation particulière de l'extrémité 3' non traduite, absence d'introns dans les gènes et transcription abondante des ARNm.

Partant de l'observation que l'extrémité 5' non traduite était anormalement courte dans l'ARNm de l'histone H4, nous avons montré que l'initiation de la traduction ne nécessitait aucune extrémité non codante, et qu'au contraire, la séquence nucléotidique correspondant à la phase codante contenait toutes les informations requises pour une traduction très efficace. Il semble donc que la cellule ait développé un mode de traduction original et particulièrement efficace pour fournir en peu de temps les grandes quantités d'histones requises pour l'empaquetage de l'ADN néosynthétisé.

Nous nous sommes appliqués à décrypter ce mécanisme de traduction atypique par une étude structure-fonction combinant le sondage en solution de l'ARNm et les tests de traduction *in vitro* d'ARNm natifs ou mutés. La combinaison de ces approches a conduit à un modèle de structure secondaire de l'ARNm et permis de proposer un modèle de fonctionnement permettant d'expliquer le mécanisme de traduction atypique et très efficace qui a été observé (**Figure 7 de l'article 5**). Nous proposons que la phase codante de l'histone H4-12 agirait en tant que site d'entrée interne de ribosome ou « IRES » (**Figure 25**) en fixant pas moins de 2 ribosomes. Un premier ribosome serait directement recruté au niveau du codon initiateur, un second au niveau de l'extrémité 3' de la phase codante. Le site d'entrée du premier ribosome serait crucial car les expériences de déplacement de l'AUG initiateur ont montré qu'il n'y avait pas de « scanning » de l'ARNm. Le second site d'entrée du ribosome serait localisé du côté 3' de la phase codante comme le démontrent les expériences de « toe print », les expériences de délétion et de microscopie électronique.

Il est bien évidemment tentant d'établir un lien entre la présence de l'« IRES » pour 2 ribosomes, la dynamique de réarrangement des structures alternatives que le sondage en solution a révélé et la très grande efficacité de traduction. Manifestement, la structure secondaire de l'ARNm joue un rôle capital dans ce mécanisme d'initiation hors du commun ce qui expliquerait la très forte pression sélective qui s'est exercée sur sa conservation au cours de l'évolution.

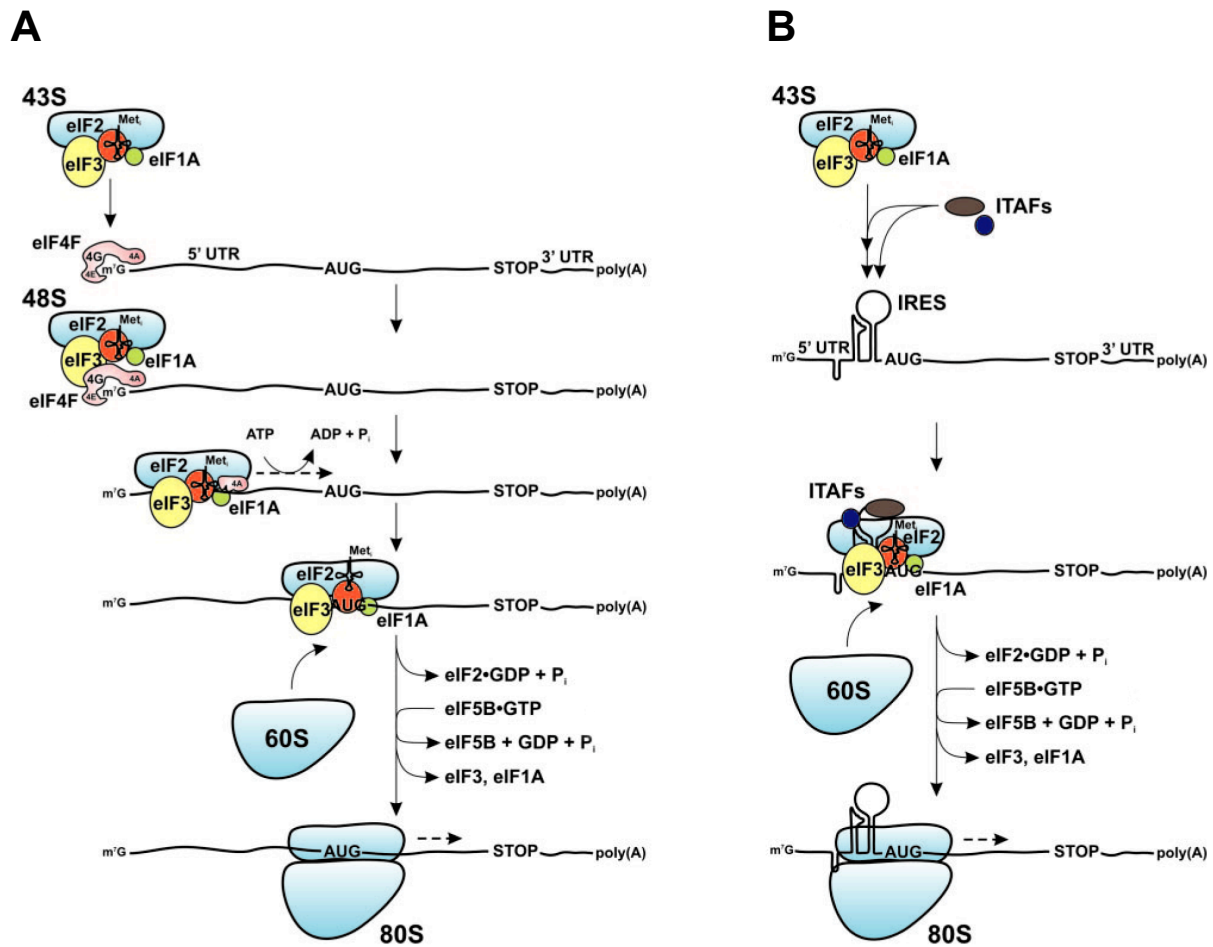


Figure 25 : Comparaison du système d'initiation de la traduction classique et du système « IRES » (Internal Ribosome Entry Site). **(A)** Le système classique fait intervenir de nombreux facteurs protéiques dont la fonction est d'assister le ribosome à se positionner puis à s'assembler correctement au niveau du codon d'initiation. Il nécessite la présence de la coiffe en 5' et un glissement ou « scanning » du ribosome. **(B)** Dans le cas de l'initiation par entrée interne du ribosome, la sous-unité 40S du ribosome est directement recrutée sur la région de l'AUG initiateur de manière indépendante de la coiffe et sans nécessiter le « scanning ». Ce recrutement s'effectue au niveau d'une structure spécifique de la région 5' non traduite de l'ARNm appelée « IRES ». Pour certains virus comme celui de l'hépatite C (HCV), ou de la paralysie du cricket (CrPV), la petite sous-unité du ribosome fixe l'ARN directement au niveau de l'IRES. Dans d'autres cas, comme dans les picornavirus, le recrutement du ribosome s'effectue à l'aide des facteurs d'initiation de la traduction classiques, parfois associés à d'autres facteurs (cellulaires ou viraux) spécifiques des IRES appelés ITAF pour « IRES-specific Trans-Acting Factors ». L'initiation de la traduction s'effectue ensuite de manière classique. Selon Komar & Hatzoglou (2005).

Si la structure secondaire de l'ARNm est importante pour expliquer la très grande efficacité de traduction, il en va probablement de même des changements dynamiques qui en résultent et qui doivent animer cette molécule. Nous avons observé de tels changements dans plusieurs régions de l'ARNm comme le montrent les coupures par les sondes antagonistes ; ces changements pourraient aussi correspondre aux différentes conformations qui ont été observées sur gel natif. C'est sur cet aspect dynamique de la structure que repose notre modèle de traduction, la régulation de l'accès du codon initiateur s'effectuant par des

changements de conformation. C'est aussi par l'inhibition de cette dynamique que pourraient s'expliquer les effets des mutations ou des petits ARNs anti-sens sur la traduction. Ainsi, l'apparition de pauses induites par l'hybridation de l'ARN 9 (représentées sur la **figure 4C de l'article 5**) pourrait refléter la stabilisation de la conformation où les régions 9-31 et 272-309 sont appariées et où le codon d'initiation n'est par conséquent pas libre. Le ribosome se fixant sur l'AUG ne pourra donc pas être recruté. Cette hypothèse de travail suggère aussi que la conformation alternative en tige boucle impliquant les résidus 265-302 (**Figure 4B** encart) ne peut se former. Une hypothèse qui accorderait les différentes données serait que le ribosome se plaçant sur l'AUG nécessite la conformation où l'appariement 9-31 avec 272-309 est non formé alors que la tige boucle 265-302 est formée. Il est possible qu'une fois la traduction démarrée et l'AUG libéré du ribosome, la conformation pourrait basculer dans l'autre forme où l'appariement 9-31 avec 272-309 est formé alors que la tige boucle 265-302 n'existe plus. Le rôle du second ribosome placé sur l'extrémité 3' de l'ARNm reste obscur. Il pourrait favoriser la fixation du premier ribosome en stabilisant par exemple la tige boucle 265-302. En retour, le démarrage de la traduction et le basculement de conformation qui entraîne la disparition de la tige boucle 265-302 pourrait être le signal, pour le second ribosome, de gagner à son tour l'AUG initiateur. On devine aisément que, dans ce schéma, le recyclage du ribosome est quasiment instantané ce qui se traduit par un gain d'efficacité dans la vitesse globale de traduction. La question du « passage » du second ribosome sur l'AUG initiateur reste sans réponse à ce jour. Cependant, le ribosome est connu pour posséder une propriété singulière et rarement observée qui est celle de sauter des boucles d'ARN ou des tiges boucles de plusieurs dizaines de nucléotides (ribosome « bypassing », « jumping » ou « shunting ») (Futterer *et al.*, 1993; Maldonado & Herr, 1998; Yueh & Schneider, 1996). Bien que ce phénomène soit encore inexpliqué, il présente des similitudes frappantes avec le saut qu'il pourrait effectuer dans le cas de l'ARNm de l'histone H4.

Les résultats très originaux de notre étude mettent en évidence le caractère essentiel et exclusif de la phase codante pour la traduction. Ils viennent contredire d'autres travaux qui attribuent à la protéine HBP un rôle dans la traduction. En effet, nous n'avons pas été en mesure de détecter un effet stimulant de la protéine HBP sur la traduction de l'histone H4. Au contraire, nous avons montré que la délétion du site de fixation d'HBP situé à l'extrémité 3' non traduite était sans incidence sur l'efficacité de traduction. Il n'est pas exclu qu'HBP ait un rôle auxiliaire comme les autres équipes ont pu le détecter. Cependant, l'effet de la protéine, s'il existe, doit être très faible, comparé à la forte efficacité de traduction de la phase codante.

Ceci permettrait d'expliquer l'interaction découverte entre HBP (fixée sur sa tige boucle) et 2 protéines ribosomiques (Dominski *et al.*, 2002). Par ailleurs, d'autres travaux plus récents ont montré que le ribosome déplaçait la protéine HBP lors de l'étape de terminaison de la traduction (Kaygun & Marzluff, 2005b). Ces résultats mettent bien en évidence l'existence d'un lien entre HBP et le ribosome. Notre modèle suggère que le ribosome pourrait être séquestré au niveau de l'extrémité 3' de la région codante de l'ARNm. Il n'est donc pas exclu que la protéine HBP stabiliserait le ribosome fixé sur son site d'ancrage voisin et par là, stimulerait la traduction. Toutefois, compte tenu du fait que l'extrémité 3' non codante peut être délétée sans effet sur la traduction, on peut penser que l'effet d'HBP reste très auxiliaire (du moins *in vitro*).

Les changements structuraux qui animent certaines régions de la molécule pourraient aussi être un moyen de réguler l'expression d'un petit peptide interne exprimé à partir de l'ARNm. En effet, il existe dans l'ARNm d'H4-12, une courte phase de lecture ouverte qui code pour le précurseur d'un peptide appelé OGP (pour « Osteogenic Growth Peptide ») démarrant au niveau du second codon Met. Le peptide en question, long de 14 résidus est impliqué dans l'ostéogenèse et l'hématopoïèse. Ce peptide stimule la croissance osseuse au niveau des cellules de moelle osseuse où il est présent à des concentrations de l'ordre du μM . N'étant pas produit par clivage de l'histone ou par sa dégradation, il a été proposé que le peptide OGP serait traduit à partir du gène entier de l'histone H4, la sélection de l'AUG initiateur interne s'effectuant par « leaky scanning » de l'ARNm. En effet, la séquence Kozak de l'AUG initiateur de l'histone n'est pas optimale alors que celle du deuxième codon AUG initiant la traduction du peptide l'est parfaitement (Bab *et al.*, 1999). Cependant, si ce phénomène peut expliquer l'initiation interne de la traduction, il n'explique pas l'expression tissulaire du peptide dans la moelle osseuse. Plus récemment, il a été proposé que le peptide OGP serait plutôt produit à partir d'un variant d'histone H4-v.1 dont l'expression augmente spécifiquement dans certains tissus et qui peut être corrélée à une forte production du peptide (Poirier *et al.*, 2005). Le variant H4-v.1 de rat est transcrit en un long ARNm polyadénylé qui présente 23 mutations silencieuses par rapport à H4-12. La plupart de ces différences sont localisées dans les boucles de notre modèle en 2 dimensions, ce qui ne devrait pas altérer significativement la structure secondaire de l'ARNm. Cependant, plusieurs différences se situent au niveau des signaux de « toe print » que nous avons identifiés, ce qui suggère que la fixation des ribosomes pourrait être modifiée. Une hypothèse plausible serait que ces

mutations reprogrammeraient la traduction de l'ARNm en faveur du peptide OGP en favorisant le « leaky scanning » de l'ARNm et en diminuant l'expression de l'histone. Rappelons que l'ARNm de H4-v.1 est pourvu d'une queue de polyA qui pourrait faire de lui un substrat plus conventionnel pour la machinerie de traduction.

Le phénomène de traduction atypique ici présenté n'a pas encore été décelé pour les histones des autres classes. Cependant, les gènes d'histones H3 ont des phases codantes très conservées et présentent un fort taux en bases G et C, avec un biais pour ces nucléotides en troisième position du codon (DeBry & Marzluff, 1994). De plus, les histones H3 et H4 sont les histones les plus conservées et ne comportent que peu de variants contrairement aux autres classes d'histones. Ces protéines sont incorporées sous la forme d'un tétramère dans le nucléosome. Leur expression doit être synchronisée pour ne pas créer de déséquilibre (Meeks-Wagner & Hartwell, 1986). Tous ces éléments suggèrent que l'histone H3 pourrait présenter le même mode de traduction que l'histone H4 dans le nucléosome, mais à ce jour, les seuls arguments expérimentaux en faveur de cette hypothèse sont les profils de polysomes obtenus à partir de cellules HeLa traitées à la cycloheximide qui révèlent la présence des ARNm de l'histone H2a dans les fractions de monosomes et de disomes (Whitfield *et al.*, 2004).

Pour conclure, je voudrais souligner la similitude existant entre l'expression des gènes d'histones et de certaines protéines virales. Cette étude a révélé que l'efficacité de traduction de la protéine H4-12 *in vitro* est du même ordre de grandeur que celle d'une protéine virale témoin, c'est-à-dire très élevée. De même, la production massive d'histones est ciblée lors de la phase S du cycle cellulaire, mimant ainsi la phase de réplication virale durant laquelle le virus produit en grande quantité ses constituants. Enfin, comme la plupart des ARNm viraux, les ARNm d'histones ne contiennent pas d'introns, ne sont pas polyadénylés et leur traduction est initiée en détournant la machinerie cellulaire vers une structure de type IRES. Ainsi, pour répondre aux mêmes impératifs d'efficacité que ceux rencontrés par le virus, la cellule utilise des procédés similaires avec le gène de l'histone H4, démontrant une évolution convergente des deux systèmes. Enfin, je voudrais terminer en formulant l'hypothèse que les protéines de coque virale et les histones pourraient avoir une origine ancestrale commune. En effet, à toutes les ressemblances déjà évoquées se rajoute la fonction même de ces protéines, les protéines de coque virale empaquetant le génome viral et les histones l'ADN de la cellule.

4 ARTICLE 5

Two IRES elements in the coding region of histone H4 mRNA enable highly efficient translation

Sharief Barends^{1,2,4}, Sophie Jaeger^{1,4}, Roman I. Koning³, Henk K. Koerten³, Gilbert Eriani¹ and Franck Martin^{1*}

¹Département "Architecture et Réactivité des ARN" UPR 9002 du CNRS, Université Louis Pasteur,

Institut de Biologie Moléculaire et Cellulaire,

15, rue René Descartes, F-67084 Strasbourg cedex, France

²Present address: Department of Biochemistry, LIC, Leiden University

PO Box 9502, 2300 RA Leiden, The Netherlands

³Department of Molecular Cell Biology, Leiden University Medical Centre

PO Box 9503, 2300 RA, Leiden, The Netherlands

⁴These authors contributed equally to this work.

*Correspondence

E-mail address: F.Martin@ibmc.u-strasbg.fr

Tel: (33)388417042

Fax: (33)388602218

Summary

During S phase, enormous amounts of histones are translated from short, capped, non-polyadenylated mRNAs. Moreover, mouse H4-12 mRNA has an unusually short 5' UTR of nine nucleotides, making conventional preinitiation complex formation on the cap seemingly impossible. Here we show that while H4 translation is cap dependent, both UTRs are dispensable for translation. Moving the start codon with respect to the ORF has a detrimental effect on translation, suggesting that the ORF functions as a scaffold for ribosome binding while the AUG is encountered without ribosomal scanning. Elucidation of mRNA 2D structure revealed base pairing between the 5' and 3' ends of the ORF, which is critical for high translation efficiency. Surprisingly, one H4 mRNA is able to recruit two ribosomes before translation starts. This is the first example of a eukaryotic, cellular mRNA recruiting two ribosomes via structural elements exclusively located in its ORF.

Running title

H4 mRNA ORF recruits ribosomes for its translation

Introduction

In metazoans, histones are massively produced during the S phase of the cell cycle from so-called replication-dependent histone genes. Their transcripts are the only eukaryotic, cellular mRNAs without poly(A) tails at their 3' ends. Instead, replication-dependent histone mRNAs end in a highly conserved 16-nt hairpin structure (Marzluff and Duronio, 2002; Marzluff, 2005; Jaeger et al., 2005), which specifically interacts with the stem-loop binding protein (SLBP also called HBP for hairpin binding protein) (Wang et al., 1996; Martin et al., 1997) and ensures coordination of histone expression with DNA synthesis (Sullivan et al., 2001; Kodama et al., 2002; Pettitt et al., 2002; Zhao et al., 2004). Furthermore, using reporter mRNAs ending with the histone hairpin, SLBP has been shown to be the functional homolog of PABP in translation (Mangus et al., 2003) by interacting with the translation initiation factors eIF3, eIF4G and Paip (Ling et al., 2002; Sanchez and Marzluff, 2002; Gorgoni et al., 2005).

Histone mRNAs are the shortest mRNAs present in eukaryotic cells, not only because they encode small proteins (102 to 135 amino acids for the core histones) but also because they have the shortest UTRs. In the 3' UTRs, the terminal histone hairpin is usually located 30 to 70 nts downstream of the stop codon (Kaygun and Marzluff, 2005), while 5' UTRs are even smaller ranging from 20 to 60 nts, the extreme example being the mouse histone H4-12 which has a 9-nt 5' UTR (Meier et al., 1989). Histones, and especially H3 and H4, are among the most conserved proteins in the eukaryotic kingdom, and in higher eukaryotic species, H4 proteins show almost identical amino-acid sequences even between distantly related species such as animals and plants (Piontkivska et al., 2002). For example, H4 proteins from humans and the annelid worm *Platynereis dumerilii* show identical amino acid sequences, even though their common ancestor diverged almost 800 million years ago (Nei et al., 2001). Alignments of the coding regions of histone H4 mRNAs also show an unusual high conservation among very different species, exceeding the need for maintaining the conserved amino-acid sequence. In contrast, the level of divergence of the 5' and 3' flanking regions is rather high for the majority of H4 mRNAs (Piontkivska et al., 2002). Furthermore, histone coding regions are GC-rich and a strong bias for G and C at the wobble position of H3 rodent mRNAs has been described (DeBry and Marzluff, 1994). Finally, equal G and C contents in histone

coding regions but not in flanking sequences suggest that the sequence conservation at the nucleotide level might reflect a selection pressure on mRNA secondary structure during evolution. This hypothesis is reinforced by the observation that mutations of G or C occurring at codon positions 1 and 2 are counterbalanced by the complementary nucleotide in the third position of the codon (Huynen et al., 1992) and this is particularly true for H4 genes. Altogether, these data suggest that the secondary structure of histone mRNA has a functional importance in the translation process and/or in its mRNA stability. However, the structure of histone mRNA and its putative role in histone biosynthesis is largely unknown.

The enormous demand for histones during S phase (about 60 million copies per core histone are synthesized whereas general DNA-binding proteins are present at ten thousand copies per cell), together with the unorthodox organization of H4 mRNA (short 5' UTR, no poly(A) tail, highly conserved nucleotide sequence) hints at a special ribosomal translation mechanism. Here we analyzed translation of mouse H4-12 mRNA in its own context, rather than using reporter constructs, and our results show that both UTRs are dispensable for efficient histone translation and that only the coding region is responsible for ribosome recruitment. Moreover, the start codon is not encountered by a scanning ribosome and its position with respect to the ORF is crucial for translation initiation. Surprisingly, biochemical data together with electron microscopy revealed that one H4 mRNA molecule is able to recruit two initiating ribosomes at distinct regions of the ORF. These data, together with the secondary structure of H4 mRNA led to a model where 5'-3' base pairing within the coding region channels ribosomes via a rolling circle principle ensuring massive histone production.

Results

H4 mRNA, with or without its UTRs is an equal excellent substrate for the ribosome

In-vitro translation of capped histone mRNA in wheat-germ extract (WGE) and rabbit reticulocyte lysate (RRL) revealed that this mRNA is a surprisingly efficient substrate for eukaryotic ribosomes even under sub-saturating mRNA concentrations. Taking into account that H4 mRNA contains only one internal methionine (Met₈₅) and that the N-terminal Met₁ is processed for 80%, we determined that the yield of in-vitro translation of H4 is comparable to the highly efficient Tobacco mosaic viral RNA (10 pmoles protein product per 100 µl reaction; data not shown). We tested different truncated H4 transcripts in order to assess the contribution of the different regions to the high translation efficiency (Figure 1). Stabilities of the capped mRNA transcripts in WGE the translation extracts has been determined by semi-quantitative RT-PCR and shows no significant differences, except for transcript F which corresponds to the precursor mRNA. However, translation efficiency of F is equal to the other transcripts indicating that the slight difference in stability and the extended 3' UTR do not interfere with translation. Interestingly, transcript H, which contains only the ORF, is also efficiently translated. These results show that the UTRs are redundant for in-vitro translation of H4 mRNA suggesting that no scanning is required. As a control, we generated capped Green Fluorescence Protein (GFP) mRNA transcripts with and without 5' UTR. In this case, GFP is poorly translated without 5' UTR and the ribosome scans the mRNA to start translation from the next AUG, which is out of frame (supplemental data S1), corroborating with previous experiments done in WGE (Kozak, 1991). In contrast to what has been reported previously, we could also show that SLBP has no influence at all on translation efficiencies of H4 transcripts, which is in good agreement with the fact that UTRs are dispensable (data not shown).

Ribosome recruitment on histone mRNA is IRES-like although it is cap dependent

Translation of H4 mRNA is inhibited in the presence of increasing concentrations of the cap analogue m⁷GpppG as for Brome mosaic viral RNA translation which shows that cap-associated factors are required for efficient translation initiation (Figure 2). The minimal inhibiting concentration is 60 mM for H4 mRNA and 300 mM for BMV

RNA. Next, we determined if the ribosome scans the H4 mRNA to reach the start codon. When mutating the start codon to a threonine (AUG to ACG, Met₁Thr), translation is completely abolished indicating that translation only starts on the first methionine codon (Figure 3). To determine the ability of the ribosome to scan for an alternative start codon, in-frame AUGs were introduced at different positions. Translation could partially be restored, however translation efficiency rapidly decreases when the AUG is introduced further downstream (please note that introduction of an AUG codon in a wild-type background results in an H4 protein with two [³⁵S]-labelled, internal methionine residues). Moreover, an upstream in-frame AUG introduced at codon position -4 of both wild type and Met₁Thr backgrounds is not used as start codon. Altogether, these results demonstrate that the ribosome enters the H4 mRNA in the vicinity of the authentic start codon and does not scan downstream or upstream for another one. In addition, introduction of an AUG in the 5' UTR leading to an out-of-frame ORF does not affect translation of H4 confirming that the ribosome recognizes exclusively the AUG at its original location (data not shown). From the fact that the UTRs are not used for AUG selection and translation initiation, we conclude that structures in the ORF are required for ribosome recruitment. To characterize these regions we developed a strategy we refer to as primer walking. In-vitro transcribed antisense RNAs of 20 nts covering almost the whole H4 ORF were annealed to H4 mRNA and the resulting duplexes were translated in RRL and WGE (Figure 4A). An elongating ribosome is able to melt any RNA duplex (Takyar et al., 2005) while an initiating ribosome is blocked by antisense RNA annealed to the 5' UTR or in the vicinity of a start codon (Johansson et al., 1994). Indeed, primer 1 complementary to the AUG area of H4 dramatically decreases translation both in WGE and RRL. In addition, primers 7 and 9 reduce translation in WGE and RRL more than 50% indicating that these regions of the ORF are involved in translation initiation. In the case of WGE, additional primers (4, 11 and 13) also have significant effects on translation. The discrepancy between these results of primer walking in RRL and WGE is not understood so far but could be due to different ionic strengths and temperatures used for the two systems.

Secondary structure of H4 mRNA

Our results indicate that the structure of H4 mRNA is important for ribosome recruitment. Therefore, we determined the secondary structure of the whole H4

mRNA by chemical and enzymatic probing (Figure 4B). This molecule is very stable since 62 of the 77 base pairs are G-C. The UTRs are located outside of a so-called core domain containing the ORF of H4, which is in good agreement with our translation results showing that UTRs are redundant for translation. The probing results also revealed three areas containing single-strand- and double-strand-specific cuts indicating “breathing” of these sequences. These sequences (9-31, 159-165 and 221-231) are probably involved in dynamic interactions. In contrast, no cuts could be detected in sequence 293-322 indicating that these nucleotides are not accessible to the probes and/or the reverse transcriptase. To further understand the effect of translation inhibition observed by the primer-walk experiment, we also probed the structures of H4 mRNA annealed to primers 1, 7 and 9. No significant differences could be observed except for primer 9 which induced the appearance of strong reverse transcriptase stops on nts 24, 25, 28, 29, 30 and 31 suggesting that the sequence complementary to primer 9 might “communicate” directly or indirectly with this part of the molecule. This result prompted us to examine the effect of mutations at these positions (Figure 4C). Surprisingly, the silent mutation C₂₅G induces a ten-fold decrease in translation efficiency, while mutations GGC₂₃₋₂₅CCG had no effect on translation. Silent mutation U₃₁G and mutations GGU₂₉₋₃₁CCG also dramatically reduce the synthesis. Using these results and probing results, we propose that sequence 9-31 can be single stranded or alternatively can base pair with the 3' end of the ORF (272-275 and 294-309) (Figure 4C). To confirm this hypothesis we generated the compensatory mutation G₂₉₄C restoring the putative long-distance interaction. And indeed, the compensatory mutation G₂₉₄C increases translation of C₂₅G from 10 to 60%. The fact that mutations GGC₂₃₋₂₅CCG, G₂₉₄C, C₂₉₄A-C₂₉₈G and U₃₀₁G have no effect on translation can be explained by alternative base pairing of nts 23-25 with 290-294, which is impossible for C₂₅G (Figure 4C). Altogether these results indicate that nucleotide at position 25 has to be base-paired for proper translation. However, probing experiments showed that this nucleotide can also be unpaired in the structure of H4 mRNA. Therefore we propose that the sequence 9-31 can adopt two alternative conformations. Native gel electrophoresis of H4 mRNA transcripts confirmed that this molecule folds into two alternative conformations suggesting that the structure of this mRNA is dynamic (data not shown).

Ribosome-H4 mRNA complexes

To better understand the translation mechanism, we performed toe-print analysis to reveal the position of the initiating ribosome on H4 mRNA. Surprisingly, four regions with strong toe prints were detected on the ribosome-H4-mRNA complex at positions 85-87, 248-250, 294 and 307-310 (by convention, numbering of toe-prints indicate their position relative to nucleotide +1, the A of the start AUG) (Figure 5). Weaker stops could also be detected at +19, +70-72, +76, +78 and +98. The positions of the toe prints confirmed our previous results showing that the ORF is directly involved in ribosome recruitment. The +19 toe print designates the start codon positioned in the P site of the ribosome and indeed, the +19 toe print disappears in the presence of edeine and the non-hydrolysable GTP analogue GMP-PNP, and is absent for Met₁Thr. The other toe prints, however, are not affected by any of the different blocking agents used, suggesting unconventional ribosome binding. Interestingly, the toe-prints are not scattered all over the H4 mRNA but seem to be grouped on both ends of the ORF. An explanation for this might be that more than one ribosome can bind H4 mRNA. Therefore, we purified H4-mRNA-ribosome initiation complexes using an original pull-down strategy (supplemental data S2 and S3). Observations by electron microscopy (EM) revealed that our purification protocol was efficient and that the majority of the particles appeared to be monosomes while about 18,8 % were disomes, containing two ribosomes on one H4 mRNA molecule (Figure 6A, 6B and supplemental data S4). The same result was obtained with the Met₁Thr mutant confirming that toe-prints are identical compared to wild-type mRNA and that two ribosomes can bind on the H4 mRNA before any translation starts (Figure S4). Toe-prints at positions 294 and 307-310 revealed that the 3' end of the ORF is important for ribosome binding. Therefore, we synthesized a deletion mutant II, which ends at nucleotides 277 (Figure 5C). This mutant does not contain sequence 294-309 that can interact with the 5' end of the molecule (figure 4C). Kinetic analysis of in-vitro translation revealed that the yield of translation of deletion II is half compared to wild-type H4 mRNA, indicating that this mutant is able to recruit only one ribosome per H4 mRNA (Figure 6C). This was supported by EM analysis, which revealed that deletion mutant II, after correction for leaky translation when using cycloheximide (Arava et al., 2003 and supplemental data table 1 and S5) does not form disomes.

Discussion

UTRs are redundant for translation

Both 5' and 3' UTRs are dispensable for H4 translation. In agreement with this, primer-walk experiments revealed that parts of the coding region participate in translation initiation and strong toe prints were found clustered in two areas of the ORF. Moreover, the same toe prints are detected on an H4 mRNA with a mutated start codon. In contrast, toe-print experiments performed on GFP mRNA gave rise to only one stop (+19), reflecting the start codon positioned in the P site of the ribosome (data not shown).

Histone genes are very ancient and branched before the eukaryotic kingdoms diverged (Piontkivska et al., 2002). The UTRs have been acquired more recently during evolution and their main purpose is to restrict histone expression to the S phase of the cell cycle. This is crucial for viability since ill-timed expression of histones is highly toxic (Meeks-Wagner and Hartwell, 1986). Different strategies have been used to reach this confined expression explaining the relatively high sequence variability in the UTRs. Yeast and plants have polyadenylated histone mRNAs while metazoans have acquired sophisticated machinery for 3' end processing comprising many factors such as snRNP U7 and SLBP. The fact that SLBP might be a modern acquisition probably explains why it does not have any effect in our translation experiments (data not shown). Previous reports claim that SLBP is required for efficient translation but it seems rather to fulfil a minor, auxiliary role in translation (Sánchez and Marzluff, 2002; Ling et al., 2002; Gorgoni et al., 2005). The involvement of SLBP in translation has been argued by Kozak (Kozak, 2004) stating that the stimulation is rather low, and that unnatural systems have been used in combination with reporter mRNAs, thus overlooking any potential role for the coding region. However, we cannot exclude that SLBP has an auxiliary function *in vivo*. This would corroborate with the fact that three-hybrid screens revealed ribosomal proteins S15a and L22 as binding partners of SLBP (Dominski et al., 2002). The influence of translation termination on histone mRNA stability also suggests a link between a terminating ribosome and SLBP (Kaygun and Marzluff, 2005). Such interactions would reinforce our model of ribosome sequestration at the 3' end of the coding region (see below) and evolution might have improved this phenomenon by adding more interactions between the ribosome and the 3' UTR via SLBP.

Translation initiation without ribosomal scanning

Although H4 translation initiation is independent of the UTRs it does rely on the presence of the cap structure, suggesting that eIF4E and perhaps other factors are required for ribosome recruitment. This is quite analogous to the translation mechanism of hepatitis A virus. This virus is unique among all picornaviruses in that its IRES-directed translation initiation is inhibited in the presence of cap analogue m⁷GpppG, while picornaviral RNAs do not contain a cap structure (Bergamini et al., 2000). It has been suggested that translation requires a direct interaction of the IRES and eIF4E/eIF4G, involving the cap-binding pocket of eIF4E, or that eIF4G is required in a conformation which can only be achieved by binding of free eIF4E (Ali et al., 2001), and a similar principle may apply for H4 translation initiation. Additionally, the ribosome is recruited in direct vicinity of the start codon of H4 mRNA without scanning, like for hepatitis C virus (Otto and Puglisi, 2004) and cricket paralysis virus (Wilson et al., 2000), which do not make use of eIF4A.

The ancient protist *Giardia lamblia* is classified as one of the deeply branched and most primitive eukaryotes (Sogin et al., 1989). This organism shares striking similarities with the modern H4; most of the genes in *Giardia* do not contain introns (Nixon et al., 2002), and mRNAs contain short 5' (0-14 nts) and 3' (10-13 nts) UTRs (Adam, 2001). Translation is cap dependent while ribosome scanning does not seem to be used since eIF4A, eIF4G and eIF4H are not encoded on the genome (Li and Wang, 2004). Interestingly, the coding region of a giardiaviral mRNA harbors a so-called downstream box of 13 nts that is complementary to the 3' end of the 16S-like ribosomal RNA, suggesting a Shine-Dalgarno type of interaction (Garlapati et al., 2001). Like in our H4-mRNA situation, the AUG is positioned in a 31-nts sequence and cannot be repositioned without translation defects (Garlapati and Wang, 2002). While a strong selection pressure both on H4 protein sequence and mRNA structure suppressed nucleotide sequence divergence, one can speculate that histone mRNA represents a primordial scaffold for primitive translation initiation.

Two ribosomal entry sites on one histone mRNA

Based on primer-walk and toe-print data we hypothesized the synchronous recruitment of two initiating ribosomes on one H4 mRNA. EM on purified ribosome

complexes revealed that about 13% were disomes. The same extent of disome formation was observed when the inactive Met₁Thr template was used. Due to the ten-fold excess of mRNA over ribosomes we can exclude the possibility of an artifact. Although this result is surprising, disome formation is not unprecedented. The 5' UTR of TMV RNA, also called the Ω fragment, is a known, efficient translation enhancer which recruits two initiating ribosomes and even trisome formation could be detected (Filipowicz and Haenni, 1979; Takahashi et al., 2005). While the first ribosome is positioned on the start codon, the second one binds at an upstream AUU triplet in an initiator tRNA^{Met}_i-dependent manner, which has been named "80S-preinitiation complex". Interestingly, mouse H4 mRNA has only one AUU triplet in its coding region, located in the -1 frame (A₃₁₀-U₃₁₂) just upstream of the strong toe prints observed in the 3' part of H4 mRNA. It is tempting to speculate that disome formation on H4 mRNA is mechanistically similar to that on TMV RNA, and indeed, both mRNAs are equally good substrates for ribosomal translation.

Model for histone translation

Data from secondary structure probing combined with mutational analysis showed that the 5' and 3' parts of the coding region interact, probably in a dynamic manner. Moreover, this end-to-end communication is important for efficient histone synthesis. Based on these results, together with the toe-print analysis and EM, we suggest a model where two ribosomes are simultaneously bound to one H4 mRNA and alternate between active translation and accommodation (or 80S-pre-initiation complex formation) at or near the AUU triplet in the 3' end (see Figure 7). The first ribosome (ribosome 1) is recruited directly to the AUG codon without any scanning. This is facilitated by the single-stranded nature of the sequence that contains the AUG, and indeed, its length (31 nts) can exactly accommodate a 40S subunit (Kozak, 1997; Sachs et al., 2002). While ribosome 1 starts translation it liberates the AUG-containing region to interact with the 3' part of the molecule. By this structural rearrangement the AUG shifts into a favourable context for the second ribosome (ribosome 2) located at the 3'-terminal IRES element, to assemble on it. And indeed, the AUU triplet shifts into close proximity of the start codon. The AUG start codon has to be located in certain window and cannot be moved outside of this frame without decreasing translation, fitting with the model for translational initiation of both

ribosomes. Ribosome 1 terminates translation and thereby dissociates 5'-3' interactions allowing ribosome 2 to enter in elongation phase. By this means, the histone H4 mRNA is continuously translated by two trapped ribosomes, which are reused without leaving the mRNA. In our model, we hypothesize that the two IRES elements located on both ends of the coding region, combined with ribosome re-channeling create high translation efficiency.

Other histone (-derived) proteins

H4 mRNA encodes only two methionine residues, the start codon and Met₈₅. Translation initiation can also occur from Met₈₅ leading to the synthesis of the 19 amino-acid long osteogenic growth peptide (OGP) (Bab et al., 1992), which stimulates bone-mass increase in bone marrow cells. The mechanism underlying the selection of Met₈₅ as start codon is still mysterious, although it has been proposed to occur via leaky scanning over Met₁ (Bab et al., 1999; Smith et al., 2005). Indeed, the nucleotide context around the first AUG has a sub-optimal Kozak sequence, whereas the second AUG is surrounded by a perfect Kozak consensus sequence. Gendron (Gendron et al., 1998) discovered a long, polyadenylated H4-mRNA variant termed H4-v.1, which has recently been found to be specifically enriched in OGP-expressing cells, suggesting that this mRNA is the actual template for OGP synthesis (Poirier et al., 2005). The H4 protein sequence encoded by H4-v.1 is identical to canonical H4 but there are significant changes on the nucleotide level, most of which are located in loops that do not alter the mRNA structure. The other changes are exclusively located around the toe prints we found, and may act as anti determinants preventing translation initiation from the first AUG. Interestingly and in contrast to the canonical H4 mRNA, there is no AUU triplet in H4-v.1 mRNA, suggesting that on this mRNA no disome formation can occur. Based on the facts that H4-v.1 mRNA is polyadenylated, AUG1 is in a sub-optimal Kozak context and the 5' UTR is of conventional length (40 nts), it seems plausible that OGP is synthesized from H4-v.1 mRNA by leaky scanning over the first AUG and that H4 mRNA is not template for OGP synthesis.

Our results show that H4 translation is mediated by interactions of the coding region with the ribosome, so the next question is: what about the other histones? In analogy to H4 mRNA, sequence comparison of rodent histone H3 mRNAs revealed an unusually high sequence conservation of coding region and a high GC content

with a strong bias to G or C at the third base (DeBry and Marzluff, 1994). In mouse, H4 genes encode a unique protein sequence while H3 genes encode only two variants, which can be attributed to the fact that H3 and H4 interact in a tetramer and initiate nucleosome assembly. In contrast H2A and H2B genes encode ~10-12 different proteins (Marzluff et al., 2002). These observations strongly suggest that H3 translation might be analogous to H4, whereas for H2A and H2B mRNAs no straightforward correlation with H4 mRNA could be deciphered. However, recent experiments have shown that in vivo H2A mRNA is mainly detected in monosomes and disomes (Whitfield et al., 2004).

Concluding remarks

Histones are in many ways unique: although restricted to eukaryotes they are encoded by genes that do not contain any introns; metazoans have a specialized, complex 3' processing machinery (snRNP U7 and SLBP) for a single cleavage event; and metazoan histone mRNAs are not polyadenylated. Additionally, the remarkable high conservation at the nucleotide level (even between only distantly related organisms) strongly suggests that the mRNA structure might play an essential role in histone biogenesis. We show here that two ribosomes are simultaneously recruited by the coding region and translate the histone mRNA by a treadmill mechanism. This might be an explanation why this atypical mRNA is such an efficient template for translation. The findings that histone translation does not make use of ribosomal scanning and uses a second, stand-by ribosome for high translation efficiency are shared by some viruses and archaic organisms, suggesting that the current histone translation mechanism might be a representation of primitive protein synthesis at the time when eukaryotes diverged.

Experimental Procedures

In-vitro transcription of H4 mRNA

The gene encoding mouse H4 (clone 12) has been amplified by PCR from mouse genomic DNA and sub cloned in the *EcoRI* site of pUC19 (Stratagene). The sequence has been verified by sequencing several clones from two independent amplifications; our sequence has one deletion (T564) in the 3' UTR compared to the previously described sequence (Acc. Numb. X13235) (Meier et al., 1989). Transcription templates were generated by PCR amplification from the plasmid pUC19-H4. The 5' primers contained the T7 promoter sequence and the 3' primers were designed in order to promote in-vitro run-off transcription at the desired position. The amplified templates were used for in-vitro transcription with T7 RNA polymerase in the presence of m⁷GpppG (New England Biolabs). The resulting transcripts were separated by PAGE (6-10%) and purified.

In-vitro translations in WGE and RRL

Translation reactions in nuclease-treated rabbit reticulocytes lysate (RRL) and wheat germ extract (WGE) were performed as described by the manufacturer (Promega). Reactions were incubated at 30°C (RRL) and 25°C (WGE) for 90 minutes and included 250 nM transcripts, 100 mM KAc (for WGE) and 3.8 µCi [³⁵S]Met. Aliquots of translation reactions were analyzed by 15% SDS-PAGE and [³⁵S]-labeled bands were quantified on a phosphor imager. The cap dependency was analyzed by preincubation in the presence of increasing m⁷GpppG concentrations ranging from 0.5 µM to 1.5 mM for 5 min at room temperature.

RNA primer walk

Anti-sense 20-mer RNA primers were in-vitro transcribed with T7 RNA polymerase and purified by denaturing PAGE (10%). A 5-fold excess of RNA primers was annealed to histone mRNA and the duplex was purified on a Micro Bio-Spin (P30) column (Bio-Rad). Translation of RNA duplexes in WGE and RRL was carried out as described above. Quantitative and proper annealing of the RNA primers at the correct positions of H4 mRNA has been checked by RT-PCR, RNase H cleavage experiments, primer extension and band-shift analysis. Selective translation inhibition

of H4 mRNA has been confirmed by checking that anti-sense primers have no effect on translation of GFP mRNA.

Toe-print analysis

For forming translation-initiation complexes, 250 nM H4 mRNA was incubated in a 15- μ l mixture containing 50% [vol./vol.] RRL, 50 mM KAc, 100 μ M of each amino acid, 1 U/ μ l RNaseOUT (Invitrogen), complemented with 2 mM GMP-PNP, 1 mg/ml cycloheximide, 2 mg/ml hygromycin or 10 μ M edeine. Complexes were formed and analyzed essentially as described (Pestova et al., 1996; Wilson et al., 2000), except that 1U/ μ l of RNase H-minus moloney murine leukemia virus reverse transcriptase (Promega) was used. After primer extension reactions, using primer 140-160 (5'-TGAGGCCGGAGATGCGCTTC-3') or primer 338-356 (5'-TTTTGTTGACAGTGGAAAATCG-3') deoxy oligonucleotides (Invitrogen), copy DNA was purified by phenol extraction and ethanol precipitation. Samples were separated by denaturing PAGE (8%), analyzed by phosphor imaging and compared with appropriate dideoxy nucleotide sequencing ladders.

Isolation of translation-initiation complex

In order to isolate translation-initiation complexes, we modified the H4 mRNA via the splint-ligation method with a 3' covalently linked, biotinylated, deoxy oligonucleotide (Isogen Life Science) (see supplemental Figure S2). For this, the 3' end of H4 mRNA was extended with 5'-(CAA)₉CAC-3' by means of PCR. In agreement with our results showing that UTRs are redundant in H4 synthesis, the modified RNA is translated as efficiently as the wild-type H4 mRNA (data not shown). In a 40- μ l reaction volume 600 pmoles of 5'-capped H4 mRNA was incubated with equimolar amounts of each of the two different splint deoxy oligonucleotides 5'-(GAG)₄GT(GTT)₄-3' and 5'-(GAG)₄NGT(GTT)₄-3', where 'N' represents any nucleotide (in order to ligate mRNAs with a 3' +1 nucleotide extension), together with an equimolar amount of a 5'-phosphorylated, 3'-biotinylated bait deoxy oligonucleotide (5'-p-(CTC)₄(A)₆-biotine-3'). The mixture was heated for 1 min at 94°C and chilled on ice for 15 min. Ligation was done with 12 units of T4 DNA ligase (Promega) and in the presence of 80 units RNaseOUT at 4°C for 20 hrs. The ligation mixture was bound to pre-washed streptavidin magnesphere paramagnetic particles (SA-PMPs, Promega) and washed

thoroughly with H₂O by magnetic separation in a stand. This method yielded typically 60-80 pmole of immobilized 5'-m⁷Gppp-H4-(CAA)₉CAC-(CTC)₄(A)₆-biotine-3'.

A 200- μ l translation reaction containing 50% [vol./vol.] RRL, 100 μ M of each amino acid, 50 mM KAc and 1 U/ μ l of RNaseOUT was pre-incubated for 20 min at 30°C and then chilled on ice for 20 min. The mixture, containing about 8 pmoles of ribosomes, was complemented with 60-80 pmoles of immobilized H4 mRNA and 1 mg/ml cycloheximide. After 5' incubation at 30°C, 10 μ l of a 320-mM Mg(Ac)₂ solution was added and chilled for 3 min on ice. Then, 200 μ l of buffer A (20 mM Tris-HCl [pH 7.6], 100 mM KAc, 2.5 mM Mg(Ac)₂, 2 mM DTT, 1 mM ATP, 0.1 mM GMP-PNP, and 0.25 mM spermidine (Wilson et al., 2000) was added and the mixture was incubated for another 5 min on ice. The SA-PMPs were washed by means of magnetic separation in a stand with ice-cold buffers (slightly modified from (Halic et al., 2004): twice with 400 μ l buffer 250 (50 mM Tris-HCl [pH 7.0], 250 mM KAc, 25 mM Mg(Ac)₂, 250 mM sucrose, 0.1% [vol./vol.] Igepal, 5 mM β -mercapthoethanol, 100 μ g/ μ l cycloheximide, and 1 mM AEBSF protease inhibitor), once with 200 μ l buffer 500 (buffer 250 containing 500 mM KAc, 50 mM sucrose and 0.05% [vol./vol.] Igepal), and once with 200 μ l buffer 50 (20 mM Tris-HCl [pH 7.0], 50 mM KAc, 10 mM Mg(Ac)₂, 1 mM DTT, 1 mM CaCl₂, 100 μ g/ μ l cycloheximide, and 1 mM AEBSF). For eluting the translation initiation complex the SA-PMPs were resuspended in 50 μ l buffer 50 complemented with 100 U of RNaseOUT and 2.5 U of RQ1 RNase-free DNase (Promega) and incubated at ambient temperature for 30 min (see supplemental Figure S3).

Electron Microscopy

Translation-initiation complexes were specifically purified for EM as described above and freshly used on the same day. The concentration of the complexes was typically ~10 nM. Carbon-coated, formvar-stabilized, 400-Mesh copper grids were negatively glow discharged in air for 2 minutes directly before use. First, 3 μ l of the translation initiation complex was positioned on the grid for about 30 seconds to adsorb the complexes to the carbon layer and excess solution is blotted off by filter paper. Subsequently, 3 μ l of a 3% [w./vol.] uranyl-acetate solution was used to negatively stain the complexes. Electron micrographs were recorded on a Philips CM10 at 100 keV and a magnification of 28,500x. Images were developed in KODAK D19

developer for 4 minutes and scanned on a Nikon super coolscan 8000 ED scanner at 2,000 dpi.

Acknowledgments

We are indebted to Mathieu Rederstorff and Lyuba Ryabova for valuable reagents, Agnès Clénet for skilful technical assistance, Dico van Meerten (Promega Benelux), Lyuba Ryabova and Eric Westhof for helpful discussions and Richard Giegé for constant support. S.B. was supported by an EMBO long-term fellowship and S.B. and R.I.K. were supported by VENI grants from the Netherlands organisation for Scientific Research (NWO).

References

- Adam, R.D. (2001). Biology of *Giardia lamblia*. *Clin. Microbiol. Rev.* **14**, 447-475.
- Ali, I.K., McKendrick, L., Morley, S.J., and Jackson, R.J. (2001). Activity of the hepatitis A virus IRES requires association between the cap-binding translation initiation factor (eIF4E) and eIF4G. *J. Virol.* **75**, 7854-7863.
- Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **100**, 3889-3894.
- Bab, I., Gazit, D., Chorev, M., Muhlrads, A., Shteyer, A., Greenberg, Z., Namdar, M., and Kahn, A. (1992). Histone H4-related osteogenic growth peptide (OGP): a novel circulating stimulator of osteoblastic activity. *EMBO J.* **11**, 1867-1873.
- Bab, I., Smith, E., Gavish, H., Attar-Namdar, M., Chorev, M., Chen, Y., Muhlrads, A., Birnbaum, M., Stein, G., and Frenkel, B. (1999). Biosynthesis of osteogenic growth peptide via alternative translational initiation at AUG85 of histone H4 mRNA. *J. Biol. Chem.* **274**, 14474-14481.
- Bergamini, G., Preiss, T., and Hentze, M.W. (2000). Picornavirus IRESes and the poly(A) tail jointly promote cap-independent translation in a mammalian cell-free system. *RNA* **6**, 1781-1790.
- Brunel, C., and Romby, P. (2000). Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods Enzymol.* **318**, 3-21.
- DeBry, R.W., and Marzluff, W.F. (1994). Selection on silent sites in the rodent H3 histone gene family. *Genetics* **138**, 191-202.
- Dominski, Z., Erkmann, J.A., Yang, X., Sánchez, R., and Marzluff, W.F. (2002). A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3'-end processing. *Genes Dev.* **16**, 58-71.
- Filipowicz, W., and Haenni, A.-L. (1979). Binding of ribosomes to 5'-terminal leader sequences of eukaryotic messenger RNAs. *Proc. Natl. Acad. Sci. USA* **76**, 3111-3115.
- Garlapati, S., Chou, J., and Wang, C.C. (2001). Specific secondary structures in the capsid-coding region of giardavirus transcript are required for its translation in *Giardia lamblia*. *J. Mol. Biol.* **308**, 623-638.

- Garlapati, S., and Wang, C.C. (2002). Identification of an essential pseudoknot in the putative downstream internal ribosome entry site in giardiavirus transcript. *RNA* 8, 601-611.
- Gendron, N., Dumont, M., Gagné, M.-F., and Lemaire, S. (1998). Poly A-containing histone H4 mRNA variant (H4-v.1): isolation and sequence determination from bovine adrenal medulla. *Biochim. Biophys. Acta* 1396, 32-38.
- Gorgoni, B., Andrews, S., Schaller, A., Schümperli, D., Gray, N., and Müller, B. (2005). The stem-loop binding protein stimulates histone translation at an early step in the initiation pathway. *RNA* 11, 1030-1042.
- Halic, M., Becker, T., Pool, M.R., Spahn, C.M.T., Grassucci, R.A., Frank, J., and Beckmann, R. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature* 427, 808-814.
- Huynen, M.A., Konings, D.A.M., and Hogeweg, P. (1992). Equal G and C contents in histone genes indicate selection pressures on mRNA secondary structure. *J. Mol. Evol.* 34, 280-291.
- Jaeger, S., Barends, S., Giegé, R., Eriani, G., and Martin, F. (2005). Expression of metazoan replication-dependent histone genes. *Biochimie* 87, 827-834.
- Johansson, H.E., Belsham, G.J., Sproat, B.S., and Hentze, M.W. (1994). Target-specific arrest of mRNA translation by antisense 2'-O-alkyloligoribonucleotides. *Nucleic Acids Res.* 22, 4591-4598.
- Kaygun, H., and Marzluff, W.F. (2005). Translation termination is involved in histone mRNA degradation when DNA replication is inhibited. *Mol. Cell Biol.* 25, 6879-6888.
- Kodama, Y., Rothman, J.H., Sugimoto, A., and Yamamoto, M. (2002). The stem-loop binding protein CDL-1 is required for chromosome condensation, progression of cell death and morphogenesis in *Caenorhabditis elegans*. *Development* 129, 187-196.
- Kozak, M. (1991). A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expr.* 1, 111-115.
- Kozak, M. (1997). Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J.* 16, 2482-2492.
- Kozak, M. (2004). How strong is the case for regulation of the initiation step of translation by elements at the 3' end of eukaryotic mRNAs? *Gene* 343, 41-54.

- Li, L., and Wang, C.C. (2004). Capped mRNA with a single nucleotide leader is optimally translated in a primitive eukaryote, *Giardia lamblia*. *J. Biol. Chem.* 279, 14656-14664.
- Ling, J., Morley, S.J., Pain, V.M., Marzluff, W.F., and Gallie, D.R. (2002). The histone 3'-terminal stem-loop-binding protein enhances translation through a functional and physical interaction with eukaryotic initiation factor 4G (eIF4G) and eIF3. *Mol. Cell. Biol.* 22, 7853-7867.
- Mangus, D.A., Evans, M.C., and Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol.* 4, 223.1-14.
- Martin, F., Schaller, A., Eglite, S., Schümperli, D., and Müller, B. (1997). The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. *EMBO J.* 16, 769-778.
- Marzluff, W.F. (2005). Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Curr. Opin. Cell Biol.* 17, 274-280.
- Marzluff, W.F., and Duronio, R.J. (2002). Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Curr. Opin. Cell Biol.* 14, 692-699.
- Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002). The human and mouse replication-dependent histone genes. *Genomics* 80, 487-498.
- Meeks-Wagner, D., and Hartwell, L. (1986). Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. *Cell* 44, 43-52.
- Meier, V.S., Bohni, R., and Schümperli, D. (1989). Nucleotide sequence of two mouse histone H4 genes. *Nucleic Acids Res.* 17, 795.
- Nei, M., Xu, P., and Glazko, G. (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl. Acad. Sci. USA* 98, 2497-2502.
- Nixon, J., Wang, A., Morrison, H., McArthur, A., Sogin, M., Loftus, B., and Samuelson, J. (2002). A spliceosomal intron in *Giardia lamblia*. *Proc. Natl. Acad. Sci. USA* 99, 3701-3705.
- Otto, G.A., and Puglisi, J.D. (2004). The pathway of HCV IRES-mediated translation initiation. *Cell* 119, 369-380.

- Pestova, T.V., Hellen, C.U.T., and Shatsky, I.N. (1996). Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol. Cell. Biol.* **16**, 6859-6869.
- Pettitt, J., Crombie, C., Schümperli, D., and Müller, B. (2002). The *Caenorhabditis elegans* histone hairpin-binding protein is required for core histone gene expression and is essential for embryonic and postembryonic cell division. *J. Cell. Sci.* **115**, 857-866.
- Piontkivska, H., Rooney, A.P., and Nei, M. (2002). Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol. Biol. Evol.* **19**, 689-697.
- Poirier, R., Lemaire, I., Dumont, M., Leduc, N., Le, H.-T., and Lemaire, S. (2005). Correlation between the expression of the histone H4 mRNA variant H4-v.1 and the levels of histone H4-(86-100) and H4-(89-102) (OGP) in various rat tissues and alveolar macrophages. *Peptides* **26**, 1503-1511.
- Sachs, M.S., Wang, Z., Gaba, A., Fang, P., Belk, J., Ganesan, R., Amrani, N., and Jacobson, A. (2002). Toeprint analysis of the positioning of translation apparatus components at initiation and termination codons of fungal mRNAs. *Methods* **26**, 105-114.
- Sánchez, R., and Marzluff, W.F. (2002). The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. *Mol. Cell. Biol.* **22**, 7093-7104.
- Smith, E., Meyerrose, T.E., Kohler, T., Namdar-Attar, M., Bab, N., Lahat, O., Noh, T., Li, J., Karaman, M.W., Hacia, J.G., *et al.* (2005). Leaky ribosomal scanning in mammalian genomes: significance of histone H4 alternative translation in vivo. *Nucleic Acids Res.* **33**, 1298-1308.
- Sogin, M.L., Gunderson, J.H., Elwood, H.J., Alonso, R.A., and Peattie, D.A. (1989). Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* **243**, 75-77.
- Sullivan, E., Santiago, C., Parker, E.D., Dominski, Z., Yang, X., Lanzotti, D.J., Ingledue, T.C., Marzluff, W.F., and Duronio, R.J. (2001). *Drosophila* stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. *Genes Dev.* **15**, 173-187.
- Takahashi, Y., Hirayama, S., and Odani, S. (2005). Ribosomal proteins cross-linked to the initiator AUG codon of a mRNA in the translation initiation complex by UV-irradiation. *J. Biochem.* **138**, 41-46.

- Takyar, S., Hickerson, R.P., and Noller, H.F. (2005). mRNA helicase activity of the ribosome. *Cell* 120, 49-58.
- Wang, Z.F., Whitfield, M.L., Ingledue, T.C.-3rd, Dominski, Z., and Marzluff, W.F. (1996). The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes Dev.* 10, 3028-3040.
- Whitfield, M.L., Kaygun, H., Erkmann, J.A., Townley-Tilson, W.H.D., Dominski, Z., and Marzluff, W.F. (2004). SLBP is associated with histone mRNA on polyribosomes as a component of the histone mRNP. *Nucleic Acids Res.* 32, 4833-4842.
- Wilson, J.E., Pestova, T.V., Hellen, C.U.T., and Sarnow, P. (2000). Initiation of protein synthesis from the A site of the ribosome. *Cell* 102, 511-520.
- Zhao, X., McKillop-Smith, S., and Müller, B. (2004). The human histone gene expression regulator HBP/SLBP is required for histone and DNA synthesis, cell cycle progression and cell proliferation in mitotic cells. *J. Cell. Sci.* 1, 6043-6051.

Figure 1

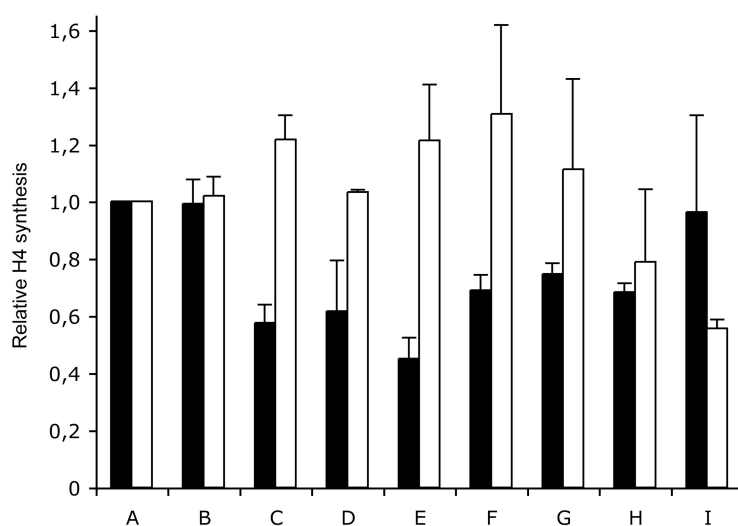
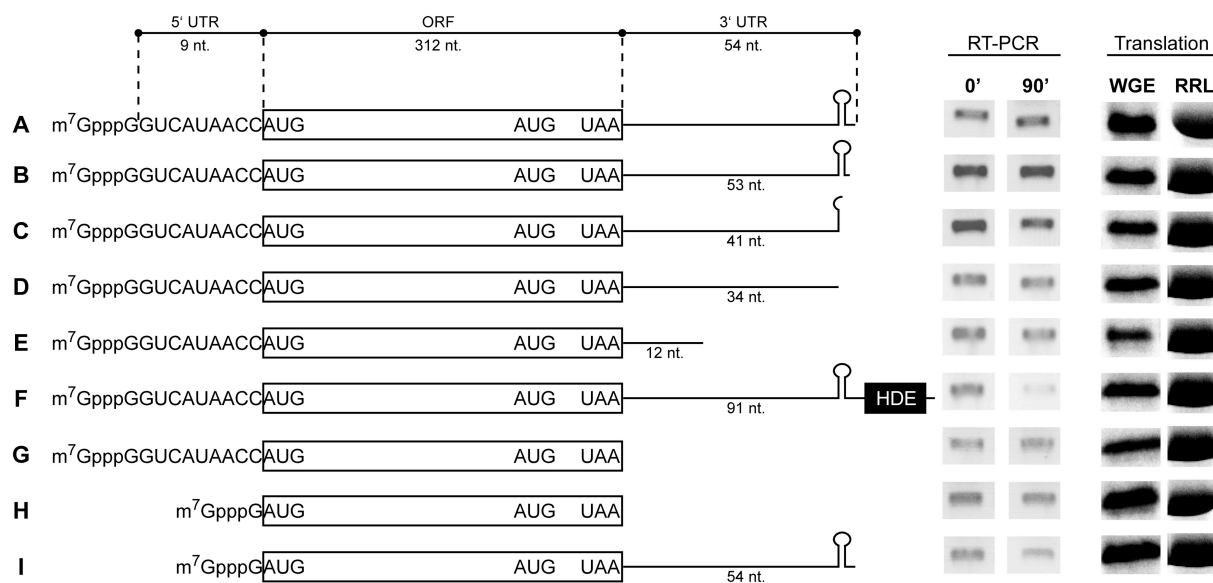


Figure 2

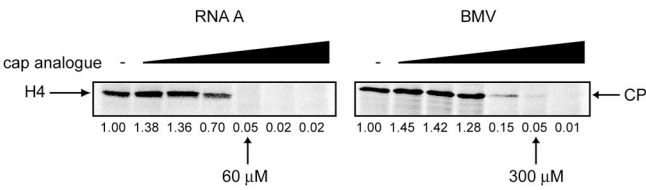


Figure 3

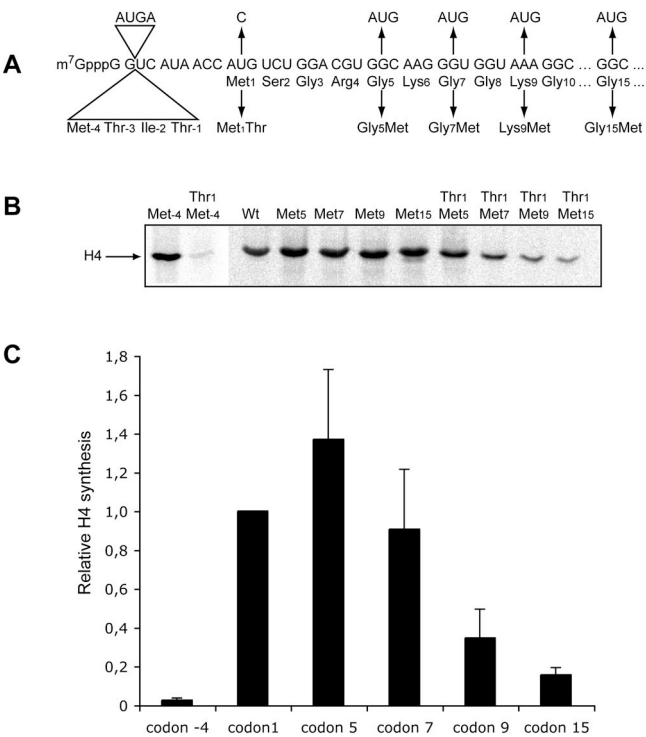


Figure 4

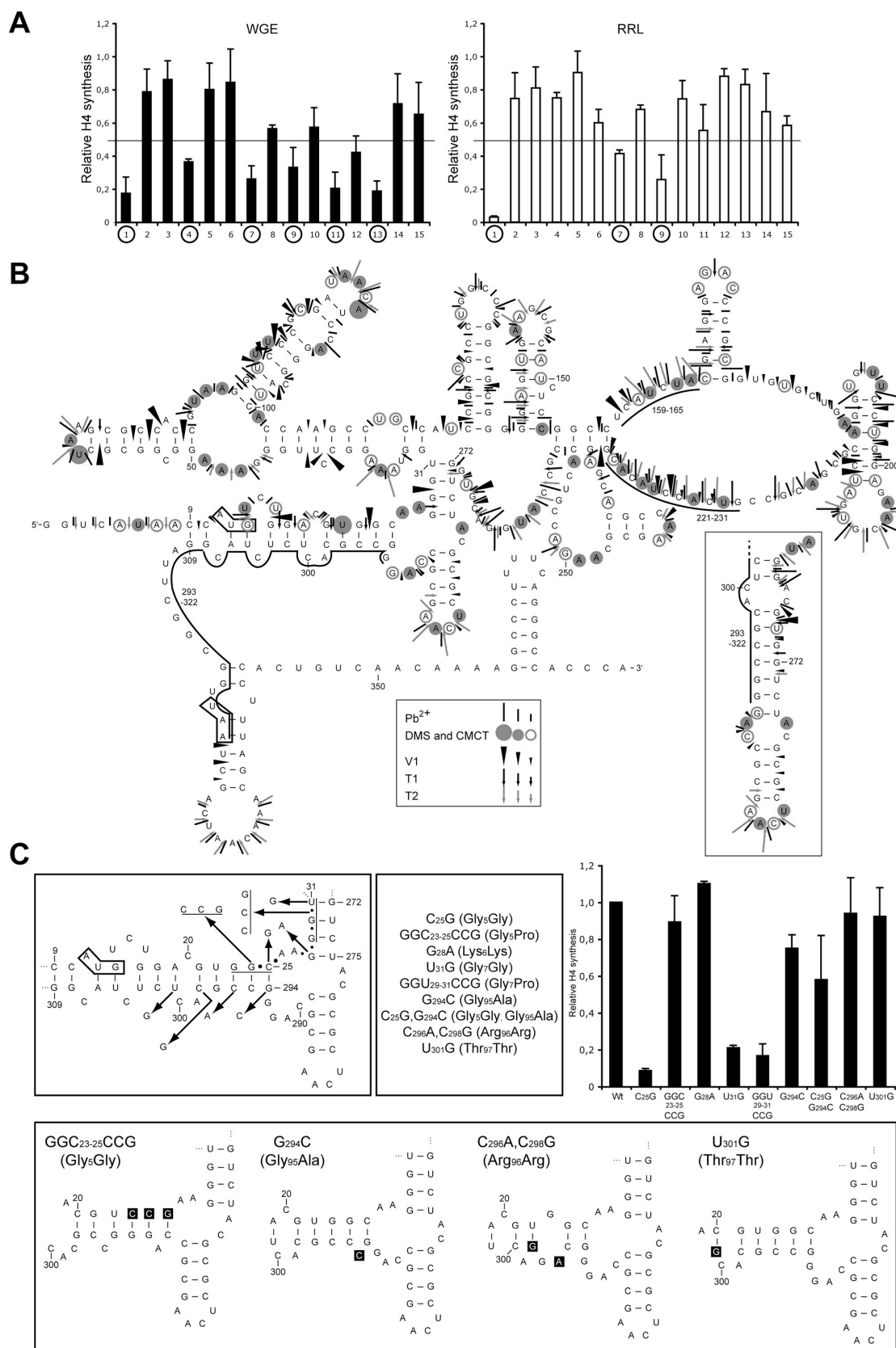


Figure 5

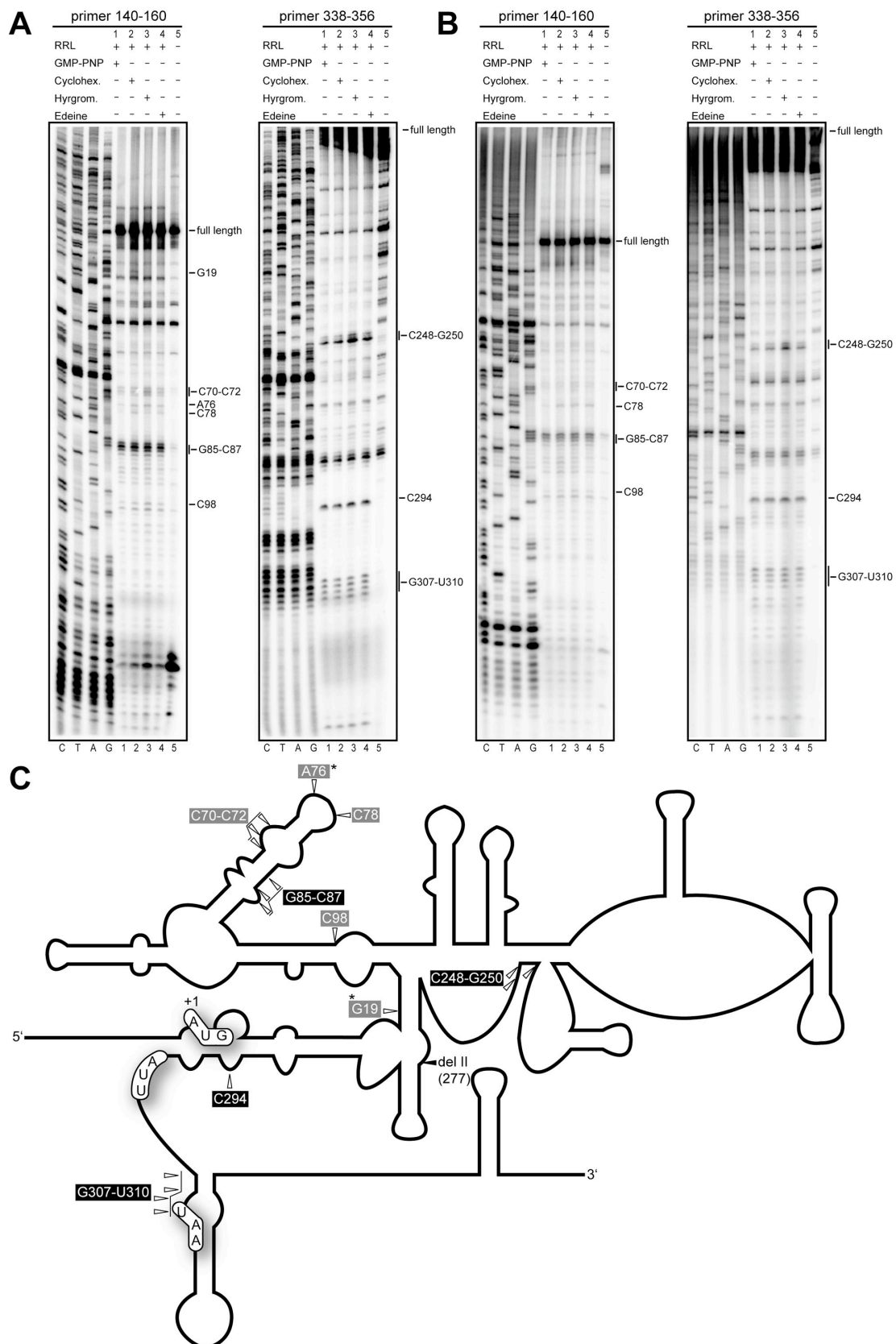


Figure 6

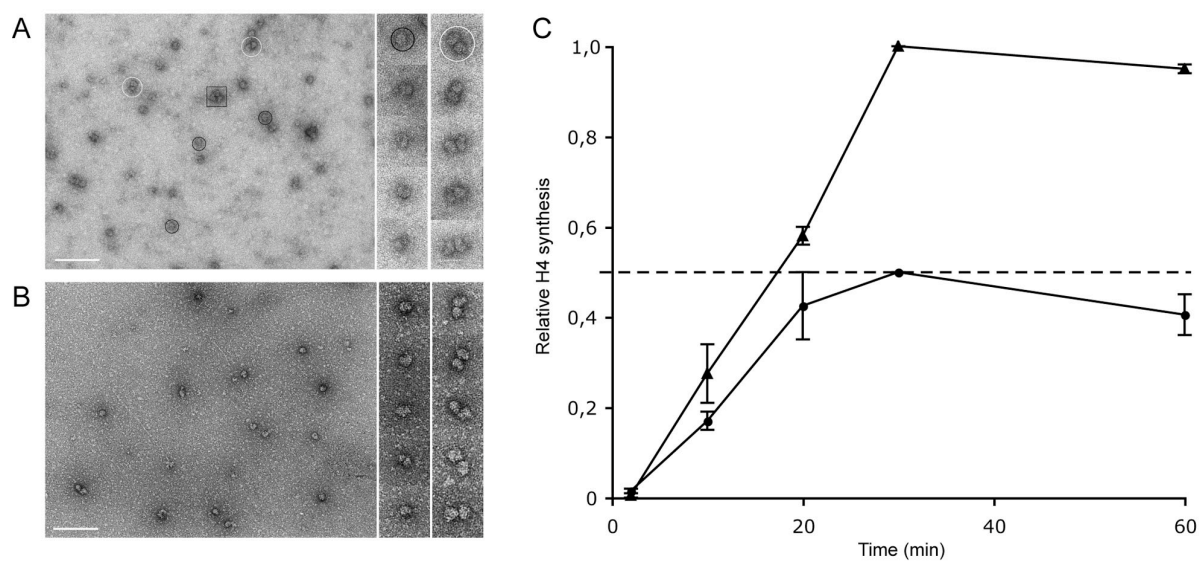
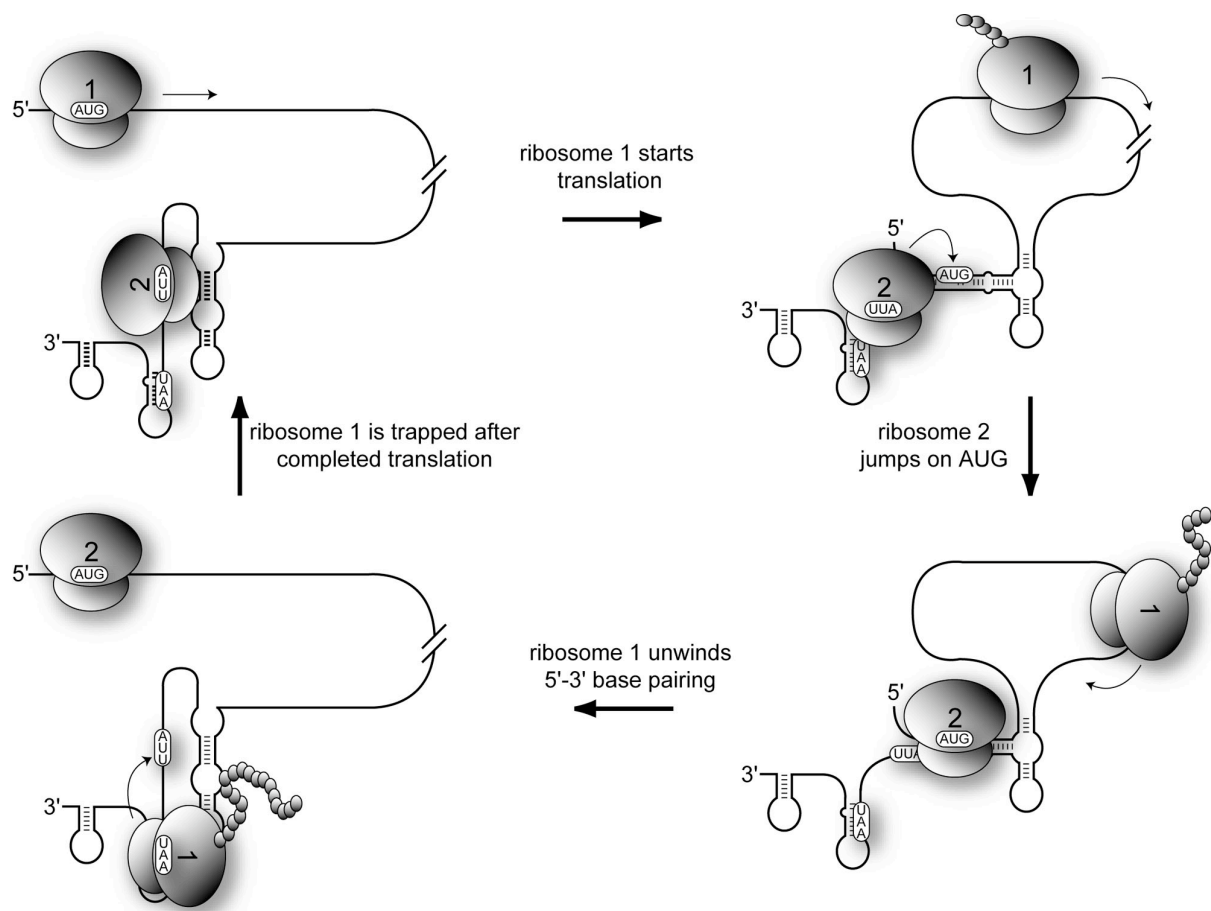
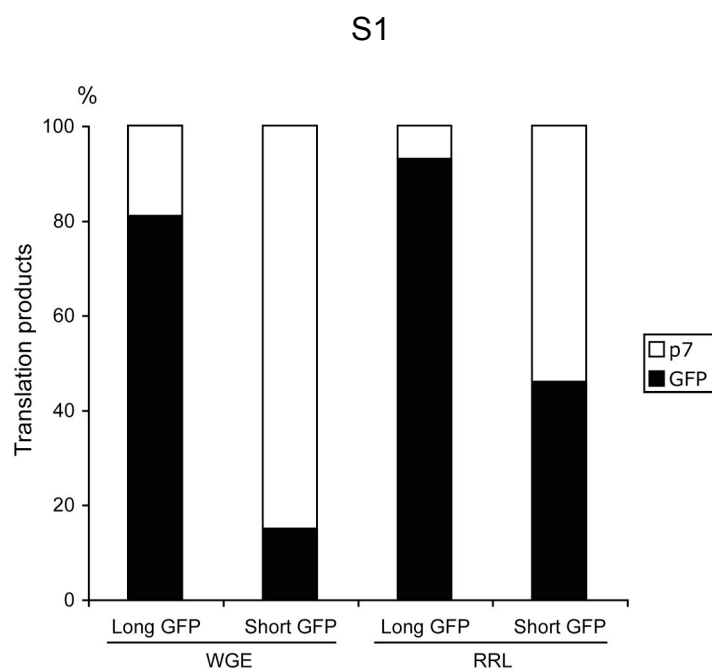
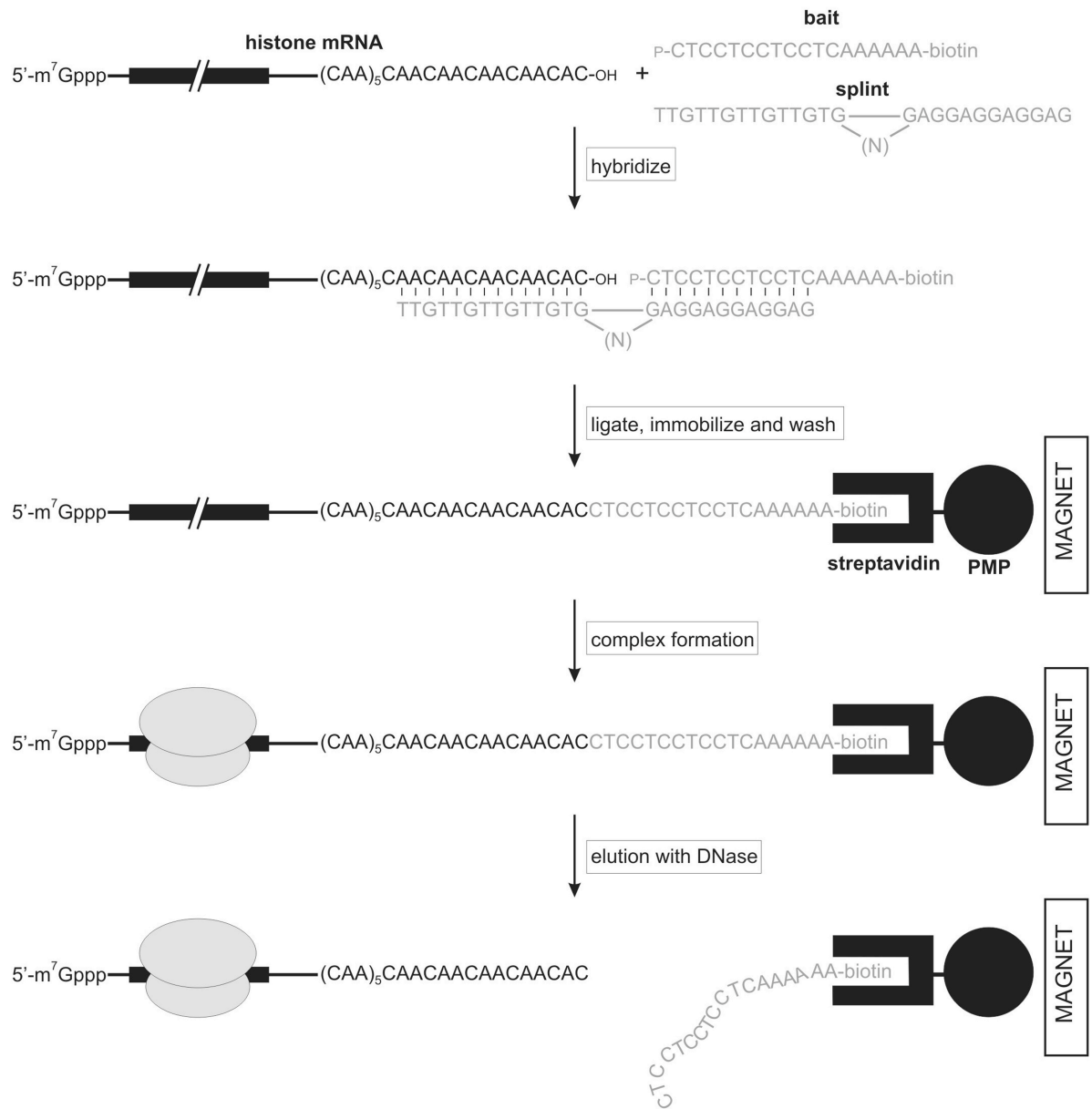


Figure 7

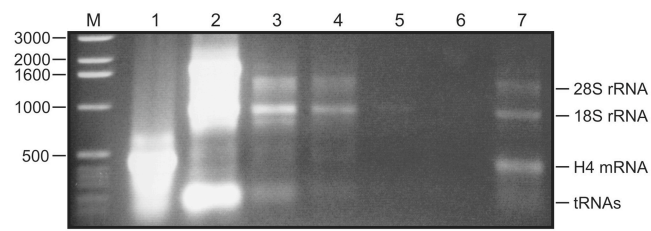




S2



S3



S4

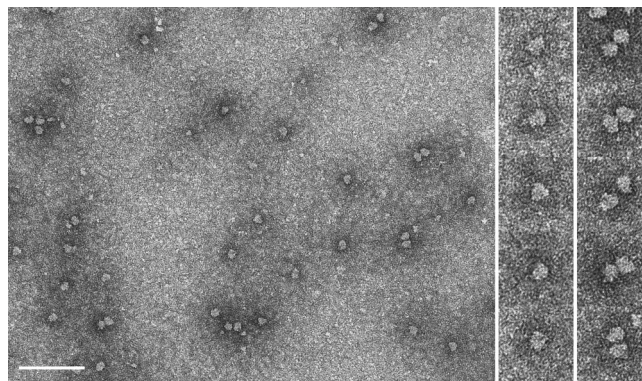
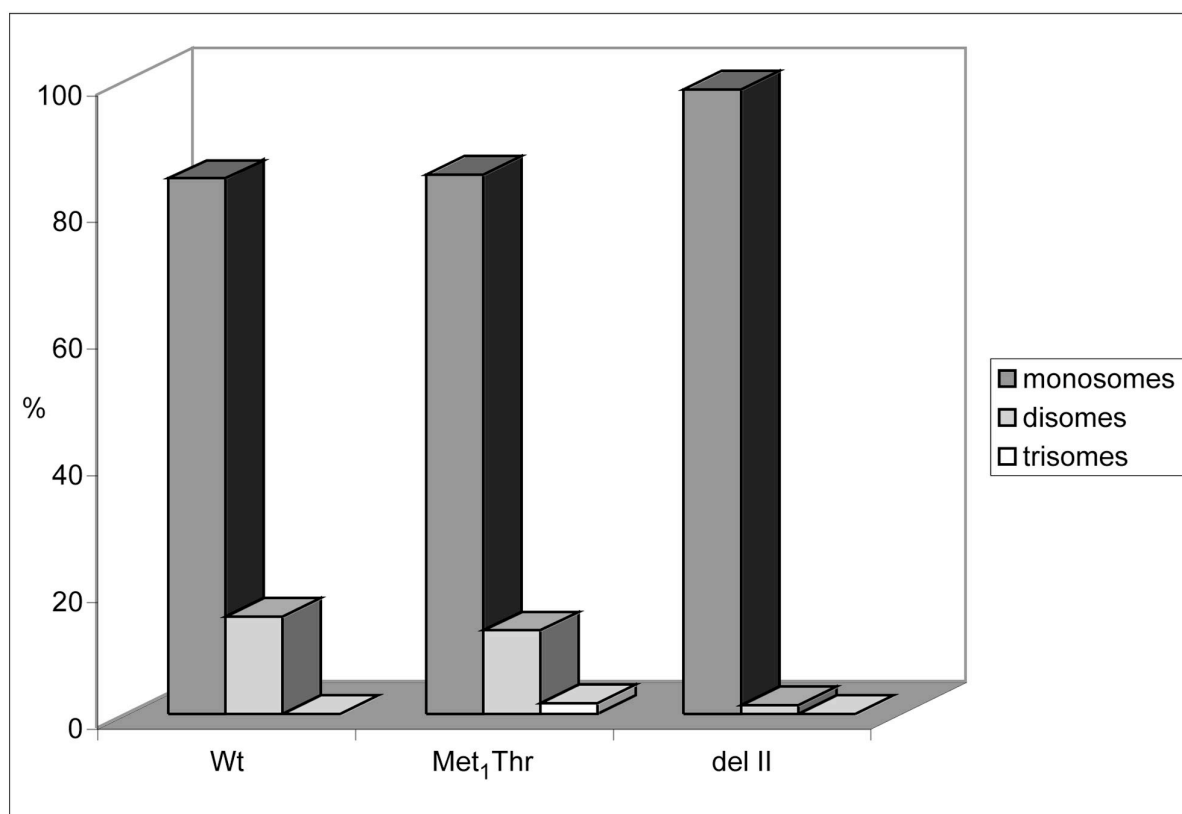


Table 1

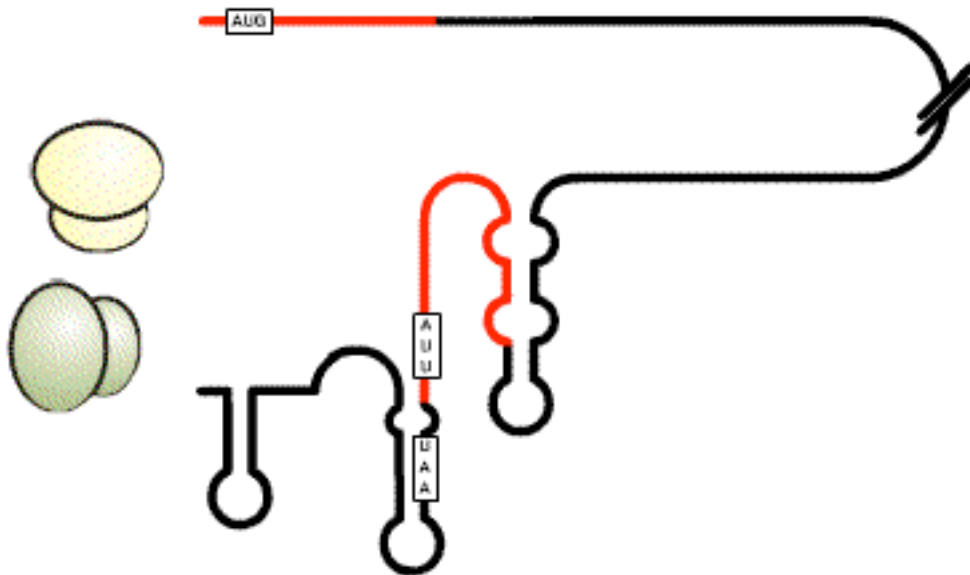
	Monosomes	Disomes	Trisomes
WT (1016 particles)	$76.5 \pm 1.3\%$	$18.8 \pm 1.2\%$	$4.7 \pm 0.7\%$
Met ₁ Thr (1078 particles)	$85.2 \pm 1.1\%$	$13.2 \pm 1.0\%$	$1.7 \pm 0.4\%$
Del II (1090 particles)	$92.5 \pm 0.8\%$	$7.1 \pm 0.8\%$	$0.4 \pm 0.2\%$

S5



S6

Both ribosomes bind the H4 mRNA ...



Copyright 2005 Roman Koning and Sharief Barends

Figure Legends

FIGURE 1. UTRs are dispensable for histone H4 mRNA translation

Schematic representations of capped, in-vitro transcribed histone H4-12 mRNAs are shown. Lengths of the UTRs and the coding region of wild-type mRNA are illustrated above transcript A and the lengths of the 3' UTRs are indicated for each transcript. Transcript F corresponds to pre-mRNA H4-12, and contains the Histone Downstream Element (HDE) at its 3' end. Physical stabilities of the transcripts have been assessed by semi-quantitative RT-PCR analysis as previously described (Li and Wang, 2004) in WGE after 0 and 90 min incubation at 25°C. Translation products formed in WGE and RRL are separated by 15% SDS-PAGE and [³⁵S]-labeled H4 was quantified by phosphor imaging. The graph represents relative expressions of H4 in WGE (black bars) and RRL (white bars). Error bars denote the standard deviation from the mean of three independent experiments.

FIGURE 2. H4-mRNA translation is cap dependent

In-vitro translation of wild-type H4 mRNA (transcript A) and BMV RNA in WGE supplemented with increasing concentrations of cap analogue (0.5, 2.5, 12.5, 60, 300 µM and 1.5 mM). Mean of relative expressions of two independent experiments is shown under the gels. H4 and coat protein (CP) and minimal inhibitory concentrations are indicated.

FIGURE 3. Optimal location of start codon

(A) Mutations are introduced in the 5'-terminal region (shown above the nucleotide sequence) and wild-type and mutated amino-acid sequences are shown under it. Insertion of the AUGA sequence in the 5' UTR generates an in-frame AUG codon at position -4. All RNAs were tested for in-vitro translation, separated by 15% SDS-PAGE (B) and quantified (C). The graph represents the ratio of relative expressions of transcripts containing the double mutation Met₁Thr-AUG_n divided by relative expressions of the corresponding transcripts containing AUG_n (n is codon position - 4, 1, 5, 7, 9 and 15). The value of codon 1 is the wild-type situation and its expression was set on 1. The values are corrected for the number of methionines incorporated in the in-vitro translated H4 protein. Error bars denote the standard deviation from the mean of three independent experiments.

FIGURE 4. H4 mRNA coding region is folded for ribosome recruitment

(A) In-vitro translation experiments of the primer walk is represented as relative expressions in WGE (black bars) and RRL (blank bars). Primers that lead to a reduction in translation more than two fold are encircled. (B) Chemical and enzymatic probing of the full-length H4-12 mRNA is plotted on the 2D model. Cleavages by chemical and enzymatic probes are detected by reverse transcription as previously described (Brunel and Romby, 2000) using primers complementary to sequences 40-60, 90-110, 140-160, 190-210, 241-260, 301-320, 338-356. Chemical compounds used are lead (single-stranded nts), DMS (accessible A and C) and CMCT (accessible U and G) and enzymatic probes are V1 (double-stranded nts), T1 (accessible G) and T2 (single-stranded nts). Three cut intensities are shown in the figure legend for each probe. Sequences 159-165 and 221-231 are underlined because they contain contradictory cuts. Sequence 293-322 is underlined because no cuts could be detected with any probe. Sequence 263-302 can also fold into an alternative structure fitting with our probing data and is boxed on the right. (C) To confirm the 5'-3' interaction, point mutations have been introduced in the putative base pairings. Mutations are represented on the 2D-model of the 5'-3' interaction model. The mutations are listed in the middle and the resulting amino-acid sequence is indicated between brackets. Reverse-transcription stops induced by primer 9 annealing are represented as black spots (big spots for strong stops and small spots for weaker stops). The bar graph shows relative expressions of H4 transcripts in WGE; error bars denote the standard deviation from the mean of three independent experiments. Alternative base pairing between the 5' and 3' part of the coding region for single mutant Gly₅Gly, Gly₉₅Ala, Arg₉₆Arg and Thr₉₇Thr are depicted.

FIGURE 5. More than one ribosome can bind to H4 mRNA ORF

Toe-print analysis was performed with two different deoxy oligonucleotides, on translation-initiation complexes of wild-type H4 mRNA (A) and Met₁Thr H4 mRNA (B) formed in RRL. Translation initiation was arrested with 2 mM GMP-PNP (lanes 1), 1 mg/ml cycloheximide (lanes 2), 2 mg/ml hygromycine (lanes 3) or 10 μ M edeine (lanes 4) and reverse transcriptase arrests were compared to those performed on H4 mRNAs in absence of RRL (lanes 5). Toe-print and full-length products are indicated on the right and sequencing ladders are shown on the left. (C) The toe-prints obtained from both wild type and Met₁Thr H4-mRNA initiation-complexes are mapped

on the 2D model of H4 mRNA. Strong toe-prints are black boxed and weak toe-prints are grey boxed. Asterisks indicate non-overlapping toe prints when comparing wild-type and Met₁Thr mRNAs. The A1 and U310 residues correspond to the first nucleotides of the start codon and stop codon, respectively. AUG codon, stop codon and AUU triplet are highlighted. The 3' end of deletion II mRNA is indicated at nucleotide 277.

FIGURE 6. H4 mRNA recruits two ribosomes

Typical negatively stained electron microscopy overviews (left) show ribosome recruitment by wild-type H4 mRNA (A) and deletion II mRNA (B). The majority of the particles with the wild-type H4 mRNA are monosomes (black circles and middle rows), while a substantial amount is disomes (white circles and right rows) and occasionally some trisomes (black square) are found. Deletion II mRNA forms mainly monosomes and sporadically disomes. Scale bar is 200 nm and ribosome gallery boxes are 113 nm. (C) At indicated time intervals in-vitro translation samples were separated by 15% SDS-PAGE, visualized and quantified by phosphor imaging. The graph represents the mean value of relative expressions of wild-type H4 mRNA (black triangles) and deletion II mRNA (black circles) in WGE and RRL.

FIGURE 7. A ribosome-channeling model for H4 mRNA translation

Ribosome 1 is recruited to the AUG codon via an IRES-like mechanism, while ribosome 2 is sequestered at the 3' end of the coding region. Ribosome 1 starts to translate the coding region enabling structural rearrangement of the AUG-containing sequence into an alternative structure. Ribosome 2 positions on the AUG codon while ribosome 1 terminates translation. After translation termination, ribosome 1 is sequestered at the 3' end of the coding region. The system is ready for a new round of translation. A movie illustrating our model is presented in supplemental Figure S6.

SUPPLEMENTAL DATA

Supplemental Figure 1. Ribosomal scanning for GFP translation

The bar graph represents the percentage of translation products in WGE and RRL from reporter GFP mRNA with a 25-nt 5' UTR (Long GFP) and without 5' UTR (Short GFP). Black bars represent the molar fraction of full-length GFP and white bars an out-of-frame peptide p7, which is translated from the second AUG codon.

Supplemental Figure 2. H4-mRNA-ribosome complex isolation

The histone mRNA was modified with a biotinylated deoxy oligonucleotide via the splint ligation method and complexes were purified by means of magnetic separation. RNA is shown in black whereas DNA is in grey. For more detail see 'Experimental Procedures'.

Supplemental Figure 3. Purification of translation-initiation complexes for electron microscopy.

Complexes were purified as described in 'Experimental Procedures' and a representative gel is shown. Lane 1 contains the modified H4 mRNA, which was used for the purification of ribosomal complexes. RNA isolated from the non-bound fraction (lane 2), the different washing steps [buffer 250 (lanes 3 and 4), buffer 500 (lane 5), buffer 50 (lane 6)] and from the eluted fraction (lane 7) was separated on a 1% agarose gel and visualized by ethidium-bromide staining. Several bands of the DNA ladder (Invitrogen) are highlighted and the positions of the ribosomal RNAs, tRNAs and H4 mRNA are indicated.

Supplemental Figure 4. Visualization of translation-initiation complexes by electron microscopy.

Typical negatively stained electron micrograph overview shows Met₁Thr mRNA in monosomes, disomes and rarely trisomes. Scale bar is 200 nm and ribosome gallery boxes are 113 nm.

Table 1: Statistical analysis of purified ribosome-mRNA particles

Ribosome-mRNA samples were freshly prepared for electron microscopy as described in 'Experimental procedure' section. From each sample about 1000 particles (mono-, di- and trisomes) were counted. Particles that were smaller than

ribosomes (ribosomal subunits) and particles that were not clearly di- or trisomes (aggregates), in total roughly 4% on top of the total amount of particles were not included. Complexes with wild-type and deletion-mutant II mRNAs have overestimated disome and trisome counts because of translation leakage: despite the presence of cycloheximide translation can start. Experiments with cycloheximide-treated samples followed by incubation at 4°C revealed at most 1% of translation leakage (Arava et al., 2003). In our purification procedure, cycloheximide is added followed by 5-minutes incubation at 30°C. Considering the fact that Met₁Thr mutant is not translated but can still efficiently recruit ribosomes as indicated by unchanged toe-prints, translation leakage in our purification conditions can be estimated by comparing the amount of disomes formed with the wild-type mRNA (18,8%) and the Met₁Thr mutant (13,2%). Therefore, translation leakage using our purification procedure is estimated to be 5,6%.

Supplemental Figure 5. Statistical analysis of purified ribosome-mRNA particles.

The graph represents the percentage of monosomes, disomes and trisomes formed with the wild-type mRNA, Met₁Thr and del II mutant mRNA visualized by electron microscopy. Taking into account the translation leakage, values for wild type and deletion II mutant have been corrected whereas Met₁Thr values are not corrected because of lack of translation.

Supplemental Figure 6. Movie summarizing the major steps of our H4 mRNA translation model.

CONCLUSION ET PERSPECTIVES

L'objectif de ce travail de thèse a été d'étudier deux étapes clé et atypiques de l'expression des gènes d'histones de type réplication-dépendants. La première concerne la maturation de l'extrémité 3' non traduite des ARN pré-messagers, la seconde la traduction des ARNm alors produits. Ceci nous a amené à décrypter les modalités de reconnaissance de la tige-boucle des ARN pré-messagers d'histones par la protéine HBP et à étudier la structure en solution de l'ARN pré-messager. Nous avons pu mettre en évidence des changements structuraux induits par la fixation de la protéine HBP sur le précurseur d'ARN. Enfin, l'étude de la traduction des ARNm d'histones nous a conduits à découvrir un mode de traduction tout à fait original et très efficace dont nous avons étudié le mécanisme.

Les domaines N- et C-terminaux de la protéine HBP contribuent à la spécificité d'interaction avec la structure en tige-boucle

L'objectif de cette première étude a été d'identifier des résidus de la protéine HBP participant à la fixation à l'ARNm. En l'absence de données structurales sur la protéine libre ou en complexe avec l'ARN, nous avons choisi une approche par mutagenèse aléatoire couplée à une sélection génétique de mutants par la technique du triple hybride dans la levure. Possédant déjà un certain nombre de mutants d'HBP ne fixant plus l'ARN, nous avons choisi de contre-sélectionner des mutations capables de compenser cet effet négatif. Ce type d'approche permet de mettre en évidence les inter-relations fonctionnelles entre différents domaines d'une protéine. Sept mutations ont ainsi été identifiées pour leur aptitude à abolir l'effet des premières mutations. Les doubles mutants correspondants ont été extraits des levures et leur phénotype a pu être confirmé *in vitro*. L'ensemble des mutations caractérisées, à l'exception d'une seule, étaient localisées dans les domaines N- et C-terminaux de la protéine qui flanquent le domaine principal de liaison à l'ARN appelé RBD. Elles permettaient de restaurer l'interaction avec l'ARN sans perte de spécificité pour la tige-boucle

sauvage. Les résultats obtenus suggèrent l'existence de connexions fonctionnelles entre le RBD et les domaines externes. Ces derniers participeraient à la mise en conformation du RBD de manière à améliorer sa spécificité de reconnaissance de la structure en tige-boucle sauvage.

La fixation de la protéine HBP sur certains ARNm d'histones induit des changements de conformation au niveau du « HDE » qui améliorent l'hybridation du snRNA U7

La précédente étude avait pour but d'identifier les résidus d'HBP impliqués dans l'interaction avec la structure typique en tige-boucle présente à l'extrémité 3' des ARNm d'histones. Dans un second temps, nous avons voulu mesurer l'effet de la fixation d'HBP sur la structure de l'extrémité 3' non traduite et non maturée de divers ARN pré-messagers d'histones. Nous avons sondé la structure des ARNs en solution, en examinant plus particulièrement la séquence d'ancrage de la snRNP U7 appelée « Histone Downstream Element » (ou HDE). Nous avons montré que les séquences HDE des ARN pré-messagers des histones H4-12, H2a-614 et H1t n'étaient pas libres mais impliquées dans diverses structures secondaires pouvant les rendre inaccessibles à la snRNP U7. Dans le cas des ARNs H4-12 et H1t, la fixation de la protéine HBP induit des variations significatives de l'accessibilité aux sondes au niveau du HDE. Ces variations pourraient révéler des changements de conformation de cette région. Enfin, nous avons pu montrer que ces changements de conformation sont associés à une amélioration de l'ancrage d'un transcrit de snRNA U7 à l'élément HDE. Cependant, ce mécanisme n'est pas généralisable à l'ensemble des gènes d'histones puisque aucune modification importante n'a pu être détectée à l'extrémité 3' non traduite d'un troisième gène d'histone (H2a-614).

À l'issue de cette étude, nous proposons que la protéine HBP, en se fixant à la structure en tige-boucle de l'ARN pré-messager, induirait des changements de conformation en aval qui amorceraient la libération de la séquence HDE. Celle-ci deviendrait alors plus accessible à l'ancrage de la snRNP U7.

La phase codante de l'ARNm H4-12 porte l'information structurale responsable de son mode de traduction atypique par recrutement interne de ribosomes

Dans le dernier volet de notre travail, nous avons analysé par sondage en solution la structure secondaire de l'ARNm de l'histone H4-12. Notre but était de comprendre par quels mécanismes cet ARNm pouvait être traduit de manière aussi efficace et originale. En effet,

nous avons montré que les séquences 5' et 3' non traduites ne sont pas requises pour la traduction, l'initiation de la traduction se fait sans le mécanisme habituel de glissement du ribosome ou « scanning », et la protéine HBP n'intervient pas de façon significative dans le processus de traduction. La coiffe au niveau de l'extrémité 5' est toutefois requise. Les sondages en solution ont permis de proposer un modèle de structure secondaire de l'ARNm où la séquence codante forme une structure centrale complexe qui est circularisée par appariement des extrémités 5' et 3' de la phase codante. Les régions 5' et 3' non traduites sont quant à elles libres et non appariées, ce qui semble confirmer leur rôle négligeable dans la traduction. D'autres sondages en solution réalisés sur des molécules bloquées dans l'initiation de la traduction par des petits ARN anti-sens nous ont permis d'identifier des résidus impliqués dans la traduction.

L'ensemble des résultats obtenus dans nos expériences de sondage en solution associés à ceux issus d'études de traduction *in vitro*, de « toe print » et de microscopie électronique, ont conduit à l'établissement d'un modèle original permettant d'expliquer la traduction atypique de l'ARNm H4-12. La phase codante recruterait directement deux ribosomes à la manière de 2 sites d'entrée interne du ribosome ou « IRES ». Le premier ribosome serait recruté au niveau du codon d'initiation, le second près de la fin de la phase codante. Par le jeu de changements de conformation et grâce à la circularisation de l'ARNm, le deuxième ribosome pourrait être dirigé très efficacement sur le codon d'initiation. Cet enchaînement des deux ribosomes sur la phase codante permettrait d'expliquer d'une part, la grande efficacité de traduction observée dans le cas de notre modèle H4-12 et d'autre part, le rôle accessoire des séquences non traduites.

Perspectives

Ci-dessous, et dans la continuité des travaux qui viennent d'être exposés, nous avons considéré un certain nombre d'expériences et de projets qui pourraient à la fois compléter ces travaux et apporter une ouverture sur de nouvelles études à mener.

Dans la suite des travaux portant sur l'interaction entre la protéine HBP et la structure en tige-boucle des ARNm d'histones, il devient nécessaire, pour progresser, de résoudre la structure tridimensionnelle de la protéine ainsi que celle du complexe. Cela suppose de résoudre les problèmes d'expression de la protéine HBP humaine ou de souris.

Concernant la maturation en 3' des pré-ARNm d'histones, il serait intéressant de compléter notre étude par une analyse de la structure en solution des extrémités 3' non traduites des gènes codant pour les histones H2b et H3 afin d'avoir une vision plus complète pour les 5 histones majeures. Par ailleurs, le gène H3-53 est très dépendant de la protéine HBP pour sa maturation malgré une forte complémentarité avec le snRNA U7, il serait donc intéressant d'examiner les modifications structurales que la protéine HBP pourrait induire sur cet ARNm.

Nous envisageons aussi d'utiliser les approches de sondage en solution, pour déterminer l'empreinte des protéines Lsm10 et Lsm11 spécifiques à la snRNP U7 sur l'extrémité 3' non traduite d'un ARN pré-messager d'histone. Nous utiliserons à cette fin des molécules chimères constituées d'une fusion entre l'extrémité 3' non codante de l'ARN pré-messager H4-12 et le snRNA U7, les deux parties étant reliées par une séquence « linker ». Une fois injectées dans le cytoplasme des ovocytes de xénopes, ces molécules s'associent aux protéines Sm et Lsm endogènes afin de former des particules ribonucléiques mimant un système de maturation fonctionnel. Il sera ainsi possible de comparer l'empreinte d'une particule contenant les protéines Lsm10 et Lsm11 spécifiques à U7 à celle obtenue avec des particules mutantes dépourvues de protéines Lsm (obtenues par mutation des « Sm-binding sites »). Enfin, nous pourrions déterminer si l'extension N-terminale de la protéine Lsm11 qui est impliquée dans la maturation est effectivement retrouvée à proximité du site de clivage de l'ARN pré-messager.

Bien entendu, du fait de son originalité, la poursuite de l'étude du mécanisme atypique de traduction est prioritaire. De nombreuses questions tant structurales que fonctionnelles restent en suspens. Comment et sur quelles structures sont recrutés les ribosomes ? Quels sont les facteurs d'initiation nécessaires, et existe-t-il des facteurs spécifiques ?

D'un point de vue structural, il conviendrait de poursuivre l'étude en solution sur des ARNm mutés ayant perdu leur activité ou l'ayant retrouvée par mutation compensatoire. Cela permettrait de confirmer le modèle structural proposé. De plus, il serait intéressant de réaliser des sondages chimiques par le DMS et la CMCT sur les ARNm mutés ou hybridés aux petits ARNs inhibiteurs. Ces sondes sont plus sensibles aux changements structuraux, éventuellement tertiaires, que les précédentes sondes utilisées qui n'avaient révélé aucun changement de structure secondaire.

Nous envisageons aussi de déterminer les empreintes des ribosomes sur l'ARNm H4-12. L'utilisation d'une petite sonde comme le plomb pourrait être adaptée à la mise en évidence des régions de l'ARN en contact avec le ou les ribosomes et pour détecter les éventuels changements de conformation induits par leur fixation.

Dans le même esprit, nous pourrions réaliser des tests de traduction de molécules contenant des délétions internes. Ces expériences s'avèreraient utiles pour déterminer la structure ou la séquence nécessaire au recrutement du second ribosome en 3'. Nous poursuivons également la collecte d'images par cryo-microscopie électronique afin d'obtenir assez d'informations pour reconstituer un modèle tridimensionnel des complexes d'initiation formés sur les ARNm H4-12. Ces résultats nous permettront de déterminer comment s'effectue le recrutement des ribosomes, mais surtout s'il s'effectue de manière similaire à celui des IRES viraux.

Il serait bien entendu d'un très grand intérêt de résoudre la structure cristalline de l'ARNm. Sa grande richesse en résidus GC pourrait être un gage de grande stabilité. Nous avons montré qu'il existe 2 conformations différentes de l'ARNm et qu'elles étaient stables une fois séparées. La structure tridimensionnelle de ces différentes molécules pourrait illustrer les différents états du modèle proposé par notre étude. L'avantage de disposer à la fois des données cristallographiques et de cryo-microscopie électronique pourrait être considérable pour la compréhension détaillée du mécanisme de traduction.

Enfin, il nous semble incontournable d'étudier plus en détail le mécanisme d'initiation de la traduction pour identifier les facteurs requis dans ce mécanisme non conventionnel.

Certains IRES requièrent des facteurs pour recruter le ribosome et initier la traduction, d'autres s'en dispensent. Le fait que la traduction de l'ARNm H4-12 soit dépendante de la présence d'une coiffe en 5' semble suggérer que certains facteurs, classiques ou spécifiques, seraient tout de même nécessaires à la traduction. En premier lieu, nous envisageons de tester l'implication de certains facteurs d'initiation de la traduction classiques en supprimant leur activité dans des extraits de traduction *in vitro*. Trois facteurs courants peuvent être facilement testés. Il existe un mutant dominant négatif du facteur d'initiation eIF4A dont l'effet sur la traduction de l'histone H4 sera aisément mesurable. Rappelons que ce facteur porte l'activité « hélicase » nécessaire au « scanning » de l'ARNm par le ribosome. Il est également possible de cliver le facteur eIF4G par la protéase 2A, ce qui le rend incapable de jouer son rôle de plate-forme d'assemblage de facteurs réunissant la coiffe et la petite sous-unité du ribosome. Un traitement des extraits de traduction par la protéase permettra d'évaluer la contribution du facteur eIF4G sur la traduction. Nous pourrions aussi dépléter les extraits de traduction du facteur eIF4E par un excès de coiffe fixée sur des billes de sépharose.

Nous envisageons aussi de tester de façon plus systématique chacun des facteurs d'initiation par ARN interférence dans des cultures de cellules HeLa. Une version de l'ARNm H4-12 comportant un « tag » HA sera utilisée dans ces expériences de traduction.

La traduction atypique de l'ARNm H4-12 pourrait également faire intervenir des facteurs spécifiques. Nous pourrions les caractériser par spectrométrie de masse des complexes d'initiation purifiés.

Une fois le mécanisme de traduction de l'ARNm H4-12 élucidé, il nous semble impératif de déterminer si le phénomène est généralisable en étendant les études de traduction aux ARNm codant les autres histones.

À plus long terme, l'étude de la traduction d'autres ARNm « atypiques » pourrait être envisagée afin de rechercher et caractériser d'autres modes de traduction hors normes. Les ARNm du protozoaire *Giardia lamblia* ont des particularités structurales semblables aux ARNm d'histones. La taille réduite, voir même l'absence d'extrémité 5' non traduite suggère à nouveau l'emploi d'un mécanisme de traduction particulier. Son étude permettra de mieux comprendre l'évolution de l'appareil de traduction.

ANNEXE

MATÉRIEL ET MÉTHODES

1 ORGANISMES ET VECTEURS UTILISÉS

1.1 Souche de *Saccharomyces cerevisiae*

L40-coat : *MATa*, *ura3-52*, *leu2-3112*, *his3Δ200*, *trp1Δ1*, *ade2*, *LYS2::(lexAop)-HIS3*, *ura3::(lexAop)-LacZ*, *LexA-MS2 coat (TRP1)* (SenGupta *et al.*, 1996). Cette souche a été utilisée dans la méthode du triple hybride. Elle contient dans son génome le gène codant pour l'hybride protéique « LexA-MS2 coat » ainsi que les gènes rapporteurs *LacZ* et *HIS3*. Ces derniers sont inductibles par le transactivateur Gal4.

1.2 Souche d'*Escherichia coli*

TB1 : *F⁻ ara Δ(lac-proAB) hsdR (r_k⁻ m_k⁺) rpsL(Str^r) [φ80 dlacΔ(lacZ)M15] thi*. Cette souche a servi de réceptrice pour les transformations par électroporation, elle a également permis d'obtenir de l'ADN plasmidique.

1.3 Vecteurs

pIII/MS2-2 : ce plasmide navette multicopie a été utilisé dans la technique du triple hybride, pour exprimer l'hybride ARN contenant le site de fixation de la protéine de capside du phage MS2 suivi de l'ARN étudié. Cette molécule est transcrite à partir du promoteur fort de l'ARN polymérase III de l'ARN de la RNase P. Le plasmide contient le marqueur

d'auxotrophie URA3 ainsi que le gène de résistance à l'ampicilline, permettant son maintien dans la levure et *E. coli* respectivement. Nous avons utilisé les constructions pIII/HP-MS2 où la séquence correspondant à la tige-boucle ou « hairpin » (HP) à tester est clonée dans la cassette du vecteur. Ainsi, le plasmide pIII/wtHP-MS2 permettra d'exprimer, dans le contexte du triple hybride, un hybride ARN contenant la tige-boucle sauvage du messager H4-12 humain codant pour l'histone H4 (**Figure 26**).

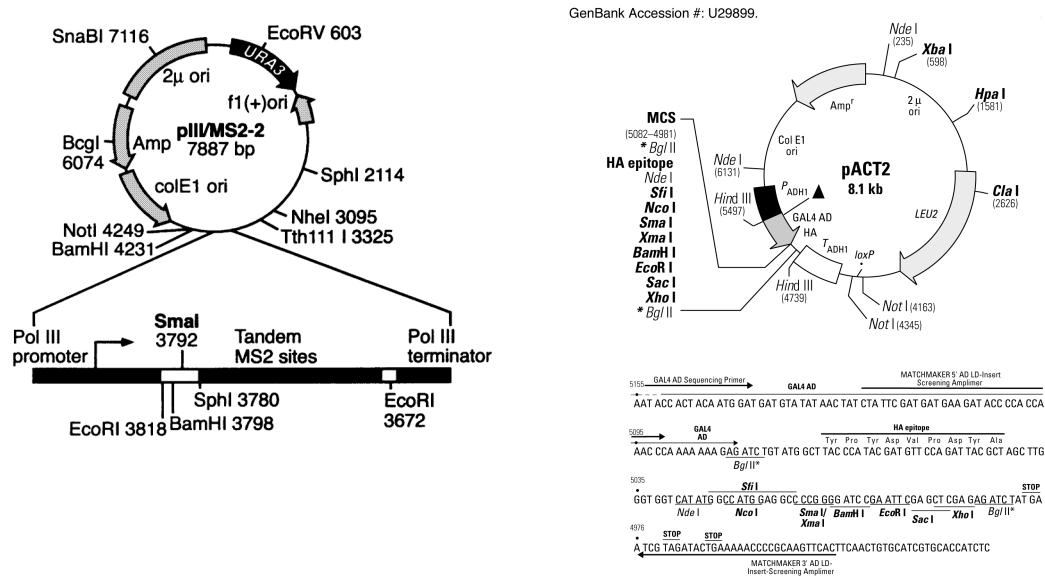


Figure 26 : Cartes de restriction des plasmides couramment utilisés.
Plasmide pIII/MS2-2 ; Plasmide pACT2.

pACT : ce plasmide navette multicopie permet l'expression, à partir du promoteur de l'alcool déshydrogénase, d'une protéine d'intérêt fusionnée au domaine activateur de Gal4 intervenant dans le triple hybride. Il contient en outre un gène de résistance à l'ampicilline pour sa sélection dans *E. coli* et un marqueur d'auxotrophie *LEU2* pour son maintien dans la levure. Nous y avons inséré le gène de la protéine HBP humaine, sauvage ou mutante, en phase avec celui du domaine d'activation de Gal4, de manière à obtenir des constructions de type pAct-HBP codant pour les protéines de fusion correspondantes.

PACT2 : ce plasmide, construit par Steve Elledge et collaborateurs, est une version plus sophistiquée de pACT. Il dispose d'un plus grand nombre de sites de restriction au sein de sa cassette de clonage, ainsi que d'un marqueur contenant l'épitope HA (**Figure 26**).

pUC19 : ce petit plasmide multicopie est un vecteur utilisé pour le clonage dans *E. coli*. Il fait partie d'une famille de vecteurs construits par Messing et collaborateurs (Yanisch-Perron *et al.*, 1985) dérivés du plasmide pBR322. Ils en portent l'origine de répllication, le gène de résistance à l'ampicilline ainsi que la partie du gène *LacZ* codant pour le fragment α de la β -galactosidase, au début duquel a été insérée une cassette de clonage ou « polylinker » qui offre des possibilités variées de clonage de fragments d'ADN. L'insertion d'un fragment d'ADN lors d'un clonage pourra être vérifiée par α complémentation. Nous avons utilisé le vecteur pUC19 afin de construire des matrices de transcription en y clonant au niveau du site *SmaI/XmaI* la séquence du promoteur de la polymérase du phage T7 suivie de la séquence correspondant à la région 3' non codante des différents ARN pré-messagers d'histones testés. Ces dernières auront, au besoin, subi de légères modifications au niveau des extrémités des fragments clonés de manière à ne reconstituer le site de restriction que du côté opposé au promoteur afin de transcrire par la méthode « run-off » après digestion du plasmide à l'enzyme *SmaI* laissant des extrémités franches.

2 CULTURE, TRANSFORMATION ET MANIPULATION DES SOUCHES DE LEVURES ET DE BACTÉRIES

2.1 Culture des cellules

Les milieux sont stérilisés par autoclavage à 120°C pendant 20 min. Les milieux gélosés sont additionnés de 20 g de bactoagar pour un litre.

Les souches d'*E. coli* sont cultivées à 37°C dans le milieu LB (Luria Broth) contenant, par litre, 10 g de bactotryptone, 5 g de NaCl, 5 g d'extrait de levure et 1 mL de NaOH 1M. Le milieu peut être complémenté de 100 mg/L d'ampicilline après autoclavage.

Les souches de *S. cerevisiae* sont cultivées dans différents milieux à 30°C. Le milieu complet YPG contient 10 g de « bacto-yeast extract », 20 g de « bactopectone » et 20 g de glucose par litre. Le milieu minimum YNB contient 7 g de « Yeast Nitrogen Base » et 20 g de glucose pour 1 litre. On y ajoute avant stérilisation 100 mL d'une solution d'éléments essentiels (acides aminés et bases) dont la composition est indiquée dans le tableau ci-dessous.

La nomenclature employée indique, pour les milieux minimums sélectifs, les éléments essentiels manquants qui permettront la sélection des marqueurs d'auxotrophie correspondants. Par exemple, le milieu YNB-Ura-Leu est un milieu minimum préparé avec la solution décrite ci-dessus dans laquelle l'uracile et la leucine sont absents ; il permet de sélectionner les clones exprimant les marqueurs d'auxotrophie *LEU2* et *URA3* contenus dans les vecteurs pACT et pIII/MS2-2 respectivement.

Éléments essentiels	Concentration (µg/mL)
Adénine	400
Arginine	200
Histidine (His)	200
Leucine (Leu)	1000
Lysine	300
Méthionine	200
Phénylalanine	500
Thréonine	2000
Tyrosine	300
Uracile (Ura)	200

Le milieu YNB-Ura-Leu+XGal utilisé pour tester l'expression de *LacZ* est un milieu YNB-Ura-Leu supplémenté en XGal (5-bromo-4-chloro-3-indoyl-β-D-galactoside) dont le pH a été ajusté à 7. Après autoclavage, on ajoute à 900 mL de milieu refroidi à 65°C, 100 mL de tampon BU ×10 pH 7 stérile (261 mM Na₂HPO₄, 250 mM NaH₂PO₄) et 4 mL d'XGal (dissout dans la diméthyl-formamide à 20 mg/mL).

Le milieu YNB-Ura-Leu-His+3AT permet de mesurer l'activité du rapporteur *HIS3*. Il contient du 3AT (3-aminotriazole), un inhibiteur des voies métaboliques de l'histidine. Le 3AT est ajouté après autoclavage au milieu refroidi. Le niveau d'expression du rapporteur *HIS3* peut-être évalué par des tests à des concentrations croissantes en 3AT dans le milieu de culture. Plus l'expression du gène *HIS3* sera élevée, plus grande sera la concentration de 3AT nécessaire pour inhiber la croissance de la souche de levure.

Le milieu YNB-Leu+5FOA est utilisé pour isoler le plasmide pACT ou pACT2 et ceci par la sélection des clones prototrophes pour la leucine ayant perdu le plasmide pIII/HP-MS2 et par conséquent le marqueur *URA3*. Il contient de l'uracile à 25 mg/mL et 0,7 g/L de 5-fluoro-orotate (5FOA), un composé dont la métabolisation par les cellules exprimant le marqueur *URA3* entraîne la formation de dérivés toxiques létaux.

2.2 Transformation d'*E. coli* par électroporation

Nous avons suivi le protocole décrit précédemment (Dower *et al.*, 1988). Une culture de la souche réceptrice est incubée à 37°C jusqu'à une absorbance de 0,8 à 700 nm. L'ensemble de la préparation s'effectue par la suite à 0°C. Le culot de cellules centrifugées 5 min à 4000 × g est lavé deux fois à l'eau stérile froide (par 1 puis 0,5 volume de la culture initiale), puis par du glycérol 10% (v/v) (1/25^e volume). Finalement, les cellules sont reprises par 1/200^e du volume initial de culture dans du glycérol 10%. La transformation est réalisée en mettant en présence 40 µL de cellules compétentes avec 1 pg d'ADN. Après électroporation (2,5 kV, 200 Ω, 25 µF), les cellules sont additionnées de 90 µL de LB puis étalées sur milieu sélectif et incubées une nuit à 37°C.

2.3 Transformation de *S. cerevisiae*

2.3.1 Transformation à l'acétate de lithium

Ce protocole est dérivé de (Gietz *et al.*, 1992). Une culture de levure de 100 mL ($A_{700nm} = 0,8-1$) est centrifugée 2 min à 1000 × g, puis les cellules sont reprises dans 1 mL d'une solution d'acétate de lithium 100 mM. Dans un tube stérile, 50 µL de ces cellules compétentes sont mélangés à 1 µg d'ADN transformant, 5 µL d'une solution à 10 µg/µL d'ADN entraîneur de sperme de saumon préalablement « soniqué » et bouilli, 69 µL d'eau stérile, 36 µL d'une solution d'acétate de lithium 1 M et 240 µL de 50% PEG 4000 (p/v). Les tubes sont incubés à 42°C pendant 20 min. Après une brève centrifugation, les cellules sont reprises par 100 µL d'eau stérile et étalées sur milieu sélectif. Les boîtes de culture sont incubées 2 ou 3 jours à 30°C.

2.3.2 Transformation à haute efficacité

Cent mL d'une culture de levure ($A_{700nm} = 0,8-0,9$) sont centrifugés 2 min à 1000 × g puis les cellules sont lavées dans 20 mL d'eau stérile, puis 10 mL d'une solution LiAc-TE (1 M LiAc, 10 mM Tris-HCl pH 7,5, 1 mM EDTA), et enfin resuspendues dans 250 µL de la même solution. Dans un tube stérile, 50 µL de ces cellules compétentes sont mélangés à 1 µg d'ADN transformant et 50 µg d'ADN entraîneur de sperme de saumon préalablement « soniqué » et bouilli. Enfin, 300 µL d'une solution PEG-LiAc-TE (50% PEG 4000 (p/v), 1 M LiAc, 10 mM Tris-HCl pH7,5, 1 mM EDTA) sont ajoutés. L'ensemble est ensuite incubé à

30°C durant 2 heures, puis les cellules sont soumises à un choc thermique par incubation à 42°C durant 20 min et refroidissement de 15 secondes à 0°C. Une brève centrifugation permettra de les sédimenter avant de les reprendre dans 1 mL d'eau stérile. Cent µL de cette solution sont ensuite étalés par boîte de pétri contenant le milieu sélectif. L'incubation se fait à 30°C pendant 2 ou 3 jours (Gietz *et al.*, 1992).

2.4 Utilisation du triple hybride

La technique dite du « triple hybride » permet d'étudier l'interaction entre un ARN et une protéine dans la levure *S. cerevisiae* (**Figure 18**) (SenGupta *et al.*, 1996). Cette technique consiste à co-exprimer trois molécules hybrides dans la souche L40-coat : la première, de nature protéique, est constituée par le domaine de liaison à l'ADN du répresseur LexA d'*E. coli* fusionné à la protéine de capsid du phage MS2. Cette dernière a la particularité de se fixer sur l'ARN génomique du phage au niveau d'une séquence de 22 nucléotides. Le second hybride est un ARN comportant ce site de fixation de la protéine de capsid suivi de la molécule X à tester. Enfin, le troisième hybride consiste en une protéine de fusion entre la protéine analysée Y et le domaine d'activation de la transcription du transactivateur Gal4 de levure. L'interaction entre l'ARN X et la protéine Y va permettre de reconstituer un transactivateur fonctionnel qui va recruter la machinerie de transcription en amont de 2 gènes rapporteurs *HIS3* et *LacZ*.

Protocole expérimental :

Dans une première étape, la souche L40-coat est transformée par le vecteur pIII/wtHP-MS2 exprimant l'hybride ARN contenant la tige-boucle sauvage de l'ARN messager H4-12. La souche ainsi obtenue exprime les deux premiers hybrides : l'hybride « LexA-MS2 coat » d'origine chromosomique (voir génotype au § 1.1) et l'hybride ARN exprimé à partir de pIII/wtHP-MS2. Pour constituer le triple hybride fonctionnel le plasmide pAct-HBP codant pour la protéine HBP (ou ses variants lors d'une sélection) fusionnée au domaine d'activation de Gal4, a été introduit dans la souche L40-coat par une deuxième transformation effectuée par la méthode à haute efficacité (§ 2.3.2). Les cellules L40-coat doublement transformées sont étalées sur boîtes de YNB-Ura-Leu-His+3AT. Sur ce milieu, seuls pousseront les clones contenant un mutant de HBP capable de reconnaître la tige-boucle, et donc capables d'activer le gène rapporteur *HIS3*. Les colonies [His⁺] peuvent ensuite être repiquées sur milieu YNB-

Ura-Leu+XGal, afin de vérifier par l'apparition de la coloration bleue, l'activation de la transcription de l'autre gène rapporteur *LacZ*.

2.5 Test en gouttes

Afin de mesurer le niveau d'expression du gène rapporteur HIS3 dans le triple hybride au sein de la cellule de levure, nous avons utilisé une technique qualitative de test en gouttes. Des gouttes de 4 µL d'une suspension de cellules à tester ($A_{700nm}=0,1$) sont déposées sur des boîtes de milieu sélectif YNB-Ura-Leu-His+3AT. Le test est réalisé sur une gamme de concentrations en 3AT variant de 0 à 100 mM, ce qui permet d'estimer le niveau d'expression de HIS3 (voir milieu 3AT § 2.1.) et, par extrapolation, l'affinité de la protéine pour l'ARN. Les boîtes sont incubées 3 jours à 30°C.

3 TAMPONS ET SOLUTIONS COURAMMENT UTILISÉS

Tampon TE : 10 mM Tris-HCl pH 8
 1 mM EDTA

Tampon TAE (×10) : 40 mM Tris-base
 5 mM acétate de sodium
 1 mM EDTA
 le tout ajusté à pH 7,9 par de l'acide acétique

Tampon TBE (×10) : 890 mM Tris-Base
 890 mM acide borique
 25 mM EDTA pH 8,3

Solution de dépôt « Formamide Dye » :
 20 mM EDTA
 Formamide 95% (v/v)
 bleu de bromophénol 0,1% (p/v)
 xylène cyanol 0,1% (p/v)

4 MÉTHODES RELATIVES À L'ÉTUDE DE L'ADN

4.1 Techniques générales de manipulation de l'ADN

4.1.1 Détermination de la concentration de l'ADN

La concentration de l'ADN bicaténaire en solution est déterminée par mesure de l'absorbance à 260 nm. Une unité $A_{260\text{nm}}$ correspond à 50 μg d'ADN double brin.

4.1.2 Précipitation de l'ADN

Cette étape permet de concentrer l'ADN et d'éliminer les substances solubles dans l'éthanol. L'ADN est précipité par addition de 3 volumes d'éthanol à la solution rendue 0,25 M final en NaCl. Après 5 min à -80°C la solution est centrifugée 10 min à $12000 \times g$, puis le culot d'ADN est lavé par de l'éthanol 80%, séché et dissout dans de l'eau milliQ.

4.1.3 Extraction des protéines au phénol

Le phénol est un agent qui dénature les protéines. Un volume de phénol ou de phénol/chloroforme saturé en tampon TE est ajouté à la solution aqueuse à extraire, et le tout est agité vigoureusement. Le tube est ensuite centrifugé 2 min, puis la phase aqueuse supérieure est soigneusement prélevée en évitant d'entraîner les protéines dénaturées situées à l'interface phénol/solution aqueuse.

4.1.4 Digestion et modification enzymatique de l'ADN

4.1.4.1 Hydrolyse par les enzymes de restriction

La digestion est conduite durant une heure à 37°C sur 0,5 à 10 μg d'ADN plasmidique additionné d'1 à 5 U d'enzyme par μg d'ADN dans un tampon préconisé par le fournisseur, durant une heure à 37°C . La digestion est suivie d'une électrophorèse sur gel d'agarose.

4.1.4.2 Déphosphorylation des extrémités 5' des vecteurs linéarisés

Cette étape empêche la recircularisation des vecteurs linéarisés lors de l'étape de ligation. Deux μg d'ADN sont traités par 2 U de phosphatase alcaline bactérienne dans le tampon préconisé par le fabricant durant 2 heures à 65°C . L'enzyme est éliminée par

extraction au phénol, suivie d'une extraction au phénol/chloroforme et d'une précipitation à l'éthanol.

4.1.4.3 Ligation de fragments d'ADN

Dix à 50 ng de vecteurs linéarisés et déphosphorylés sont incubés avec le fragment possédant les extrémités compatibles, dans un rapport vecteur/fragment=1/4, en présence d'une U de T4 ADN ligase dans un tampon fourni par le fabricant. Le milieu est incubé 12 heures à 16°C. Deux µL de ce mélange peuvent être utilisés directement pour transformer la souche bactérienne TB1 par électroporation.

4.1.5 Séparation et purification de fragments d'ADN

4.1.5.1 Séparation sur gel d'agarose

Cette technique permet de séparer des fragments d'ADN entre 0,2 et 30 kpb. Le gel est préparé à partir de 100 mL d'agarose 1% (p/v) dans le tampon TAE. Dix µL d'échantillon additionnés de 4 µL de solution de dépôt (glycérol 20% (v/v), EDTA 25 mM, sarkosyl 2% (p/v), bleu de bromophénol et xylène cyanol 0,3% (p/v)) sont déposés dans les poches du gel. L'électrophorèse est conduite sous 70 V dans le tampon TAE, puis l'ADN est visualisé sous UV après immersion du gel dans une solution de bromure d'éthidium. La taille des fragments est estimée par comparaison avec une échelle de poids moléculaires.

4.1.5.2 Éluion des fragments d'ADN séparés sur gel d'agarose

Le fragment d'intérêt, séparé sur gel d'agarose, est visualisé sous UV, excisé et filtré sur laine de verre dans un microtube par une centrifugation de 5 min à 12000 × g. Le filtrat est traité au phénol puis au phénol/chloroforme, avant d'être soumis à une précipitation à l'éthanol. Le culot est repris dans 40 µL d'eau milliQ.

4.1.6 Séquençage automatique

Les réactions de séquençage ont été réalisées à l'aide de l'AMPLITAQ FS polymérase en utilisant la solution de terminaison BIGDYE™ TERMINATORS (Applied Biosystems). Les séquences obtenues ont été analysées par un séquenceur de type APPLIED BIOSYSTEM 373. Ces travaux ont été effectués à l'IBMP par M. Philippe Hammann.

4.2 Extraction et purification de l'ADN

4.2.1 Minipréparation d'ADN plasmidique par la méthode de la lyse alcaline

Nous avons utilisé la méthode de préparation d'ADN mise au point par Birnboim et Doly (1979). Elle permet d'obtenir rapidement de l'ADN plasmidique pour des analyses aux enzymes de restriction ou pour le séquençage. Cet ADN reste cependant contaminé par des ARN cellulaires.

4.2.2 Préparation à grande échelle d'ADN plasmidique (Maxi-préparation)

Cinq cents mL de milieu LB additionnés par 100 mg/mL d'ampiciline sont ensemencés par des transformants d'*E. coli* TB1 puis cultivés pendant une nuit à 37°C. Les cellules sont récupérées par centrifugation, resuspendues et soumises à une lyse alcaline. Après neutralisation, l'ADN est fractionné sur une cartouche échangeuse d'anions jetable selon le kit « Jet Star Maxi 2 » commercialisé par Genomed.

4.2.3 Préparation d'ADN plasmidique de levure

Nous avons utilisé le protocole décrit par Hoffman et *al.* (1987). Une ansée de cellules de levure est resuspendue dans 1 mL d'eau stérile puis centrifugée 2 min à 12000 × g. Le culot de cellules est mis en suspension dans 250 µL de tampon de broyage (Triton X-100 2% (v/v), SDS 1% (p/v), 100 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA). La suspension est agitée vigoureusement pendant 20 min en présence de 300 mg de billes de verre (0,5 mm de diamètre) et de 250 µL de phénol/chloroforme. Après centrifugation du mélange (5 min à 12000 × g), la phase aqueuse est récupérée puis les acides nucléiques sont précipités par 0,6 volume d'isopropanol. Après lavage à l'éthanol 80% et séchage, le culot est repris par 50 µL d'eau milliQ. Deux µL de cette solution permettent de transformer *E. coli* par électroporation.

4.2.4 Mutagenèse aléatoire de l'ADN par traitement à l'hydroxylamine

Le principe de cette méthode de mutagenèse repose sur le fait que l'hydroxylamine engendre des lésions plus ou moins importantes dans la double hélice d'ADN, ces dernières étant par la suite réparées dans la cellule, créant ainsi des mutations au niveau de ces lésions

(Isackson & Bertrand, 1985). Deux transitions sont retrouvées majoritairement : G→A et C→T. Nous avons fixé les conditions de réaction de manière à obtenir statistiquement une ou deux mutations par molécule d'ADN. Dix µg de plasmide portant le gène que l'on désire muter sont précipités à l'éthanol, repris par 500 µL d'une solution 20 mM phosphate de sodium pH 7, 0,45 M NaOH, 1 M hydroxylamine, puis incubés à 37°C durant 24 heures. La réaction de mutagenèse est arrêtée par l'ajout de 10 µL de NaCl 5 M puis 20 µL de BSA (1 mg/mL). L'ADN est tout d'abord précipité par 1 mL d'éthanol 100%. Après centrifugation à 12000 × g pendant 10 min, le culot est repris par 200 µL de tampon TE puis soumis à une deuxième précipitation à l'éthanol suivie d'un lavage par 500 µL d'éthanol 80% et d'un séchage dans un dessiccateur. Le culot sec est enfin repris par 50 µL de tampon TE.

5 MÉTHODES RELATIVES À L'ÉTUDE DE L'ARN

5.1 Techniques générales de manipulation de l'ARN

5.1.1 Détermination de la concentration de l'ARN

La concentration de l'ARN en solution est déterminée par mesure de l'absorbance à 260 nm. Une unité $A_{260\text{nm}}$ correspond à 40 µg d'ARN par mL.

5.1.2 Précipitation de l'ARN

Cette étape permet de concentrer l'ARN et d'éliminer les substances solubles dans le mélange eau-éthanol. L'ARN est précipité par addition de 2,5 volumes d'éthanol et 1 µg de glycogène à la solution. Après 20 min à -20°C la solution est centrifugée 10 min à 12000 × g.

5.2 Séparation des molécules d'ARN sur gel de polyacrylamide dénaturant

La séparation des molécules d'ARN s'effectue en fonction de leur taille par migration électrophorétique sur des gels de 10% polyacrylamide (rapport acrylamide/bisacrylamide = 19/1) contenant 8 M d'urée dans du tampon TBE. La polymérisation du gel est amorcée par ajout de 240 µL d'une solution de persulfate d'ammonium 10% (p/v) et 60 µL de TEMED (pour 60 mL de solution de polyacrylamide). Lors de la purification d'ARN synthétisés, ces gels sont appelés « préparatifs » et leur épaisseur est de 1 mm. Ils permettent de séparer

l'ARN des autres constituants du milieu de transcription. Dans le cas de séquençage ou de sondage en solution, ces gels qualifiés d'« analytiques » n'ont qu'une épaisseur de 0,4 mm. Ils permettent la séparation de fragments d'ARN en fonction de leur taille avec une résolution allant jusqu'au nucléotide. Avant leur dépôt sur le gel, les ARNs sont repris dans le bleu de dépôt « Formamide Dye » et chauffés 3 min à 65°C pour les dénaturer.

En fonction de la nature des molécules d'ARN à séparer, la durée de migration des gels est variable. Une migration « courte » permet de détecter les plus petits fragments. Elle correspond à l'arrivée du bleu de bromophénol à 5 cm du bas. La migration « longue » laisse descendre le xylène cyanol à 5 cm du bas. Afin de préserver des conditions dénaturantes au cours de la migration, l'électrophorèse se fait à une puissance de 75 watts (pour une plaque de 30 × 60 cm) pour permettre un échauffement du gel suffisant (environ 40-50°C).

5.3 Préparation des molécules d'ARN par transcription *in vitro*

5.3.1 Préparation de l'ADN matrice pour la transcription par l'ARN polymérase du phage T7

Les ARNs de grande taille sont généralement transcrits à partir d'une matrice ADN double brin, qui est clonée dans un plasmide et amplifiée dans des cellules d'*E. coli*. De l'ADN double brin obtenu par PCR peut également servir de matrice. Une matrice partiellement simple brin peut aussi être utilisée pour la synthèse de petits ARNs comme le snRNA U7 ou les ARNs anti-sens utilisés lors des expériences de « primer walk ». En effet, l'ARN polymérase du phage T7 peut transcrire des ADNs simple brin, mais elle nécessite un promoteur double brin pour démarrer la transcription. Dans ce cas, la matrice ADN est obtenue par hybridation d'un 17-mer correspondant au promoteur avec un second oligonucléotide de plus grande taille qui contient la séquence à transcrire en plus de la séquence complémentaire au 17-mer. Les deux oligonucléotides sont obtenus par synthèse chimique (Sigma).

Pour les transcriptions à grande échelle, la matrice privilégiée est d'origine plasmidique. Les gènes sont clonés dans un plasmide de type pUC19. Le clonage du gène débute par la coupure du plasmide pUC19 par l'enzyme XmaI et la déphosphorylation de ses extrémités. Cette étape de linéarisation est suivie de l'insertion de la séquence d'intérêt composée de plusieurs oligonucléotides chevauchants phosphorylés qui sont hybridés puis ligaturés dans le vecteur pUC19. Un choix adéquat des séquences des oligonucléotides nous a

permis de créer un site SmaI en 3' du gène que nous avons ensuite utilisé pour effectuer une transcription « run off ».

Nous avons ainsi construit des vecteurs qui nous ont permis d'obtenir les ARN correspondant aux extrémités 3' non traduites des ARNs pré-messagers d'histones murins H4-12 (Numéro d'accès X13235), H2a-614 (Numéro d'accès AY158924) et H1t (Numéro d'accès AY158908). Les extrémités 3' des ARN pré-messagers choisis ont été définies lors d'études préalables (Dominski & Marzluff, 1999; Streit *et al.*, 1993).

Les plasmides construits sont amplifiés par culture dans *E. coli* TB1. Avant transcription, les vecteurs sont linéarisés par l'enzyme de restriction SmaI. Lors de la transcription, la polymérase arrive au bout de la matrice et « tombe » ce qui provoque son arrêt net (transcription « run-off »). Les transcrits obtenus à l'issue de cette opération commencent par deux résidus guanosine (pour augmenter l'efficacité du démarrage de la transcription) et se terminent par les 3 résidus cytosines du site SmaI. Occasionnellement, la polymérase ajoute un ou deux nucléotides non-codés à l'extrémité 3' du transcrit (Milligan & Uhlenbeck, 1989).

5.3.1.1 Clonage d'une matrice par assemblage d'oligonucléotides

Linéarisation du vecteur pUC19

Cinq µg de plasmide pUC19 sont incubés en présence de 20 U de XmaI dans un volume de 30 µL de tampon NEB4 préconisé par le fournisseur (New England Biolabs) pendant 1 h 30 à 37°C. L'ADN est ensuite déphosphorylé dans 100 µL de tampon BAP (50 mM Tris-HCl pH 9, 1 mM MgCl₂) par 4 U de phosphatase alcaline bactérienne (ou BAP, « Bacterial Alkaline Phosphatase », Amersham) pendant 3 h à 65°C. La réaction est arrêtée par extraction au phénol et au phénol/chloroforme suivie d'une précipitation à l'éthanol.

Assemblage des oligonucléotides et clonage

Les oligonucléotides d'ADN obtenus par synthèse chimique sont amenés à une concentration de 100 µM. Deux µL de chaque oligonucléotide sont phosphorylés séparément dans un volume final de 30 µL de tampon PNK (70 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 5 mM DTT) en présence de 1 mM d'ATP et 5 U de T4 polynucléotide kinase à 37°C pendant 30 min. L'enzyme est inactivée par chauffage à 65°C pendant 10 min.

Dix µL de chaque solution d'oligonucléotide phosphorylé sont ensuite mélangés. L'ensemble est hybridé par incubation à 70°C pendant 5 min puis à 37°C durant 1 heure.

Sept μL de ce mélange et 50 ng de vecteur linéarisé et déphosphorylé sont mélangés et incubés dans un volume final de 10 μL de « tampon ligase » (50 mM Tris-HCl pH 7.6, 10 mM MgCl_2 , 10 mM DTT, 1 mM ATP) pendant une nuit à 16°C en présence de 5 U de T4 ADN ligase. Après précipitation à l'éthanol, le mélange de ligation est repris dans 10 μL d'eau. Une aliquote de 2 μL servira à électroporer des cellules électrocompétentes *E. coli* TB1. Les clones individuels obtenus sur LB-ampicilline après une nuit d'incubation à 37°C sont analysés par minipréparation de leur ADN plasmidique et analyse par digestion *Sma*I. La présence de l'insert peut aussi être vérifiée par amplification par PCR. En dernière étape, les clones sont vérifiés par séquençage de l'ADN plasmidique.

Linéarisation de la matrice plasmidique par l'enzyme *Sma*I

Cinquante μg d'ADN plasmidique purifiés grâce au kit « Jet Star Maxi 2 » sont linéarisés par 200 U de *Sma*I dans 60 μL de tampon NEB4 préconisé par le fournisseur (New England Biolabs) pendant 3 heures à 25°C. L'intégrité de la digestion est vérifiée par électrophorèse d'une aliquote de 1 μL sur un gel d'agarose (1%) avec un témoin plasmidique non digéré. La réaction est arrêtée par extraction au phénol et au phénol/chloroforme suivie d'une précipitation à l'éthanol.

5.3.1.2 Préparation d'une matrice constituée d'oligonucléotides

La matrice est obtenue par hybridation de deux oligonucléotides de synthèse. L'oligonucléotide porteur du gène d'intérêt est hybridé à un 17-mer complémentaire au promoteur de l'ARN polymérase du phage T7 afin de reconstituer une séquence promotrice double brin. Cinquante μg d'oligonucléotide porteur du gène d'intérêt sont incubés avec une quantité équimoléculaire de 17-mer dans un tampon 10 mM Tris-HCl pH 7.5, 50 mM acétate de potassium. L'hybridation s'effectue par incubation du mélange durant 5 min à 95°C, 5 min à 65°C puis 15 min à 25°C.

5.3.1.3 Obtention d'une matrice à transcrire par amplification par PCR

La matrice employée dans la réaction de PCR est le plasmide pUC19 contenant le gène H4-12. À l'aide d'amorces spécifiques, il a été possible d'amplifier différentes formes mutées du gène H4-12 (formes délétées ou formes mutées ponctuellement).

5.3.2 Transcription par la T7 ARN polymérase

Une transcription courante est effectuée par 5-10 µg de T7 ARN polymérase dans 500 µL de tampon TMSDT (40 mM Tris-HCl pH 8.1, 22 mM MgCl₂, 1mM spermidine, 5 mM DTE, Triton X-100 0,01% (v/v)) en présence de 4 mM de chaque NTP, de 8 mM de GMP, 40 U de RNasin et 0,1 µg/µL d'ADN matrice. Le mélange est incubé pendant 3 h à 37°C. Un précipité blanc de pyrophosphate inorganique se forme fréquemment après 1-2 h. Le cas échéant, on ajoute alors 5 µL de pyrophosphatase afin de dégrader le pyrophosphate et déplacer l'équilibre vers la synthèse d'ARN. Dans les cas de matrices non plasmidiques, la réaction de transcription est suivie d'une digestion des matrices de transcription par la DNase I. La réaction est arrêtée par une extraction au phénol-phénol/chloroforme, suivie d'une précipitation à l'éthanol.

5.3.3 Purification des transcrits

Les transcrits sont séparés des NTPs, des ARNs inachevés et de l'ADN par électrophorèse sur un gel préparatif de 10% polyacrylamide dénaturant. Les bandes de gel contenant les transcrits sont repérées sous UV, découpées et placées dans un appareil d'électroélution muni de trois membranes semi-perméables, dont deux séparent un compartiment de rétention d'ARN d'un volume de 600-800 µL. Après 1 h de migration dans un champ électrique de 250 V dans du tampon TBE, le tampon contenu dans le compartiment de rétention est prélevé. L'élution est répétée une fois puis l'ARN est précipité par l'éthanol. Les culots sont repris dans de l'eau milliQ et leurs concentrations sont mesurées.

Dans le cas d'une purification sur gel en conditions natives (sans urée, migration à 4°C), l'ARN est élué des bandes de gel par diffusion passive dans un tampon « cracking buffer » (0,3 M NaCl, 0,5 mM EDTA, 10 mM Tris-HCl pH 7.5). L'élution se fera par agitation lente sur une roue durant une nuit à 4°C.

5.4 Marquage radioactif de l'ARN

5.4.1 Marquage de l'extrémité 5'

Les acides nucléiques peuvent être phosphorylés en 5' par la T4 polynucléotide kinase. Le [γ -³²P] ATP est le substrat de la kinase qui permet de marquer l'extrémité 5' de l'ARN. Dans le cas des ARNs obtenus par transcription *in vitro*, l'extrémité 5' doit

préalablement être libérée des mono- ou triphosphates présents par action de la phosphatase alcaline (BAP) d'*E. coli*.

5.4.1.1 Déphosphorylation des extrémités 5' des transcrits par la BAP

Soixante picomoles d'ARN sont dénaturées à 80°C pendant 3 min dans 45 µL d'eau milliQ, puis déphosphorylées à 37°C pendant 30 min dans un tampon BAP (50 mM Tris-HCl pH 9, 1 mM MgCl₂) en présence de 1 U de BAP. La réaction est arrêtée par une extraction au phénol et au phénol/chloroforme suivie d'une précipitation à l'éthanol.

5.4.1.2 Marquage de l'extrémité 5' par la T4 polynucléotide kinase

Les culots d'ARN déphosphorylés seront repris dans 25 µL d'un tampon PNK (70 mM Tris-HCl, 10 mM MgCl₂, 5 mM DTT pH 7.6 à 25°C) et incubés 30 min à 37°C en présence de 40 µCi de [γ -³²P] ATP (3000 Ci/mme, Amersham) et de 10 U de T4 polynucléotide kinase. La réaction est arrêtée par addition de 2 volumes de solution de dépôt « Formamide Dye » et chauffage des échantillons durant 3 min à 65°C. L'ensemble est ensuite purifié par électrophorèse sur gel préparatif de polyacrylamide 10% dénaturant.

5.4.2 Marquage de l'extrémité 3' par la T4 ARN ligase

Le marquage radioactif d'ARN à son extrémité 3'-terminale peut être réalisé par ligation de diphospho-cytidine radioactive ([5'-³²P]pCp, 3000 Ci/mme, Amersham) par la T4 ARN ligase. Cette enzyme requiert la présence d'une extrémité 3'-OH « acceptrice » sur la molécule à marquer, et un monophosphate en 5' sur la molécule ARN donatrice, ici le [5'-³²P]pCp. Le [5'-³²P]pCp possède en outre un phosphate en 3' qui empêche l'ARN marqué au pCp de devenir substrat pour une nouvelle réaction de marquage. De l'ATP est nécessaire pour fournir de l'énergie à la réaction.

Protocole expérimental :

Un µg d'ARN est incubé dans 20 µL d'un tampon de marquage (50 mM HEPES-NaOH pH 7.5, 10 mM MgCl₂, 5 mM DTE) en présence de DMSO 10% (v/v), BSA 0,1 µg/µL, 100 µM ATP, 30 µCi de [5'-³²P]pCp et 4 U d'ARN ligase à 4°C pendant une nuit. La réaction est arrêtée par addition de deux volumes de solution de dépôt « Formamide Dye » et chauffage des échantillons durant 3 min à 65°C. Le mélange est ensuite purifié par électrophorèse sur gel préparatif de polyacrylamide 10% dénaturant.

5.5 Purification des transcrits marqués

Les transcrits marqués radioactivement sont purifiés par électrophorèse sur gel préparatif de polyacrylamide 10% dénaturant (§ 5.3.3). La superposition du gel avec son autoradiographie permet de repérer l'acide nucléique marqué, qui est excisé, puis élué du gel par diffusion passive sur une roue dans 400 µL de tampon « cracking buffer » (0,3 M NaCl, 0,5 mM EDTA, 10 mM Tris-HCl pH 7.5). L'élution se déroule durant une nuit à 4°C.

L'ARN est ensuite précipité par l'éthanol. Les culots seront repris dans de l'eau milliQ, puis une aliquote d'1 µL est comptée en mode Cerenkov sur 0,1 minute. La solution d'ARN est ensuite ajustée au nombre de cpm désirés par ajout d'eau milliQ.

6 MÉTHODES RELATIVES À LA MANIPULATION DES PROTÉINES

6.1 Concentration des protéines

Pour réaliser les études d'empreinte en solution, nous avons été amenés à concentrer les solutions protéiques d'HBP recombinante produite dans le système baculovirus. Les protéines ont été concentrées par filtration sur membrane microporeuse calibrée retenant les molécules à concentrer. Elle est effectuée par centrifugation. Cette méthode permet aussi l'échange du tampon initial contre un tampon plus adapté aux réactions qui suivront. En fonction des volumes à concentrer nous avons utilisé différents types d'appareils.

Les dispositifs de concentration de type Centricon (Millipore) permettent de concentrer par centrifugation de petits volumes jusqu'à 100 µL. Nous avons utilisé les modèles munis d'une membrane de type YM10 qui retient les molécules d'une masse moléculaire supérieure à 10.000 Da. Ces concentrateurs s'utilisent sous une force centrifuge de 4000 × g pendant une durée qui dépend du volume et de la composition de la solution à filtrer.

Les dispositifs de concentration de type Microcon (Millipore) permettent de concentrer des solutions jusqu'à un volume de 10 µL. Nous avons utilisé les modèles équipés de membranes YM10. Ils s'utilisent sous une force centrifuge de 14.000 × g.

6.2 Gels de retardement

Les expériences de gel-retardement (« bandshift » familièrement) consistent à co-incuber des macromolécules susceptibles de former un ou plusieurs complexes puis de les séparer en fonction de leur taille et de leur encombrement par électrophorèse en conditions natives. Les complexes de haut poids moléculaire et de fort encombrement seront retardés dans leur migration par rapport aux petites molécules compactes. Le marquage radioactif d'un des composants du mélange permet de repérer sa présence dans un complexe par autoradiographie du gel.

Nous avons utilisé la technique de gel-retardement pour tester l'activité de fixation des échantillons protéiques utilisés et vérifier si leur concentration est adéquate. Lors des études d'empreinte en solution, la concentration protéique a été ajustée de façon à fixer un minimum de 80% d'ARN marqué. Les expériences de gel-retardement ont également permis d'apprécier le rôle de la protéine HBP sur la fixation du snRNA U7 à l'ARN pré-messager d'histone. Dans ce cas, c'est un complexe ternaire HBP+ARN pré-messager+snRNA U7 qui a été séparé sur gel de polyacrylamide natif (voir chapitre 2).

Expérimentalement, la formation du complexe ARN:protéine a été effectuée dans les conditions de tampon et de force ionique utilisées lors des expériences de mesure des empreintes en solution, mais en l'absence de la sonde (§ 7.2.4). En règle générale, 50.000 cpm d'ARN marqué ont été utilisés dans un volume final de 15 à 20 μ L. Après formation du complexe par incubation dans la glace, le mélange a été fractionné sur gel de polyacrylamide non dénaturant à faible réticulation (5%) dans un tampon Tris-Glycine (50mM Tris-base, 50 mM Glycine). La taille standard d'un gel utilisé pour les expériences de gel-retardement est de 14×14×0.1 cm (100 mL de gel sont polymérisés par addition de 600 μ L d'une solution de persulfate d'ammonium 10% (p/v) et 100 μ L de TEMED). Le dépôt des échantillons ne requiert pas un tampon spécifique, car ils contiennent du glycérol à hauteur de 10%. La migration s'effectue à 4°C en imposant une puissance de 3W qui évitera un échauffement du gel.

7 MÉTHODES RELATIVES À L'ÉTUDE DE LA STRUCTURE EN SOLUTION DES ARNs ET DE L'EMPREINTE DE LA PROTÉINE HBP

7.1 Principe

La structure en solution des ARNs est étudiée à l'aide de sondes de structure chimiques et enzymatiques qui détectent les zones libres ou appariées de la molécule d'ARN. Les sondes présentent différentes réactivités. Elles peuvent être spécifiques des zones simple brin ou double brin, mais aussi de résidus donnés ou d'une position précise dans les nucléotides (liaison phosphodiester, base). Les sondes enzymatiques induisent des coupures spécifiques dans l'ARN, elles sont plus sensibles à l'encombrement stérique que les sondes chimiques car leur masse moléculaire est plus élevée. Les sondes chimiques ne clivent pas la molécule (excepté le plomb), mais modifient des atomes dans des nucléotides donnés. Une combinaison des résultats obtenus avec des sondes diverses permet d'établir la structure secondaire de la molécule d'ARN en solution. Dans chaque étude, il est nécessaire de réaliser une expérience « témoin » en l'absence de sonde afin d'identifier les coupures spontanées pouvant apparaître au cours des manipulations.

Dans les études d'empreinte protéique (ou « footprint ») sur des molécules d'ARN, c'est le complexe ARN:protéine qui est sondé. La fixation de la protéine sur l'ARN provoque un encombrement stérique qui empêche l'accès des sondes et protège les régions potentiellement réactives. L'action des sondes est donc inhibée au niveau de la région d'interaction. De même, si la fixation de la protéine induit des remaniements structuraux, ils seront repérables par des changements d'accessibilité aux sondes par rapport aux données obtenues avec l'ARN libre. Ainsi, c'est en confrontant les résultats obtenus sur l'ARN libre avec ceux du complexe, que l'on peut définir les zones protégées ou remaniées lors de la fixation de la protéine sur la molécule. Ces expériences s'accompagnent à nouveau d'un certain nombre de « témoins » réalisés en l'absence de sonde qui permettent d'identifier les coupures pouvant être induites par l'ajout de la protéine à l'ARN.

7.2 Les différents sondages

7.2.1 Sondage enzymatique

Le sondage enzymatique est basé sur la digestion statistique de l'ARN par des endoribonucléases (ou RNases), dans des conditions réactionnelles imposant un maximum d'une coupure par molécule testée. La spécificité des coupures varie en fonction des sondes. La majorité des RNases utilisées clivent au niveau des liaisons phosphodiester dans des régions non appariées. Ainsi, la RNase T1 (IRL) coupe les liaisons phosphodiester en 3' des guanines non appariées. La RNase T2 (Invitrogen) est une endonucléase spécifique des régions simple brin sans aucune spécificité de résidu. L'unique sonde enzymatique coupant préférentiellement dans les régions double brin est la RNase V1 (Kemotex).

Protocole expérimental :

Les transcrits marqués par la radioactivité (50.000 cpm) ou froids (2 pmoles) sont préalablement dénaturés 2 minutes à 80°C, renaturés par lente diminution de la température jusqu'à atteindre 35°C puis placés dans la glace. Ils sont ensuite incubés dans le tampon B (10 mM Tris-HCl pH 7.5, 50 mM KCl, 5 mM MgCl₂, 1 mM DTT, glycérol 10 %(v/v)) durant 20 minutes à 0°C en présence de 2 mM de Vanadyl (un analogue de ribonucléoside, inhibiteur de RNase, New England Biolabs), accompagné de RNasin 2U/μL (un second inhibiteur de RNase, Promega), et de tRNA total de levure (Boehringer) 0.67 pmol/μL. L'hydrolyse enzymatique de l'ARN par les RNases se déroule durant 10 minutes à 20°C en présence de chacune des sondes enzymatiques. Dans les conditions statistiques que nous avons mises au point, les quantités d'enzymes utilisées ont été les suivantes : 59.10⁻⁶ U de RNase V1, 1,27 U de RNase T1 ou 5,6 U de RNase T2. Les réactions sont arrêtées par un refroidissement immédiat du milieu réactionnel dans la glace et addition de 15 μL de tampon « stop mix A » (0.6 M NaAc pH 6, 3 mM EDTA et tRNA total 0,1 μg/μL). Après arrêt de l'hydrolyse, les milieux réactionnels sont extraits au phénol puis au phénol/chloroforme. Ils sont ensuite précipités par l'éthanol en présence d'1 μg de glycogène. Les culots sont lavés à deux reprises dans de l'éthanol à 80% puis séchés.

7.2.2 Sondage au plomb

Le sondage au plomb repose sur des clivages chimiques statistiques de l'ARN induits par les ions Pb²⁺ d'une solution d'acétate de plomb. Seules les zones simple brin sont clivées

au niveau des liaisons phosphodiester. Le plomb induit également de fortes coupures au niveau des sites de fixation d'ions divalents. La petite taille des ions Pb^{2+} leur permet d'accéder aux régions les plus encombrées. Ainsi, les ions Pb^{2+} permettent de sonder la structure en solution des molécules d'ARN de manière plus précise.

Protocole expérimental :

Les ARNs à sonder sont dénaturés, renaturés, puis repris dans le tampon B comme décrit au paragraphe précédent (§ 7.2.1). Le protocole diffère à partir de l'étape d'hydrolyse statistique où une incubation de 5 minutes à 20°C est effectuée en présence d'une solution d'acétate de plomb fraîchement préparée à une concentration finale de 10,7 mM. Les réactions sont arrêtées par un refroidissement immédiat du milieu réactionnel dans la glace et addition d'EDTA à une concentration finale de 17 mM. Après arrêt de l'hydrolyse, les milieux réactionnels subissent une extraction au phénol suivie d'une extraction au phénol/chloroforme puis d'une précipitation à l'éthanol. Les culots sont lavés à deux reprises dans de l'éthanol à 80% et séchés.

7.2.3 Sondage chimique

Le principe du sondage chimique repose sur la modification statistique des bases de l'ARN par des composés chimiques, dans des conditions expérimentales limitant le nombre de modifications par molécule testée à un nombre de un. Chaque sonde a une spécificité de modification qui lui est propre et modifie chimiquement un atome précis d'un nucléotide donné (**Figure 27**). La majorité des sondes chimiques utilisées dans les études en solution modifient les atomes des bases libres (non appariées). Ainsi, la CMCT (dérivé de la carbodiimide) modifie préférentiellement, par addition d'un groupement chimique complexe, la position N3 des résidus uraciles et plus faiblement la position N1 des résidus guanines. Le DMS (diméthyl sulfate) méthyle spécifiquement la position N1 des résidus adénines et plus faiblement la position N3 des résidus cytosines. Ainsi, l'utilisation de ces deux composés permet de sonder les 4 bases de l'ARN pour leur accessibilité ou leur implication dans la formation d'un appariement.

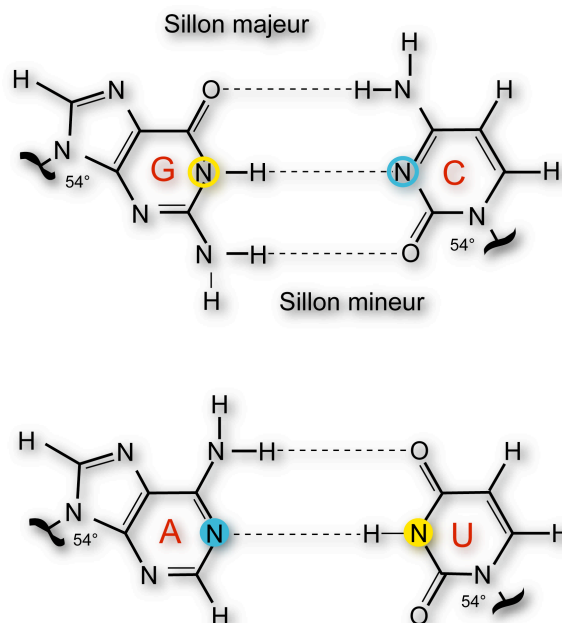


Figure 27 : Représentations schématisques des paires de bases A:U et G:C montrant les atomes modifiés par les sondes chimiques. Les positions encerclées sont réactives aux sondes chimiques (CMCT en jaune et DMS en bleu) lorsque les bases sont dans des régions non appariées et non dissimulées par la protéine. Les ronds symbolisent une réactivité forte, les cercles, une réactivité plus faible.

Protocole expérimental :

Deux pmoles de transcrit froid sont préalablement dénaturées 2 minutes à 80°C, renaturées par lente diminution de la température jusqu'à 35°C puis placées dans la glace. Elles sont ensuite incubées dans un tampon spécifique à chaque sonde. Dans le cas d'un sondage au DMS, il s'agit du tampon B décrit précédemment (§ 7.2.1). La CMCT (Merck) nécessite l'emploi d'un tampon spécifique de pH plus élevé, le tampon B/CMCT (20 mM Tris-HCl pH 8,4, 50 mM KCl, 5 mM MgCl₂, 1 mM DTT, glycérol 10 %). Cette incubation s'effectue pendant 20 minutes à 0°C en présence de 2 mM de Vanadyl (New England Biolabs), de 2 U/μL de RNasin (Promega), et de tRNA total de levure 1 pmol/μL. Après les 20 minutes, la réaction de modification est déclenchée pour une durée de 20 minutes à 20°C. Pour obtenir des conditions statistiques, nous avons utilisé la CMCT à une concentration finale de 5 μg/μL dans le milieu réactionnel. Pour le DMS (Fluka), nous avons utilisé 1 μL d'une dilution au 1/40 de DMS dans de l'éthanol 100% dans un volume réactionnel de 20 μL (dilution finale du DMS au 1/800). Les réactions sont arrêtées par un refroidissement immédiat du milieu réactionnel dans la glace, une extraction au phénol suivie d'une extraction au phénol/chloroforme et d'une précipitation à l'éthanol 100%. Les culots sont ensuite lavés à l'éthanol 80% puis séchés.

7.2.4 Empreinte en solution de la protéine HBP sur les molécules d'ARN

Les conditions expérimentales utilisées pour rechercher l'empreinte de la protéine HBP sur l'ARN sont identiques à celles décrites ci-dessus pour l'ARN libre. La protéine HBP est préalablement dialysée contre le tampon de la réaction et concentrée de façon à pouvoir fixer au moins 80% des molécules marquées dans les conditions du sondage. La protéine HBP est ajoutée après l'étape de dénaturation-renaturation de l'ARN. Elle est suivie de l'incubation de 20 minutes dans la glace pour permettre la formation du complexe HBP:ARN puis de la réaction de clivage ou de modification.

7.2.5 Sondage en solution de l'ARNm bloqué par des ARNs anti-sens

Un protocole identique au sondage en solution est employé en pré-incubant au préalable 20 pmoles de petits ARNs anti-sens avec 2 pmoles de transcrit froid d'ARNm. Dans cette étude, seules les sondes enzymatiques et les ions Pb^{2+} ont été employés. L'hybridation des petits ARNs a été contrôlée par des expériences de digestion à la RNase H (voir § suivant).

7.2.6 Vérification de la bonne hybridation des petits ARNs par digestion à la RNase H

L'expérience de « primer walk » a consisté à tester l'effet inhibiteur de petits ARNs anti-sens sur la traduction de l'ARNm. Il est cependant nécessaire de vérifier la bonne hybridation des ARNs anti-sens, soit par séquençage direct par les oligonucléotides, soit par digestion de l'hétéroduplexe ARN:ADN par la RNase H qui coupe spécifiquement les ARNs hybridés avec les ADN.

Vingt pmoles d'oligonucléotide ADN sont hybridées à 2 pmoles d'ARNm contenant 50.000 cpm d'ARN marqué au ^{32}P au niveau de l'extrémité 5', par un cycle de dénaturation à 80°C durant 2 min, suivi d'un refroidissement lent jusqu'à 35°C puis brutal à 0°C. Le mélange est ensuite incubé dans un tampon RNase H (20 mM Hepes-KOH pH 8, 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTE) durant 5 min à 37°C. L'hydrolyse enzymatique de l'ARN des hétéroduplexes ARN-ADN par la RNase H (Roche) se déroule durant 30 minutes à 37°C dans un volume de 100 μ L en présence d'une unité d'enzyme. Un contrôle de digestion en l'absence d'oligonucléotide ADN est réalisé en parallèle. Les réactions sont arrêtées par ajout

de 13 μL de SDS 10% et 1 U de protéinase K (Roche) et incubation à 37°C durant 10 minutes. Après extraction au phénol puis au phénol/chloroforme, l'ARN est précipité par l'éthanol en présence d'1 μg de glycogène. Les culots sont lavés à deux reprises dans de l'éthanol à 80% puis séchés. Ils sont finalement repris par 20 μL de la solution de dépôt « Formamide Dye ». Des aliquotes de 5 μL sont déposées sur gel dénaturant pour séparer les fragments obtenus par digestion à la RNase H. Après autoradiographie, la position des coupures est assignée par la co-migration de deux échelles d'ARN obtenues par hydrolyse en conditions dénaturantes : l'échelle alcaline et l'échelle T1 (voir § suivant).

7.3 Détection des coupures ou modifications obtenues au cours des sondages en solution

7.3.1 Détection directe des coupures sur un ARN radioactif

Les culots contenant les différents fragments d'ARN sont repris dans une solution de Formamide Dye (20 mM EDTA, Formamide 95% (v/v), bleu de bromophénol 0,1% (p/v), xylène cyanol (p/v)) et placés dans un autre tube. La radioactivité des échantillons est ensuite évaluée par comptage en mode Cerenkov. Des quantités de radioactivité identiques sont ensuite déposées sur gel analytique de polyacrylamide 10 % dénaturant.

Les différents fragments obtenus après clivage sont séparés par électrophorèse en conditions dénaturantes. Après autoradiographie, le profil de coupure de l'ARN libre est analysé et comparé à celui obtenu en présence de protéine. La disparition de certaines coupures définit l'empreinte de la protéine. Elle permet d'identifier les résidus de l'ARN en proche contact avec la protéine.

Les positions des sites de coupures peuvent être assignées en faisant co-migrer deux échelles d'ARN obtenues par hydrolyse en conditions dénaturantes.

Une première échelle est obtenue par hydrolyse à la RNase T1 en conditions dénaturantes en présence de 2,5 U d'enzyme (Peattie & Gilbert, 1980). En principe, cette échelle permet de détecter tous les résidus guanine de la molécule. Cependant, certains résidus guanine ne sont pas réactifs et demeurent invisibles sur l'autoradiographie.

La seconde échelle est obtenue par hydrolyse alcaline de l'ARN en conditions dénaturantes et statistiques. Chaque position peut faire l'objet d'une coupure indépendamment de la nature de la base. Elle est réalisée en incubant 1 μL de transcrit marqué (50.000 cpm) avec 1 μL de tRNA total de levure à 1 $\mu\text{g}/\mu\text{L}$ durant 10 minutes à 80°C dans un

tampon d'hydrolyse (80 mM de Na_2CO_3 et 80 mM de NaHCO_3 le tout ajusté à pH 9). La réaction est arrêtée dans la glace. Puis l'ARN fait l'objet d'une précipitation par l'éthanol, de 2 lavages, d'un séchage avant d'être re-dissout dans une solution de Formamide Dye pour migration.

Ainsi, l'échelle T1 permet l'identification des résidus guanine, qui reportés sur l'échelle alcaline permettront alors l'identification de tous les autres résidus.

7.3.2 Détection indirecte des coupures ou modifications par transcription inverse d'un ARN non marqué

La détection indirecte est utilisée pour localiser les modifications de l'ARN induites par les sondes chimiques. Elle permet également de détecter des coupures et se révèle très utile pour le sondage de longues molécules dont l'analyse par détection directe se révèle être laborieuse ou impossible.

L'ARN modifié ou clivé est hybridé à un oligonucléotide radioactif qui sert d'amorce pour l'élongation par la transcriptase inverse du virus de la myéloblastose aviaire (AMV-RT). En présence des quatre dNTP, la transcriptase inverse synthétise des chaînes d'ADNc dans le sens 5' vers 3', jusqu'à ce que l'élongation de la chaîne soit arrêtée par une modification ou une coupure dans l'ARN (**Figure 28**). Les fragments d'ADNc synthétisés sont ensuite séparés selon leur taille par électrophorèse sur gel analytique de polyacrylamide 10% dénaturant ce qui permet de détecter la position des modifications ou des coupures. Une expérience de contrôle est effectuée parallèlement sur de l'ARN non modifié afin de pouvoir identifier les arrêts de la transcriptase inverse dus à des coupures spontanées dans l'ARN et/ou à des régions de structure secondaire stable. Un séquençage du même ARN est également effectué selon le protocole de Sanger en utilisant la même amorce. Ce séquençage permet de lire la séquence et d'identifier sans ambiguïté les signaux de modification ou de coupure.

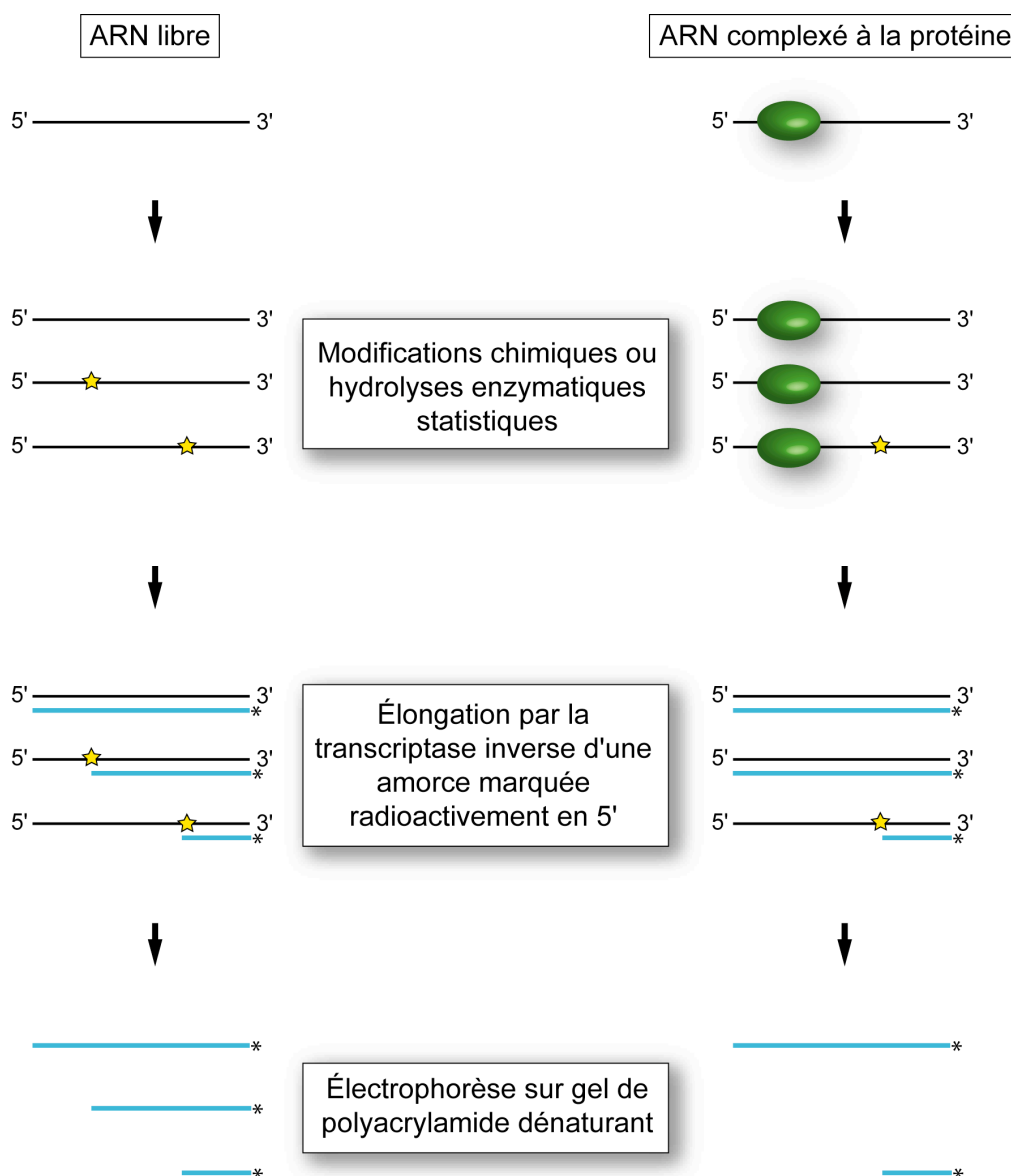


Figure 28 : Principe de détection indirecte des modifications par extension d’amorce. L’ARN libre en solution ou complexé à la protéine (représentée en vert) est soumis à des modifications par des sondes chimiques ou enzymatiques (étoiles jaunes). Ces modifications sont détectées de manière indirecte par la technique d’élongation à la transcriptase inverse (brin bleu). Une séquence est réalisée en parallèle afin d’identifier les bases modifiées. L’empreinte générée par la protéine est détectée par comparaison des profils de migrations des fragments radioactifs.

7.3.2.1 Marquage de l’oligonucléotide amorce pour la détection indirecte

Les oligonucléotides choisis pour s’hybrider à l’ARN ne possèdent pas d’extrémité 5’ phosphate, il n’est donc pas nécessaire de les déphosphoryler, le marquage peut s’effectuer directement par phosphorylation à la T4 polynucléotide kinase. Dix pmoles d’oligonucléotide synthétique sont incubées dans un volume final de 25 µL de tampon PNK (70 mM Tris-HCl pH 7,6, 10 mM MgCl₂, 5 mM DTT) en présence de 40 µCi de [γ -³²P] ATP et 5 U de T4 polynucléotide kinase durant 30 min à 37°C. L’enzyme est inactivée par addition de 125 µL

d'eau milliQ et chauffage à 65°C pendant 10 min. Les oligonucléotides marqués seront ensuite séparés du [γ -³²P] ATP non incorporé par filtration sur colonne d'exclusion. Une suspension de 1 mL de résine Sephadex-G25 (Pharmacia) dans de l'eau milliQ est transférée dans le corps d'une seringue de 1 mL dont le fond est tapissé d'une fine couche de laine de verre. Le surplus d'eau contenue dans la résine est évacué par centrifugation pendant 5 minutes à 1000 rpm. Le milieu de marquage est ensuite chargé sur la résine puis centrifugé 5 minutes à 1000 rpm dans un tube de 15 mL. Un volume d'éluat équivalent au volume de la réaction de marquage est recueilli. Un μ L de l'éluat est placé dans un tube Eppendorf et son activité est comptée en mode Cerenkov. La solution d'oligonucléotide marquée est ensuite ajustée à la concentration en radioactivité désirée par ajout d'eau milliQ.

7.3.2.2 Élongation de l'amorce marquée par transcription inverse

L'amorce d'ADN marquée, d'une longueur moyenne de 15 à 20 nucléotides, est complémentaire à l'ARN que l'on désire « lire » pour ses modifications ou coupures. Les ARNs froids modifiés servent quant à eux de matrice à cette amorce radioactive.

Les culots secs obtenus en fin de sondage enzymatique ou chimique contiennent 2 pmol d'ARN sondé. Ils seront repris dans 4 μ L d'eau milliQ. Une aliquote d'1 μ L (100.000 cpm) de l'amorce est hybridée aux 2 pmoles d'ARN matrice par chauffage à 90°C pendant 1 min, refroidissement dans la glace pendant 1 min et incubation de 5 min à température ambiante. L'élongation s'effectue dans 15 μ L de tampon RTB (50 mM Tris-HCl pH 8.3, 40 mM KCl, 6 mM MgCl₂) à 37°C pendant 30 min en présence de 2 U de AMV-RT (Q-biogene) et 333 μ M de chacun des dNTP. La réaction est arrêtée par hydrolyse alcaline de l'ARN matrice (§ 7.3.2.4).

7.3.2.3 Séquençage par transcription inverse

Cette méthode suit le principe du séquençage de l'ADN (Sanger *et al.*, 1977). Une amorce radioactive hybridée à la matrice à séquencer est allongée par une polymérase. Dans quatre réactions parallèles, l'élongation de l'amorce est statistiquement arrêtée par incorporation d'un didésoxynucléotide spécifique. Dans notre cas, les ADN polymérases habituellement utilisées pour le séquençage d'ADN ont été remplacées par la transcriptase inverse du virus de la myéloblastose aviaire (AMV-RT). Le séquençage conduit à la lecture de l'enchaînement des nucléotides du fragment d'ARN ce qui permet ensuite l'identification des modifications ou coupures induites lors des expériences de sondage.

Une aliquote de 4 μL de l'amorce (100.000 cpm/ μL) est hybridée à 4 μL d'une solution contenant 0.1 $\mu\text{g}/\mu\text{L}$ d'ARN matrice à séquencer. L'hybridation se déroule dans un volume de 12 μL par chauffage du mélange à 90°C pendant 1 min, refroidissement dans la glace pendant 1 min et incubation durant 15 min dans un tampon RTB à température ambiante. Le mélange est divisé en 4 tubes à raison de 5 μL par tube. Chaque tube représente la séquence spécifique à un des 4 nucléotides.

L'élongation se fait dans un volume réactionnel de 15 μL de tampon RTB pendant 30 min à 37°C en présence de 1 U d'AMV-RT (Q-biogene), 100 μM de chacun des dNTP (sauf dXTP), 25 μM dXTP et 3,33 μM ddXTP. La réaction est arrêtée par hydrolyse alcaline de l'ARN matrice (§ 7.3.2.4).

7.3.2.4 Hydrolyse alcaline de l'ARN matrice

Cette étape permet d'éliminer la matrice ARN des réactions de transcription inverse. Il s'en suit une meilleure résolution des fragments d'ADNc lors de la migration sur gel dénaturant de polyacrylamide 10%. L'hydrolyse consiste à ajouter 3 μL de KOH 3 M et 20 μL d'un tampon de lyse (50 mM Tris-HCl pH 8.0, 7,5 mM EDTA, SDS 0,5%). Le mélange est incubé 3 min à 90°C puis une nuit à 37°C. Il est ensuite neutralisé par ajout de 6 μL d'acide acétique 6 M, puis précipité à l'éthanol 100%. Les culots sont lavés à deux reprises dans de l'éthanol 80% et séchés. Les fragments d'ADNc sont finalement repris dans une solution de Formamide Dye puis fractionnés sur gel dénaturant de polyacrylamide 10%.

Références

- Adams, C. & Kamakaka, R. (1999). Chromatin assembly: biochemical identities and genetic redundancy. *Curr Opin Genet Dev* **9**, 185-90.
- Adamson, T. & Price, D. (2003). Cotranscriptional processing of Drosophila histone mRNAs. *Mol Cell Biol* **23**, 4046-55.
- Ahmad, K. & Henikoff, S. (2002). Epigenetic consequences of nucleosome dynamics. *Cell* **111**, 281-4.
- Annunziato, A. (2005). Split decision: what happens to nucleosomes during DNA replication? *J Biol Chem* **280**, 12065-8.
- Ausio, J., Abbott, D., Wang, X. & Moore, S. (2001). Histone variants and histone modifications: a structural perspective. *Biochem Cell Biol* **79**, 693-708.
- Azzouz, T., Gruber, A. & Schümperli, D. (2005a). U7 snRNP-specific Lsm11 protein: dual binding contacts with the 100 kDa zinc finger processing factor (ZFP100) and a ZFP100-independent function in histone RNA 3' end processing. *Nucleic Acids Res* **33**, 2106-17.
- Azzouz, T., Pillai, R., Dapp, C., Chari, A., Meister, G., Kambach, C., Fischer, U. & Schümperli, D. (2005b). Towards an assembly line for U7 snRNPs: Interactions of U7-specific Lsm proteins with PRMT5-and SMN-complexes. *J Biol Chem* **280**, 34435-40.
- Bab, I., Smith, E., Gavish, H., Attar-Namdar, M., Chorev, M., Chen, Y., Muhlrads, A., Birnbaum, M., Stein, G. & Frenkel, B. (1999). Biosynthesis of osteogenic growth peptide via alternative translational initiation at AUG85 of histone H4 mRNA. *J Biol Chem* **274**, 14474-81.
- Battle, D. & Doudna, J. (2001). The stem-loop binding protein forms a highly stable and specific complex with the 3' stem-loop of histone mRNAs. *RNA* **7**, 123-32.
- Birnboim, H. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**, 1513-23.
- Bowman, T., Kaludov, N., Klein, M. & Hurt, M. (1996). An H3 coding region regulatory element is common to all four nucleosomal classes of mouse histone-encoding genes. *Gene* **176**, 1-8.
- Campbell, S., Li Del Olmo, M., Beglan, P. & Bond, U. (2002). A sequence element downstream of the yeast HTB1 gene contributes to mRNA 3' processing and cell cycle regulation. *Mol Cell Biol* **22**, 8415-25.

- Chakravarthy, S., Park, Y., Chodaparambil, J. & Edayathumangalam, R. (2005). Structure and dynamic properties of nucleosome core particles. *FEBS Lett* **579**, 895-8.
- Cheng, Y. & Patel, D. (2004). Crystallographic structure of the nuclease domain of 3'hExo, a DEDDh family member, bound to rAMP. *J Mol Biol* **343**, 305-12.
- Chodchoy, N., Pandey, N. & Marzluff, W. (1991). An intact histone 3'-processing site is required for transcription termination in a mouse histone H2a gene. *Mol Cell Biol* **11**, 497-509.
- Clark-Adams, C., Norris, D., Osley, M., Fassler, J. & Winston, F. (1988). Changes in histone gene dosage alter transcription in yeast. *Genes Dev* **2**, 150-9.
- Davey, C., Sargent, D., Luger, K., Maeder, A. & Richmond, T. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol* **319**, 1097-113.
- DeBry, R. & Marzluff, W. (1994). Selection on silent sites in the rodent H3 histone gene family. *Genetics* **138**, 191-202.
- DeJong, E., Marzluff, W. & Nikonowicz, E. (2002). NMR structure and dynamics of the RNA-binding site for the histone mRNA stem-loop binding protein. *RNA* **8**, 83-96.
- Dominski, Z., Erkmann, J., Greenland, J. & Marzluff, W. (2001). Mutations in the RNA binding domain of stem-loop binding protein define separable requirements for RNA binding and for histone pre-mRNA processing. *Mol Cell Biol* **21**, 2008-17.
- Dominski, Z., Erkmann, J., Yang, X., Sanchez, R. & Marzluff, W. (2002). A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3'-end processing. *Genes Dev* **16**, 58-71.
- Dominski, Z. & Marzluff, W. (1999). Formation of the 3' end of histone mRNA. *Gene* **239**, 1-14.
- Dominski, Z., Sumerel, J., Hanson, R. & Marzluff, W. (1995). The polyribosomal protein bound to the 3' end of histone mRNA can function in histone pre-mRNA processing. *RNA* **1**, 915-23.
- Dominski, Z., Yang, X., Kaygun, H., Dadlez, M. & Marzluff, W. (2003). A 3' exonuclease that specifically interacts with the 3' end of histone mRNA. *Mol Cell* **12**, 295-305.
- Dominski, Z., Yang, X. & Marzluff, W. (2005). The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* **123**, 37-48.
- Dominski, Z., Zheng, L., Sanchez, R. & Marzluff, W. (1999). Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. *Mol Cell Biol* **19**, 3561-70.

- Dower, W., Miller, J. & Ragsdale, C. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16**, 6127-45.
- Eckner, R., Ellmeier, W. & Birnstiel, M. L. (1991). Mature mRNA 3' end formation stimulates RNA export from the nucleus. *EMBO J* **10**, 3513-22.
- Eliassen, K., Baldwin, A., Sikorski, E. & Hurt, M. (1998). Role for a YY1-binding element in replication-dependent mouse histone gene expression. *Mol Cell Biol* **18**, 7106-18.
- Erkman, J., Sanchez, R., Treichel, N., Marzluff, W. & Kutay, U. (2005a). Nuclear export of metazoan replication-dependent histone mRNAs is dependent on RNA length and is mediated by TAP. *RNA* **11**, 45-58.
- Erkman, J., Wagner, E., Dong, J., Zhang, Y., Kutay, U. & Marzluff, W. (2005b). Nuclear import of the stem-loop binding protein and localization during the cell cycle. *Mol Biol Cell* **16**, 2960-71.
- Fan, H., He, X., Kingston, R. & Narlikar, G. (2003). Distinct strategies to make nucleosomal DNA accessible. *Mol Cell* **11**, 1311-22.
- Furger, A., Schaller, A. & Schümperli, D. (1998). Functional importance of conserved nucleotides at the histone RNA 3' processing site. *RNA* **4**, 246-56.
- Futterer, J., Kiss-Laszlo, Z. & Hohn, T. (1993). Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. *Cell* **73**, 789-802.
- Gallie, D., Lewis, N. & Marzluff, W. (1996). The histone 3'-terminal stem-loop is necessary for translation in Chinese hamster ovary cells. *Nucleic Acids Res* **24**, 1954-62.
- Gietz, D., St. Jean, A., Woods, R. & Schiestl, R. (1992). Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res* **20**, 1425.
- Gorgoni, B., Andrews, S., Schaller, A., Schümperli, D., Gray, N. & Müller, B. (2005). The stem-loop binding protein stimulates histone translation at an early step in the initiation pathway. *RNA* **11**, 1030-42.
- Govin, J., Caron, C., Lestrat, C., Rousseaux, S. & Khochbin, S. (2004). The role of histones in chromatin remodelling during mammalian spermiogenesis. *Eur J Biochem* **271**, 3459-69.
- Graves, R., Pandey, N., Chodchoy, N. & Marzluff, W. (1987). Translation is required for regulation of histone mRNA degradation. *Cell* **48**, 615-26.
- Gu, X. & Marzluff, W. (1996). 3' Processing and termination of mouse histone transcripts synthesized *in vitro* by RNA polymerase II. *Nucleic Acids Res* **24**, 3797-805.
- Gunjan, A., Paik, J. & Verreault, A. (2005). Regulation of histone synthesis and nucleosome assembly. *Biochimie* **87**, 625-35.

- Gunjan, A. & Verreault, A. (2003). A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in *S. cerevisiae*. *Cell* **115**, 537-49.
- Hassa, P. & Hottiger, M. (2005). An epigenetic code for DNA damage repair pathways? *Biochem Cell Biol* **83**, 270-85.
- Henikoff, S., Furuyama, T. & Ahmad, K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. *Trends Genet* **20**, 320-6.
- Hoffman, C. S. & Winston, F. (1987). A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* **57**, 267-72.
- Horvath, G., Dasgupta, A., Kistler, M. & Kistler, W. (2003). The rat histone H1d gene has intragenic activating sequences that are absent from the testis-specific variant H1t. *Biochim Biophys Acta* **1625**, 165-72.
- Huang, Y. & Carmichael, G. (1997). The mouse histone H2a gene contains a small element that facilitates cytoplasmic accumulation of intronless gene transcripts and of unspliced HIV-1-related mRNAs. *Proc Natl Acad Sci U S A* **94**, 10104-9.
- Huang, Y., Gattoni, R., Stevenin, J. & Steitz, J. (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell* **11**, 837-43.
- Huang, Y. & Steitz, J. (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol Cell* **7**, 899-905.
- Huynen, M., Konings, D. & Hogeweg, P. (1992). Equal G and C contents in histone genes indicate selection pressures on mRNA secondary structure. *J Mol Evol* **34**, 280-91.
- Isackson, P. & Bertrand, K. (1985). Dominant negative mutations in the Tn10 tet repressor: evidence for use of the conserved helix-turn-helix motif in DNA binding. *Proc Natl Acad Sci U S A* **82**, 6226-30.
- Jaeger, S., Barends, S., Giegé, R., Eriani, G. & Martin, F. (2005). Expression of metazoan replication-dependent histone genes. *Biochimie* **87**, 827-34.
- Kaludov, N., Pabon-Pena, L. & Hurt, M. (1996). Identification of a second conserved element within the coding sequence of a mouse H3 histone gene that interacts with nuclear factors and is necessary for normal expression. *Nucleic Acids Res* **24**, 523-31.
- Kamakaka, R. & Biggins, S. (2005). Histone variants: deviants? *Genes Dev* **19**, 295-310.
- Kapp, L. & Lorsch, J. (2004). The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* **73**, 657-70.
- Kassabov, S., Zhang, B., Persinger, J. & Bartholomew, B. (2003). SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol Cell* **11**, 391-403.

- Kaygun, H. & Marzluff, W. (2005a). Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. *Nat Struct Mol Biol* **12**, 794-800.
- Kaygun, H. & Marzluff, W. (2005b). Translation termination is involved in histone mRNA degradation when DNA replication is inhibited. *Mol Cell Biol* **25**, 6879-88.
- Kennedy, S., Wang, D. & Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**, 645-9.
- Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. *Cell* **116**, 259-72.
- Kim, Y., Furic, L., Desgroseillers, L. & Maquat, L. (2005). Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. *Cell* **120**, 195-208.
- Kodama, Y., Rothman, J. H., Sugimoto, A. & Yamamoto, M. (2002). The stem-loop binding protein CDL-1 is required for chromosome condensation, progression of cell death and morphogenesis in *Caenorhabditis elegans*. *Development* **129**, 187-96.
- Koessler, H., Kahle, J., Bode, C., Doenecke, D. & Albig, W. (2004). Human replication-dependent histone H3 genes are activated by a tandemly arranged pair of two CCAAT boxes. *Biochem J* **384**, 317-26.
- Kozak, M. (2004). How strong is the case for regulation of the initiation step of translation by elements at the 3' end of eukaryotic mRNAs? *Gene* **343**, 41-54.
- Ling, J., Morley, S., Pain, V., Marzluff, W. & Gallie, D. (2002). The histone 3'-terminal stem-loop-binding protein enhances translation through a functional and physical interaction with eukaryotic initiation factor 4G (eIF4G) and eIF3. *Mol Cell Biol* **22**, 7853-67.
- Lippman, Z. & Martienssen, R. (2004). The role of RNA interference in heterochromatic silencing. *Nature* **431**, 364-70.
- Liu, T., Levine, B., Skoultchi, A. & Marzluff, W. (1989). The efficiency of 3'-end formation contributes to the relative levels of different histone mRNAs. *Mol Cell Biol* **9**, 3499-508.
- Lowndes, N. & Toh, G. (2005). DNA repair: the importance of phosphorylating histone H2AX. *Curr Biol* **15**, R99-R102.
- Maldonado, R. & Herr, A. (1998). Efficiency of T4 gene 60 translational bypassing. *J Bacteriol* **180**, 1822-30.
- Martin, F., Michel, F., Zenklusen, D., Muller, B. & Schümperli, D. (2000). Positive and negative mutant selection in the human histone hairpin-binding protein using the yeast three-hybrid system. *Nucleic Acids Res* **28**, 1594-603.

- Martin, F., Schaller, A., Eglite, S., Schümperli, D. & Müller, B. (1997). The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. *EMBO J* **16**, 769-78.
- Marzluff, W. (2005). Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Curr Opin Cell Biol* **17**, 274-80.
- Marzluff, W. & Duronio, R. (2002). Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Curr Opin Cell Biol* **14**, 692-9.
- Marzluff, W., Gongidi, P., Woods, K., Jin, J. & Maltais, L. (2002). The human and mouse replication-dependent histone genes. *Genomics* **80**, 487-98.
- Meeks-Wagner, D. & Hartwell, L. (1986). Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. *Cell* **44**, 43-52.
- Meier, V. S., Bohni, R. & Schümperli, D. (1989). Nucleotide sequence of two mouse histone H4 genes. *Nucleic Acids Res* **17**, 795.
- Melin, L., Soldati, D., Mital, R., Streit, A. & Schümperli, D. (1992). Biochemical demonstration of complex formation of histone pre-mRNA with U7 small nuclear ribonucleoprotein and hairpin binding factors. *EMBO J* **11**, 691-7.
- Mello, J. & Almouzni, G. (2001). The ins and outs of nucleosome assembly. *Curr Opin Genet Dev* **11**, 136-41.
- Miele, A., Braastad, C., Holmes, W., Mitra, P., Medina, R., Xie, R., Zaidi, S., Ye, X., Wei, Y., Harper, J., van Wijnen, A., Stein, J. & Stein, G. (2005). HiNF-P directly links the cyclin E/CDK2/p220NPAT pathway to histone H4 gene regulation at the G1/S phase cell cycle transition. *Mol Cell Biol* **25**, 6140-53.
- Milligan, J. & Uhlenbeck, O. (1989). Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* **180**, 51-62.
- Nelson, D., Ye, X., Hall, C., Santos, H., Ma, T., Kao, G., Yen, T., Harper, J. & Adams, P. (2002). Coupling of DNA synthesis and histone synthesis in S phase independent of cyclin/cdk2 activity. *Mol Cell Biol* **22**, 7459-72.
- Norris, D., Dunn, B. & Osley, M. (1988). The effect of histone gene deletions on chromatin structure in *Saccharomyces cerevisiae*. *Science* **242**, 759-61.
- Palko, L., Bass, H., Beyrouthy, M. & Hurt, M. (2004). The Yin Yang-1 (YY1) protein undergoes a DNA-replication-associated switch in localization from the cytoplasm to the nucleus at the onset of S phase. *J Cell Sci* **117**, 465-76.

- Pandey, N., Williams, A., Sun, J., Brown, V., Bond, U. & Marzluff, W. (1994). Point mutations in the stem-loop at the 3' end of mouse histone mRNA reduce expression by reducing the efficiency of 3' end formation. *Mol Cell Biol* **14**, 1709-20.
- Peattie, D. & Gilbert, W. (1980). Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci U S A* **77**, 4679-82.
- Peltz, S., Brewer, G., Groppi, V. & Ross, J. (1989). Exonuclease activity that degrades histone mRNA is stable when DNA or protein synthesis is inhibited. *Mol Biol Med* **6**, 227-38.
- Peterson, C. & Laniel, M. (2004). Histones and histone modifications. *Curr Biol* **14**, R546-51.
- Pettitt, J., Crombie, C., Schümperli, D. & Müller, B. (2002). The *Caenorhabditis elegans* histone hairpin-binding protein is required for core histone gene expression and is essential for embryonic and postembryonic cell division. *J Cell Sci* **115**, 857-66.
- Pillai, R., Grimmmler, M., Meister, G., Will, C., Lührmann, R., Fischer, U. & Schümperli, D. (2003). Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. *Genes Dev* **17**, 2321-33.
- Pillai, R. S., Will, C. L., Lührmann, R., Schümperli, D. & Müller, B. (2001). Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein. *EMBO J* **20**, 5470-9.
- Piontkivska, H., Rooney, A. & Nei, M. (2002). Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol Biol Evol* **19**, 689-97.
- Poirier, R., Lemaire, I., Dumont, M., Leduc, N., Le, H. & Lemaire, S. (2005). Correlation between the expression of the histone H4 mRNA variant H4-v.1 and the levels of histone H4-(86-100) and H4-(89-102) (OGP) in various rat tissues and alveolar macrophages. *Peptides* **26**, 1503-11.
- Reichheld, J., Gigot, C. & Chaubet-Gigot, N. (1998). Multilevel regulation of histone gene expression during the cell cycle in tobacco cells. *Nucleic Acids Res* **26**, 3255-62.
- Richmond, T. & Davey, C. (2003). The structure of DNA in the nucleosome core. *Nature* **423**, 145-50.
- Ross, J., Kobs, G., Brewer, G. & Peltz, S. (1987). Properties of the exonuclease activity that degrades H4 histone mRNA. *J Biol Chem* **262**, 9374-81.
- Sanchez, R. & Marzluff, W. (2002). The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. *Mol Cell Biol* **22**, 7093-104.
- Sanger, F., Nicklen, S. & Coulson, A. (1977). DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-7.

- Sarma, K. & Reinberg, D. (2005). Histone variants meet their match. *Nat Rev Mol Cell Biol* **6**, 139-49.
- Scharl, E. & Steitz, J. (1994). The site of 3' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. *EMBO J* **13**, 2432-40.
- Schümperli, D. & Pillai, R. (2004). The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein. *Cell Mol Life Sci* **61**, 2560-70.
- SenGupta, D. J., Zhang, B., Kraemer, B., Pochart, P., Fields, S. & Wickens, M. (1996). A three-hybrid system to detect RNA-protein interactions in vivo. *Proc Natl Acad Sci U S A* **93**, 8496-501.
- Spycher, C., Streit, A., Stefanovic, B., Albrecht, D., Koning, T. & Schümperli, D. (1994). 3' end processing of mouse histone pre-mRNA: evidence for additional base-pairing between U7 snRNA and pre-mRNA. *Nucleic Acids Res* **22**, 4023-30.
- Streit, A., Koning, T., Soldati, D., Melin, L. & Schümperli, D. (1993). Variable effects of the conserved RNA hairpin element upon 3' end processing of histone pre-mRNA in vitro. *Nucleic Acids Res* **21**, 1569-75.
- Sullivan, E., Santiago, C., Parker, E. D., Dominski, Z., Yang, X., Lanzotti, D. J., Ingledue, T. C., Marzluff, W. & Duronio, R. (2001). Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. *Genes Dev* **15**, 173-87.
- Sun, J., Pilch, D. & Marzluff, W. (1992). The histone mRNA 3' end is required for localization of histone mRNA to polyribosomes. *Nucleic Acids Res* **20**, 6057-66.
- Takyar, S., Hickerson, R. & Noller, H. (2005). mRNA helicase activity of the ribosome. *Cell* **120**, 49-58.
- Thapar, R., Marzluff, W. & Redinbo, M. (2004a). Electrostatic contribution of serine phosphorylation to the Drosophila SLBP-histone mRNA complex. *Biochemistry* **43**, 9401-12.
- Thapar, R., Mueller, G. & Marzluff, W. (2004b). The N-terminal domain of the Drosophila histone mRNA binding protein, SLBP, is intrinsically disordered with nascent helical structure. *Biochemistry* **43**, 9390-400.
- Urrutia, R. (2003). KRAB-containing zinc-finger repressor proteins. *Genome Biol* **4**, 231.
- Vasserot, A., Schaufele, F. & Birnstiel, M. (1989). Conserved terminal hairpin sequences of histone mRNA precursors are not involved in duplex formation with the U7 RNA but act as a target site for a distinct processing factor. *Proc Natl Acad Sci U S A* **86**, 4345-9.

- Wagner, E., Berkow, A. & Marzluff, W. (2005). Expression of an RNAi-resistant SLBP restores proper S-phase progression. *Biochem Soc Trans* **33**, 471-3.
- Walther, T., Wittop Koning, T., Schümperli, D. & Müller, B. (1998). A 5'-3' exonuclease activity involved in forming the 3' products of histone pre-mRNA processing *in vitro*. *RNA* **4**, 1034-46.
- Wang, Z. , Whitfield, M. , Ingledue, T., Dominski, Z. & Marzluff, W. (1996). The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes Dev* **10**, 3028-40.
- Wei, Y., Jin, J. & Harper, J. (2003). The cyclin E/Cdk2 substrate and Cajal body component p220(NPAT) activates histone transcription through a novel LisH-like domain. *Mol Cell Biol* **23**, 3669-80.
- Whitfield, M., Kaygun, H., Erkmann, J., Townley-Tilson, W., Dominski, Z. & Marzluff, W. (2004). SLBP is associated with histone mRNA on polyribosomes as a component of the histone mRNP. *Nucleic Acids Res* **32**, 4833-42.
- Whitfield, M., Zheng, L., Baldwin, A., Ohta, T., Hurt, M. & Marzluff, W. (2000). Stem-loop binding protein, the protein that binds the 3' end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms. *Mol Cell Biol* **20**, 4188-98.
- Williams, A., Ingledue, T., Kay, B. & Marzluff, W. (1994). Changes in the stem-loop at the 3' terminus of histone mRNA affects its nucleocytoplasmic transport and cytoplasmic regulation. *Nucleic Acids Res* **22**, 4660-6.
- Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strain: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-19.
- Yueh, A. & Schneider, R. (1996). Selective translation initiation by ribosome jumping in adenovirus-infected and heat-shocked cells. *Genes Dev* **10**, 1557-67.
- Zanier, K., Luyten, I., Crombie, C., Müller, B., Schümperli, D., Linge, J., Nilges, M. & Sattler, M. (2002). Structure of the histone mRNA hairpin required for cell cycle regulation of histone gene expression. *RNA* **8**, 29-46.
- Zhao, J. (2004). Coordination of DNA synthesis and histone gene expression during normal cell cycle progression and after DNA damage. *Cell Cycle* **3**, 695-7.
- Zhao, J., Kennedy, B., Lawrence, B., Barbie, D., Matera, A., Fletcher, J. & Harlow, E. (2000). NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. *Genes Dev* **14**, 2283-97.

- Zhao, X., McKillop-Smith, S. & Müller, B. (2004). The human histone gene expression regulator HBP/SLBP is required for histone and DNA synthesis, cell cycle progression and cell proliferation in mitotic cells. *J Cell Sci* **1**, 6043-51.
- Zheng, L., Dominski, Z., Yang, X., Elms, P., Raska, C., Borchers, C. & Marzluff, W. (2003). Phosphorylation of stem-loop binding protein (SLBP) on two threonines triggers degradation of SLBP, the sole cell cycle-regulated factor required for regulation of histone mRNA processing, at the end of S phase. *Mol Cell Biol* **23**, 1590-601.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-15.

RÉSUMÉ

Les recherches présentées dans ce mémoire ont porté sur l'expression des gènes d'histones de type réplication-dépendants. Ces gènes sont particuliers car leurs ARNm sont dépourvus d'introns et de queue poly A en 3', l'extrémité 3' étant générée par coupure endonucléolytique au cours d'un processus de maturation original impliquant plusieurs protéines et la snRNP U7.

Lors d'une première étude, nous avons étudié l'étape initiale de la réaction de maturation qui consiste en la fixation de la protéine HBP sur une structure de l'ARN pré-messenger. À partir de mutants de HBP abolissant la fixation sur l'ARN nous avons sélectionné par la technique du triple hybride dans la levure des supprimeurs intragéniques permettant de restaurer cette fixation. La plupart des mutations isolées se situaient dans les domaines N- et C-terminaux de la protéine, en dehors du domaine central impliqué dans l'interaction avec l'ARN. Cette restauration s'effectuait sans perte de spécificité pour la séquence de fixation à l'ARN, suggérant que les domaines N- et C-terminaux sont impliqués dans le processus de reconnaissance de l'ARN.

Dans un second volet de notre étude portant sur la réaction de maturation de l'extrémité 3' de l'ARNm d'histone, nous avons examiné l'impact structural induit par la protéine HBP lors de sa fixation sur les extrémités 3' non traduites des ARNs pré-messagers des histones H4-12, H1t et H2a-614. En utilisant les techniques de sondage en solution nous avons montré que ces extrémités présentent de fortes structures secondaires qui pourraient empêcher l'accès à la particule snRNP U7. Puis, nous avons montré que la fixation de la protéine HBP engendrait des changements de conformation de l'ARN au niveau de la séquence d'hybridation au snRNA U7. Enfin, nous avons pu montrer que ces changements de conformation étaient associés à une amélioration de l'ancrage du snRNA U7 à l'ARN pré-messenger. Cependant, ce mécanisme n'est pas généralisable à l'ensemble des gènes d'histones puisque aucune modification importante n'a pu être détectée à l'extrémité 3' non traduite du gène d'histone H2a-614.

Enfin, nous avons étudié la traduction in vitro de l'ARNm H4-12 et montré qu'elle s'effectuait de façon très efficace en l'absence des régions 5' et 3' non codantes, suggérant que l'initiation de la traduction s'effectuerait par recrutement des ribosomes à l'intérieur de la phase codante. Par sondage en solution de l'ARNm entier, nous avons proposé un modèle de repliement secondaire dans lequel la séquence codante est circularisée par l'hybridation de ses extrémités. À l'aide d'ARNs anti-sens nous avons identifié un certain nombre de nucléotides essentiels pour le maintien d'une haute efficacité de traduction. L'ensemble des résultats obtenus lors des expériences de sondage en solution associés à ceux issus des études de traduction in vitro, de « toe print » et de microscopie électronique, ont conduit à l'établissement d'un modèle original permettant d'expliquer la traduction atypique de l'ARNm H4-12. La phase codante recruterait directement deux ribosomes à la manière de 2 sites d'entrée interne du ribosome ou « IRES ». Le premier ribosome serait recruté au niveau du codon d'initiation, le second près de la fin de la phase codante. Par le jeu de changements de conformation et grâce à la circularisation de l'ARNm, le deuxième ribosome pourrait être dirigé très efficacement sur le codon d'initiation. Cet enchaînement des deux ribosomes sur la phase codante permettrait d'expliquer d'une part, la grande efficacité de traduction observée dans le cas de notre modèle H4-12 et d'autre part, le rôle accessoire des séquences non traduites.

Mots clés : protéine HBP humaine, ARNm de l'histone H4, triple hybride, sélections génétiques, structure en solution, maturation en 3', snRNP U7, traduction in vitro, IRES, relations structure-fonction.